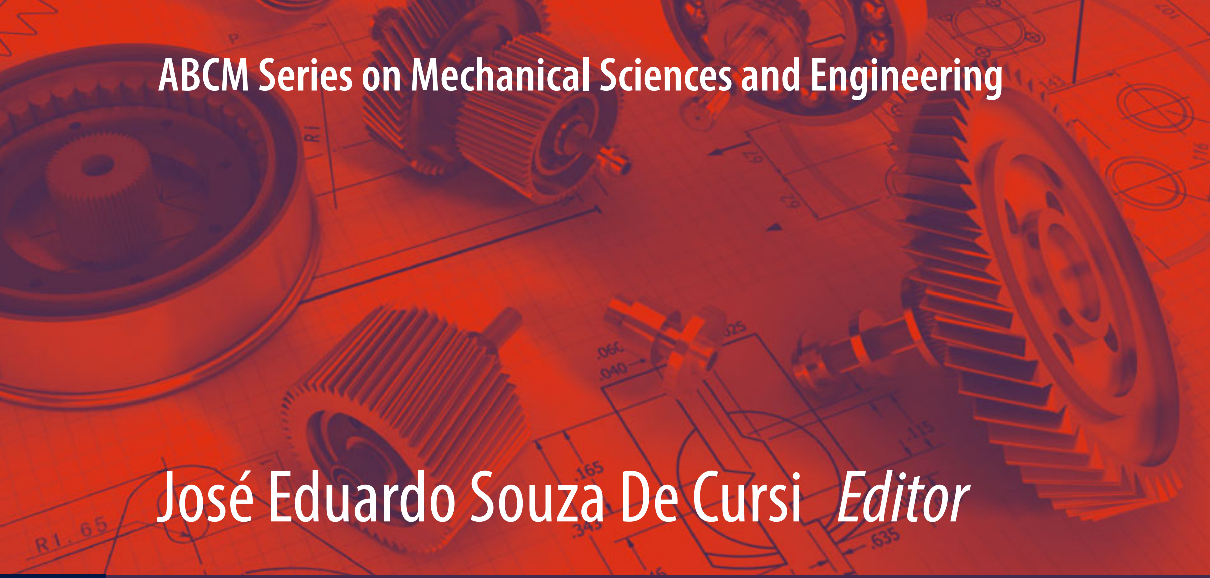


ABCM Series on Mechanical Sciences and Engineering

José Eduardo Souza De Cursi *Editor*

Proceedings of the 6th International Symposium on Uncertainty Quantification and Stochastic Modelling

Uncertainties 2023

ABCM
Associação Brasileira de
Engenharia e Ciências Mecânicas

Springer

Lecture Notes in Mechanical Engineering

ABCM Series on Mechanical Sciences and Engineering

This series publishes selected papers as well as full proceedings of events organized and/or promoted by the Brazilian Society for Mechanical Sciences and Engineering (ABCM) on an international level. These include the International Congress of Mechanical Engineering (COBEM) and the International Symposium on Dynamic Problems of Mechanics (DINAME), among others.

José Eduardo Souza De Cursi
Editor

Proceedings of the 6th International Symposium on Uncertainty Quantification and Stochastic Modelling

Uncertainties 2023

ABCM
Associação Brasileira de
Engenharia e Ciências Mecânicas

Springer

Editor
José Eduardo Souza De Cursi
Department Mechanics/Civil Engineering, INSA
Rouen Normandie
Saint-Etienne du Rouvray, France

ISSN 2195-4356 ISSN 2195-4364 (electronic)
Lecture Notes in Mechanical Engineering
ISSN 2662-3021 ISSN 2662-303X (electronic)
ABCM Series on Mechanical Sciences and Engineering
ISBN 978-3-031-47035-6 ISBN 978-3-031-47036-3 (eBook)
https://doi.org/10.1007/978-3-031-47036-3

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Paper in this product is recyclable.

Preface

This book assembles the full papers accepted by the 6th International Symposium on Uncertainty Quantification and Stochastic Modeling (Uncertainties 2023), jointly organized by INSA Rouen Normandy and the Universidade Federal do Ceara (UFC, Fortaleza, Brazil). The congress was held from July 30 to August 4, 2023. This congress was originally scheduled to be held in 2022, but the pandemic crisis due to COVID forced it to be postponed to 2023, under a 100% virtual form.

After the first meeting held in Maresias, Brazil, in 2012, Uncertainties became a biannual event in order to create a permanent forum for the discussion of academic, scientific, and technical aspects of uncertainty quantification in mechanical systems. Previous meetings were held in Rouen (2014), Maresias (2016), Florianópolis (2018), and Rouen (2020). The pandemic prevented the conference from being held in 2022, so the conference was postponed to 2023.

The main goal of Uncertainties is to provide a forum for discussion and presentation of recent developments concerning academic, scientific, and technical aspects of uncertainty quantification in engineering, including research, development, and practical applications, which are strongly encouraged.

Uncertainties 2023 is a sequel of the 5th International Symposium on Uncertainty Quantification and Stochastic Modeling (Uncertainties 2020), held in Rouen (France) from June 29th to July 3rd, 2020.

Uncertainties 2023 is organized on behalf of the Scientific Committee on Uncertainty in Mechanics (Mécanique et Incertain) of the AFM (French Society of Mechanical Sciences) Scientific Committee on Stochastic Modeling and Uncertainty Quantification of the ABCM (Brazilian Society of Mechanical Sciences) SBMAC (Brazilian Society of Applied Mathematics).

José Eduardo Souza De Cursi

Organization

Program Committee Chair

Souza de Cursi, Eduardo	INSA Rouen, Saint-Etienne du Rouvray, France

Program Committee Members

Arruda, Jose Roberto	UNICAMP, Brazil
Beck, Andre T.	USP Sao Carlos, Brazil
Beer, Michael	Leibniz University, Hannover, Germany
Besnier, Philippe	IET Rennes, France
Campos Velho, Haroldo	INPE, Brazil
Canavero, Flavio	Instituto Politecnico di Torino, Italy
Cavalcante, Charles Casimiro Cavalcante	UFC, Brazil
Cortes, Juan Carlos Cortes	Universitat Politecnica de Valencia, Spain
Ellaia, Rachid	EMI, Maroc
Ettore, Pierre	INP Grenoble, France
Fabro, Adriano Fabro	UNB, Brazil
Hachimi, Hanaa Hachimi	ENSA Kenitra, Maroc
Holdorf, Rafel	UFSC, Brazil
Iaccarino, Gianluca	Stanford University, USA
Kougioumtzoglou, Ioannis A.	Columbia University, USA
Lavor, Carlile	UNICAMP, Brazil
Guang Li	Purdue University, USA
Chu Liu	Nantong University, China
Lima, Antonio M. G.	UFU, Brazil
Lima, Roberta	PUC-Rio, Brazil
Maciel, Tarcisio Ferreira	UFC, Brazil
Moens, David	KU Leuven, Belgium
Nobile, Fabio	EPFL, Switzerland
Prieur, Clementine	University Grenoble Alpes, France
Mota, João César Moura Mota	UFC, Brazil
Piovan, Marcelo	UTN Bahia Blanca, Argentina
Rade, Domingos	ITA, Brazil
Ritto, Thiago	UFRJ, Brazil
Rosales, Marta	UNS, Argentina

Rochinha, Fernando A.	UFRJ, Brazil
Sampaio, Rubens	PUC-Rio, Brazil
Sudret, Bruno	ETH Zurich, Swiss
Trindade, Marcelo	USP Sao Carlos, Brazil
Villanueva, Rafael Jacinto	Universitat Politecnica de Valencia, Spain
Schoefs, Franck Schoefs	University of Nantes, France

Organizing Committee Chairs

Souza de Cursi, Eduardo	INSA Rouen, Saint-Etienne du Rouvray, France
Maciel, Tarcisio Ferreira	UFC, Brazil

Organizing Committee Members

Sousa, Diego Aguiar Sousa	IFCE, Brazil
Cavalcante, Charles Casimiro Cavalcante	UFC, Brazil
Mota, João César Moura Motta	UFC, Brazil
Nóbrega, Kleber Zuza	UFC, Brazil
Moreira, Nicolas de Araújo	UFC, Brazil
Carvalho Júnior, José Raimundo de Araújo	UFC, Brazil

Contents

Uncertainties of Numerical Simulation for Static Liquefaction Potential of Saturated Soils

W. H. Huang[1(✉)], Y. Shamas[1], K. H. Tran[2,3], S. Imanzadeh[1,2], S. Taibi[2], and E. Souza de Cursi[1]

[1] Normandie Univ., INSA Rouen Normandie, Laboratoire de Mécanique de Normandie,
76801 Saint-Etienne du Rouvray, France
wenhao.huang@insa-rouen.fr

[2] Normandie Université, UNIHAVRE, Laboratoire Ondes et Milieux Complexes,
CNRS UMR 6294, Le Havre, France

[3] Faculty of Civil Engineering and Environment,
Thai Nguyen University of Technology, Thai Nguyen City, Thai Nguyen, Vietnam

Abstract. The wind turbine foundations are subject to dynamic and static loadings. Some studies have shown that it is necessary to study the liquefaction of the soil under the foundations due to the impact of these loadings. However, in literature there are not many studies focusing on the comparison between the experimental results and numerical modeling of static liquefaction. In this study, NorSand model was used to simulate the static liquefaction behavior of soil. First, a series of experimental data were selected from the literature to study the experimental behavior leading to liquefaction of saturated Hostun sand RF. Then, these experimental data were compared to the results simulated by using NorSand model. This study focuses on the uncertainties of experimental results leading to the uncertainties of numerical modeling results. Furthermore, the numerical model integrates physical parameters related to the nature of the soil and also mathematical parameters. To modelize the static liquefaction of soil, some input parameters are needed to be determined based on the experimental tests. The corresponding uncertainties are evaluated to quantify the effect of the model parameters on the liquefaction potential soil of Hostun sand RF. An analysis of the uncertainties linked to the choice of these parameters makes it possible to reduce the difference between the experimental results and their simulation and as well the uncertainties.

Keywords: Static liquefaction · Saturated soil · NorSand model · Triaxial undrained tests · Hostun sand RF · Statistical errors · Uncertainties

1 Introduction

The concept of static liquefaction was first proposed by Castro et al., [1–4] 1975. However, it was not until the 1990s that static liquefaction gradually attracted attention [5]. The occurrence of flow landslide damage caused by static liquefaction is very sudden and extremely harmful to engineering, such as the landslide accident that occurred during the construction of Wachusett Dam in Massachusetts, USA, in 1907, which caused huge

J. E. S. De Cursi (Ed.): UQSM 2023, LNME, pp. 1–15, 2024.
https://doi.org/10.1007/978-3-031-47036-3_1

human injuries and property damages, thus drawing the attention of scholars worldwide. The main reason for liquefaction is that the effective stress and shear strength of saturated sand decrease significantly or even reach zero due to the increase of pore water pressure [6]. This phenomenon is called liquefaction. There are two reasons for liquefaction, one is the dynamic cycle of earthquake [7–9], the other is static load. The liquefaction caused by earthquake is called cyclic liquefaction, and the liquefaction caused by static load is called static liquefaction. The research of this paper mainly focuses on the simulation of the whole process of mechanics and deformation development of sand in the process of static liquefaction by using the numerical simulation method.

At present, the NorSand constitutive model (NS) [10] is familiarly used to simulate the static liquefaction phenomenon of saturated sandy soil under undrained conditions. The advantage of this model is that only few soil parameters are required, which can be estimated from routine laboratory or in-situ tests, and is able to explicitly capture the behavior of the entire soil, from static liquefaction of loose sands to swelling of dense sands. However, few researchers have studied the uncertainty generated in numerical simulation. In this paper, the selection of state parameters in the simulation process is systematically analyzed and compared, and the uncertainty of simulation results caused thereby is emphatically analyzed.

2 Material

The material used in this paper is a fine sandy soil (Hostun sand RF), where Fig. 1 shows the particle shape of this sandy soil under the microscope [11, 12], and Fig. 2 shows the particle distribution of this material and make comparison with limitation of liquefiable soil [13].

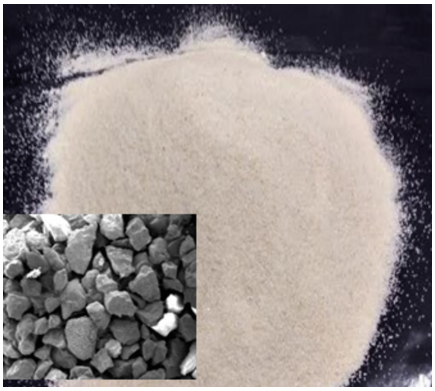


Fig. 1. Microscopic picture of Hostun sand RF [11, 12]

Table 1 shows the particle size composition and basic physical property index of Hostun sand RF [14, 15]. The specific gravity of Hostun sand RF is 2.65, the friction angle is about 40°, and the maximum and minimum void ratios are 1.041 and 0.648, respectively. Where D_{10}, D_{50} and D_{60} in Table 1 are the effective particle size, average particle size and limiting particle size of Hostun sand RF respectively.

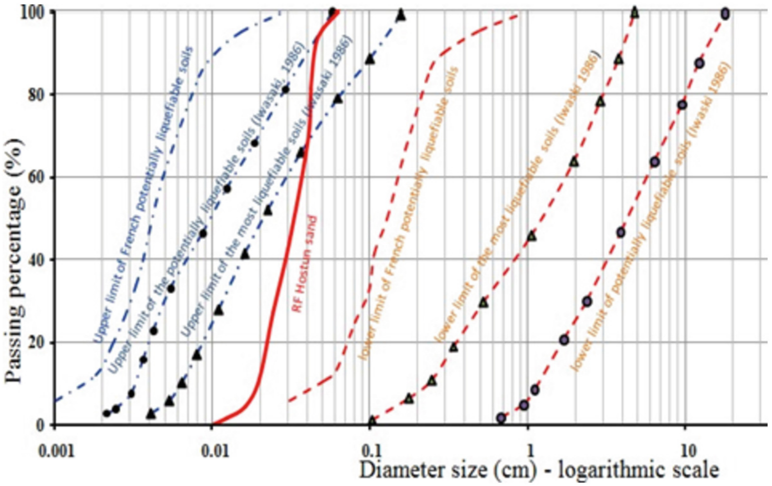


Fig. 2. Comparison between the particle distribution of Hostun sand RF and other liquefied sand [13]

Table 1. Particle size composition and basic physical property index of Hostun sand RF [14, 15]

Grain specific weight ρ (g/cm^3)	e_{max}	e_{min}	D_{10}(μm)	D_{50}(μm)	D_{60}(μm)	Friction angle φ (°)
2.65	1.041	0.648	200	300	400	40

3 Triaxial Experimental Tests

In order to verify the applicability of the NorSand constitutive model, the main parameters of the model must be obtained on the basis of laboratory tests. Therefore, it is necessary to know the instruments required for the test. This paper will refer to the apparatus used to carry out the static liquefaction tests [11].

3.1 Experimental Instruments

The static liquefaction test of sandy soils is carried out using a special type of autonomous triaxial unit whose originality comes from the loading pattern of the specimen, which is carried out by means of a piston located at the bottom of the lower support and which can be moved vertically in one direction or the other. The movement of the piston can be achieved by displacement or by controlled pressure, hydraulically ensured by a pressure chamber containing degassed water (Fig. 3).

3.2 Experimental Static Liquefaction of Hostun Sand RF

The size of the sample is a standard triaxial sample with a height of 140 mm and a diameter of 70 mm, and the sample's saturation is completely saturated, this test adopts

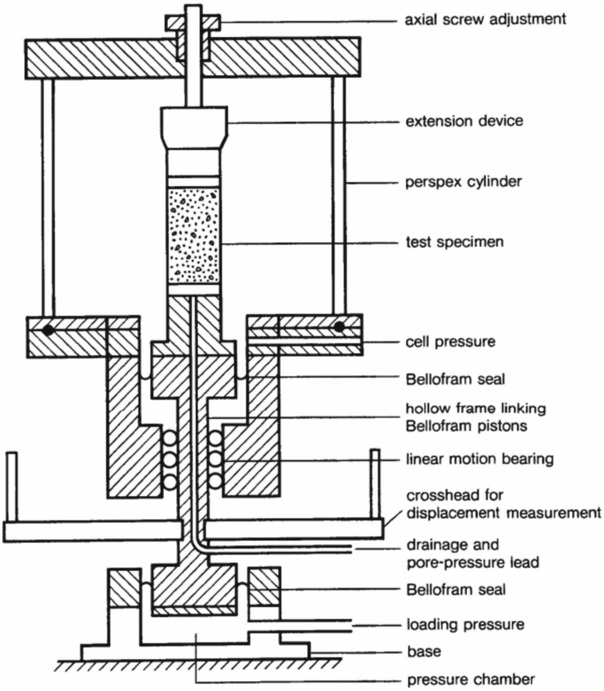


Fig. 3. Schematic diagram of triaxial device for static liquefaction test [11]

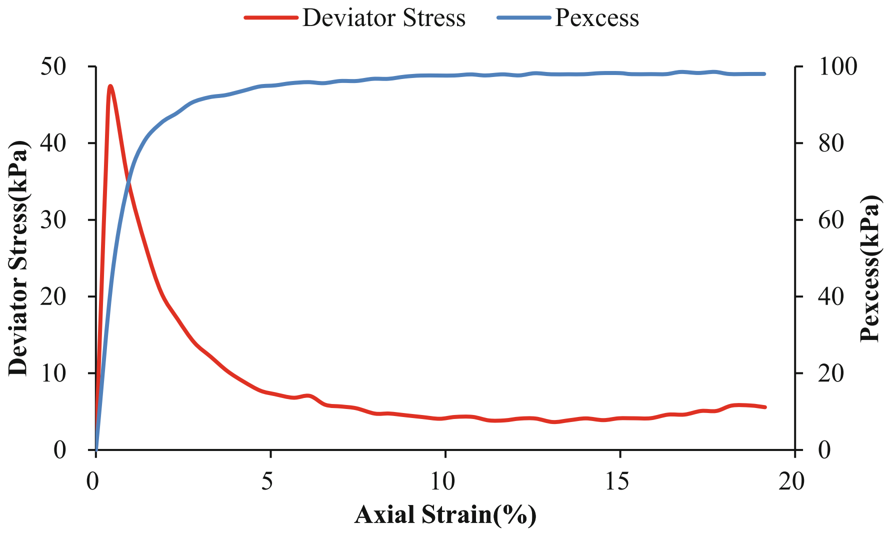


Fig. 4. Experimental relation curve between measured excess pore water pressure, deviatoric stress with axial strain of fully saturated specimen [11]

the undrained test, and the confining pressure is controlled at 100 kPa, then the axial stress is continuously increased until the sample is sheared to failure.

The experimental results are shown in Fig. 4. It can be seen from the figure that with the constant increase of axial strain, the deviatoric stress will continue to increase to the maximum value, and then rapidly decrease to close to 0, the reason for this phenomenon is that the increase of pore water pressure resulted in the decrease the effective stress of soil, which cause a significant decline in the shear strength of the sample. This characteristic of sand is called static liquefaction phenomenon.

4 Numerical Simulation

4.1 Constitutive Model of Soil—NorSand Model

The NorSand model was first proposed by Mike Jefferies [20] and is a constitutive model developed based on the critical state soil mechanics theory. NorSand is a generalized critical state model that can accurately capture the effect of soil porosity on soil behavior and, therefore, can simulate soil static liquefaction behavior very accurately within a certain range [16, 17]. In addition, the NorSand model has the advantage that few parameters are required and most of them can be obtained from within the laboratory.

The parameters of the NorSand model (Table 2) are independent of the void ratio and the confining pressure, which means that the values are kept constant for each type of soil. In general, the optimization of parameters is achieved by simulating soil behavior under different conditions through laboratory triaxial tests, this is very helpful for future numerical simulation work.

Table 2. Parameters of NorSand model [20]

Description	Parameters
Critical state locus	‘Altitude’ of CSL: $\varGamma$
	Slope of CSL: λ_e
Dilation limit	Material parameter: χ_{tc}
Strength parameter	Critical friction ratio: M_{tc}
	Material parameter: N
Plastic hardening	Hardening parameter: H_o
	Hardening parameter: H_ψ
Elasticity	Reference value of the shear modulus at the reference pressure: G_{ref}
	Exponent of the power-law elasticity: n_G
	Poisson’s ratio: ν
Over-consolidation ratio	R
Softening parameter	S
State parameter	ψ

4.2 Input Parameters Used in the Model

4.2.1 State Parameter

The state parameter, as shown in the Fig. 5, as a key controlling variable of the NorSand model, means the difference between the soil pore ratio and its critical pore ratio at a current pressure.

$$\psi = e - e_c \tag{1}$$

In Eq. 1, ψ is the state parameter of the soil, e is the current pore ratio of the soil, and e_c is the void ratio of the critical state at the current mean stress. On the other hand, it can be seen that the state parameter contains two influencing parameters, density and pressure. When ψ >0, the pore ratio of the current soil is greater than the critical pore ratio in the same state, and the soil is in a loose state at this time, and when ψ <0, the pore ratio of the current soil is less than the critical pore ratio in the same state, and the soil is in a dense state at this time. In other words, the larger ψ is, the looser the soil is, and the smaller ψ is, the denser the soil is.

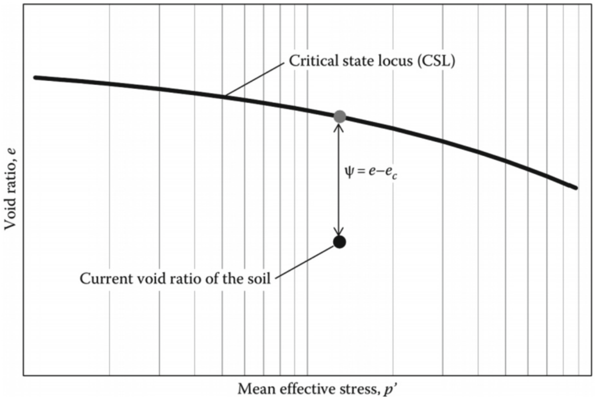

Critical state locus (CSL)
$\psi = e - e_c$
Current void ratio of the soil
Void ratio, e
Mean effective stress, p'

Fig. 5. Definition of state parameter ψ [10]

The CSL in Fig. 5 is determined by the fitting line CSL in Fig. 6 from the experimental tests. This figure presents the relationship between void ratio at critical state and the effective mean stress p'. It can be seen that there is some variability between the fitting line and the experimental points of void ratio at critical state for different value of effective mean stress.

The void ratio of normally consolidated soil remaining constant, with the increase in axial stress, the pore water pressure will increase accordingly, resulting in a decrease in effective stress until the specimen fails, at which time the soil reaches a critical state under this void ratio, and then a series of critical state points are obtained by testing the specimens under different void ratios.

From Figs. 6 and 7, a fitted straight line can be obtained, and this line is the critical state line (CSL) of Hostun RF sand with the absolute value of slope 0.03 and intercept

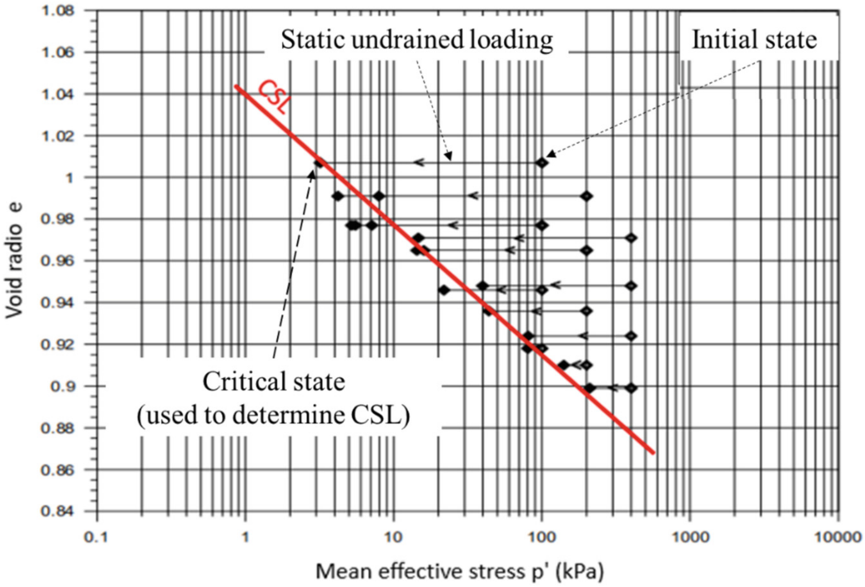


Fig. 6. Representation of the experimental data to determine the void ratio at critical state for different values of mean effective stress for Hostun sand RF [11]

1.035. Therefore, the slope λ_e and 'Altitude' Γ of the critical state line (CSL) can be obtained as shown in Fig. 7.

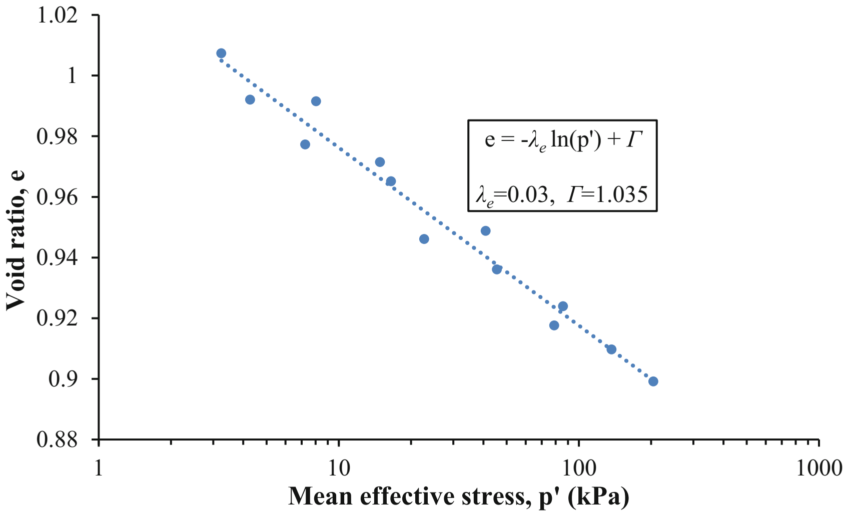


Fig. 7. Numerical fitting of critical state line (CSL)

4.2.2 Others Parameters

The various parameters of Hostun RF sand need to be determined before conducting numerical simulations, as shown in Table 3, $G_{\mathrm{ref}}/P_{\mathrm{ref}}$ represents Shear rigidity, where G_{ref} is the scaling constant for elastic shear modulus and P_{ref} is the reference mean pressure (typically the common value of 100 kPa is used); v denotes the elastic Poisson's ratio; Γ and λ_e are the intercept and slope of the critical state line, respectively. M_{tc} refers to the critical friction ratio at triaxial compression; χ_{tc} represents the dilatancy rate parameter; N is the volumetric coupling parameter in stress dilatancy; H_0 denotes the plastic hardening modulus. The Table 3 also shows the values of each parameter for Hostun RF sand under the application of NorSand model, in comparison to the typical range of values suggested by Mike et al. [10]. Some of them are taken from the state parameter presented in Sect. 4.2.1 and some of them are taken from the preceded adjustment modelization.

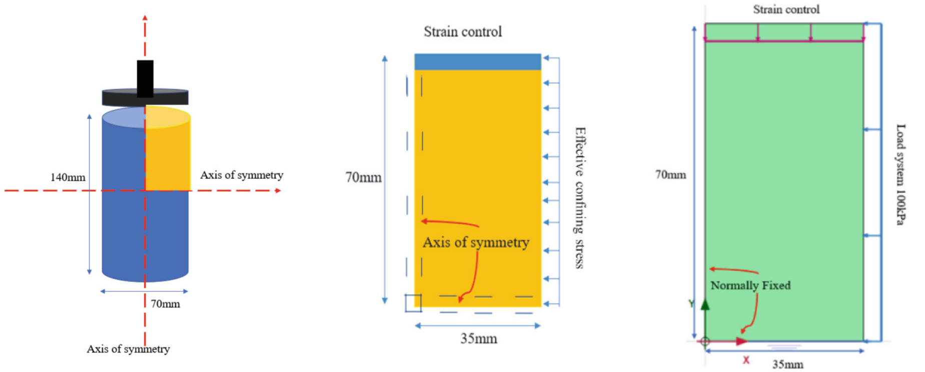
Table 3. Parameter values for Hostun RF sand and typical ranges for each parameter

Parameters	$G_{\mathrm{ref}}/P_{\mathrm{ref}}$	v	Γ	λ_e	M_{tc}	χ_{tc}	N	H_0
Hostun RF sand	150	0.2	1	0.03	1.5	4	0.35	100
Typical range	100–160	0.1–0.3	0.9–1.4	0.01–0.07	1.2–1.5	2–5	0.2–0.5	25–500

4.3 Modelization of Triaxial Test

To validate the model, finite element software was used to simulate the triaxial consolidated undrained test, and the simulated results were compared with the experimental results.

As shown in Fig. 8 (a), the dimensions of the specimen are 140 mm in height and 70 mm in diameter. As using axisymmetric coordinate, only one quarter of the specimen is used for simulation (Fig. 8 a and b) [18, 19]. The method of strain control was used to shear the sample, the calculation stops after reaching the specified axial strain (20%). The effective confining stress is 100 kPa. The other two boundaries are normally fixed. It's means that the points on these two boundaries can only move along the axis (Fig. 8 b and c).



a) Concept diagram of triaxial test

b) Axisymmetric coordinate

c) Numerical simulation

Fig. 8. Simulation of undrained triaxial test

4.4 Results

Figure 9 shows the liquefaction area in the soil when the axial strain reaches 20%. The purple inverted triangle is the liquefaction point, where the effective stress is almost equal to 0. It can be seen that liquefaction has occurred in the sample after shearing the modeling sample. The liquefaction points on the top of the sample are denser than the bottom one due to the mesh of finite element method.

In order to better understand the change of stress inside the specimen after shear failure, any one liquefaction point is taken as the calculation point to obtain the relationship curve of stress with axial strain as shown in Fig. 10, it can be seen that as the axial strain increases, the deviator stress increases to a maximum value at a small strain ($<1\%$), and then decreases rapidly until it approaches 0, at the same time, the pore water pressure inside the specimen increases with the increase of axial strain, and then remains constant at about 100 kPa, which is equal to the confining stress of the specimen, which means that the effective stress of the specimen is zero at this time, thus leading to liquefaction.

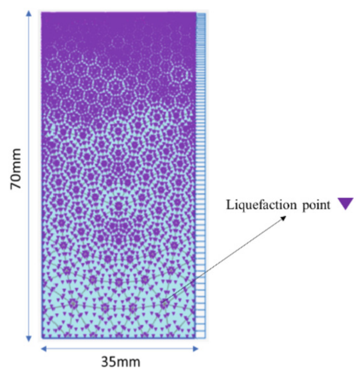


Fig. 9. Liquefaction appearance inside the sample

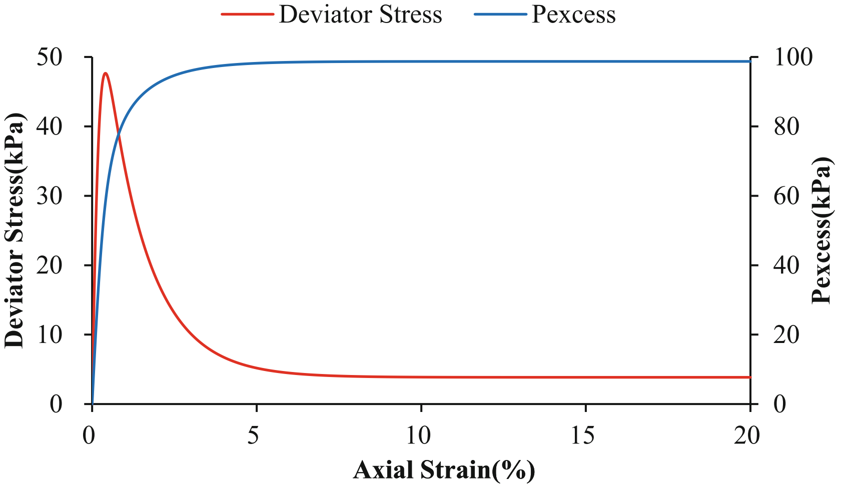


Fig. 10. Numerical relation Curve between excess pore water pressure, deviatoric stress with axial strain of fully saturated specimen

5 Discussions

5.1 Error Analysis

The static liquefaction potential is an important parameter for measuring the resistance of soil to liquefaction and is closely related to the porosity (or void ratio) of the soil. Generally, the denser the soil, the smaller the porosity, and the stronger the resistance to liquefaction (as in Case 3 in Fig. 11), resulting in a smaller static liquefaction potential. Conversely, when the soil is looser, the porosity is larger, and the resistance to liquefaction is weaker (as in Case 2 in Fig. 11), resulting in a larger static liquefaction potential. To determine these two boundary conditions, which is considered as the extremities and the critical cases, it is necessary to calculate the maximum and minimum value.

In order to verify the accuracy of the fitted CSL, an error analysis was performed on its fitted straight line, and the error between the void ratio of each experimental specimen and the void ratio on the fitted straight line under the same stress was calculated, as shown in Eq. 2,

$$\text{Error} = e_e - e_f \tag{2}$$

e_e is the void ratio obtained from the specimen experiment, e_f is the void ratio calculated from the fitted straight line, and its maximum positive error and maximum negative error are shown in Table 4.

It can be seen from Table 4 that when the void ratio of the sample is 0.991, the positive error with the fitted straight line (CSL) is the largest; when the void ratio of the sample is 0.946, the negative error with the fitted straight line (CSL) is the largest; while keeping the slope of the fitted straight line λ_e unchanged, the maximum positive error straight line and the minimum negative error straight line are respectively fitted, as shown in Fig. 11. In order to separate these lines, three cases are defined: case 1 correspond

Table 4. Maximum positive error and minimum negative error

Types	Experimental data [11]	Fitting	Error
Void ratio	0.991	0.983	+0.008
	0.946	0.957	−0.011

the fitting line, case 2 correspond the maximum positive error, case 3 correspond the minimum negative error.

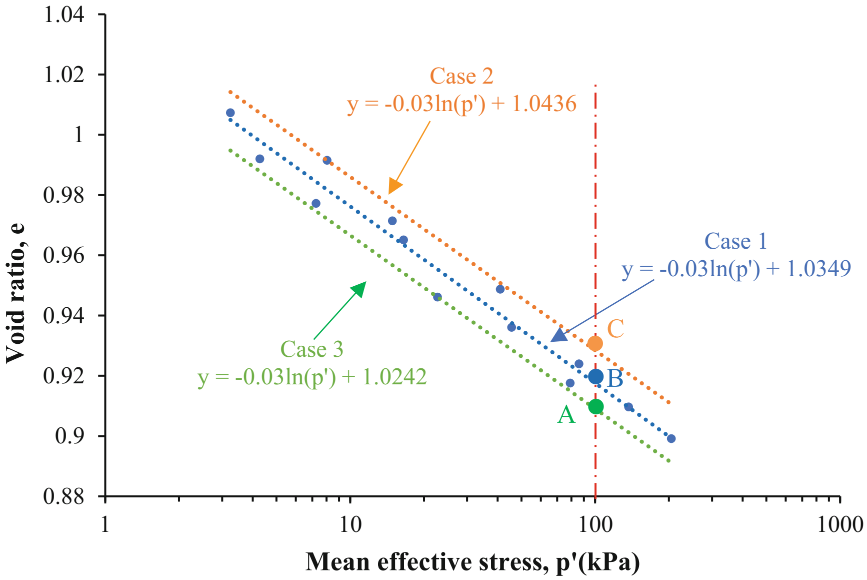


Fig. 11. Error comparison of fitting Critical State Line

5.2 Discussion

From three fitting straight lines in Fig. 11, the void ratio at the critical state with mean effective stress of 100 kPa was deducted (Points A, B, C in Fig. 11). Table 5 shows the value of state parameter ψ calculated based on these deducted void ratios following Eq. 1. When the soil's void ratio change also results in the change of altitude of CSL (Γ) which is presenting in Fig. 11 and Table 5.

From these values of state parameter and altitude of CSL, three models corresponding to three values of void ratio at critical state were carried out to evaluate the effect of void ratio on the liquefaction behavior. This calculation is necessary when the soil in situ is always not uniform, the void ratio of the soil often distributes around a mean value. It can be seen from Fig. 12 that when the strain is about 0.5%, the deviatoric stress of the

Table 5. State parameters obtained from different fitting straight lines under the confining pressure of 100 kPa

Types	Void ratio (100 kPa)	State parameter ψ	Altitude of CSL Γ
Modelized soil (Current void ratio in Eq. 1)	$e = 1.007$ [21, 22]	-	-
Case 1	Point B: $e_c = 0.897$	0.110	1.035
Case 2	Point C: $e_c = 0.905$	0.102	1.044
Case 3	Point A: $e_c = 0.886$	0.121	1.024

sample reaches the maximum. Compare the maximum values of the three simulation results with the experimental results, and calculate the error according to Eq. 3,

$$Error = \frac{q_{Me} - q_{Mm}}{q_{Me}} \times 100\% \tag{3}$$

where q_{Me} is the maximum of the deviatoric stress obtained from the experiment, q_{Mm} is the maximum of the deviatoric stress obtained from the simulation, the calculation results are shown in Table 6.

Table 6. Error between the maximum deviatoric stress of simulation results and experimental results

Types	Experimental data (q_{Me})	Case 1 (q_{Mm})	Case 2 (q_{Mm})	Case 3 (q_{Mm})
Deviatoric stress (kPa)	47.4	47.6	48.9	45.9
Error (%)	-	0.42	−3.05	3.43

It can be seen that the maximum deviatoric stress obtained by case 1 has the smallest error between the considered cases compared with the experimental results. Moreover, its decreasing trend after reaching the maximum deviatoric stress; and its final stable state after reaching an axial strain of 10%, graph of case 1 remain almost the same as that of the experimental data. This shows that the case 1 results are very close to the experimental results, and also shows that the NorSand model can well simulate the static liquefaction of soil.

Similarly, Fig. 13 shows the relationship curve between the excess pore water pressure and the axial strain. It can be seen that although the state parameters are different, the differences between the three simulation results are very small, and the overall trend is consistent with the experimental results. With the increase of the axial strain, the

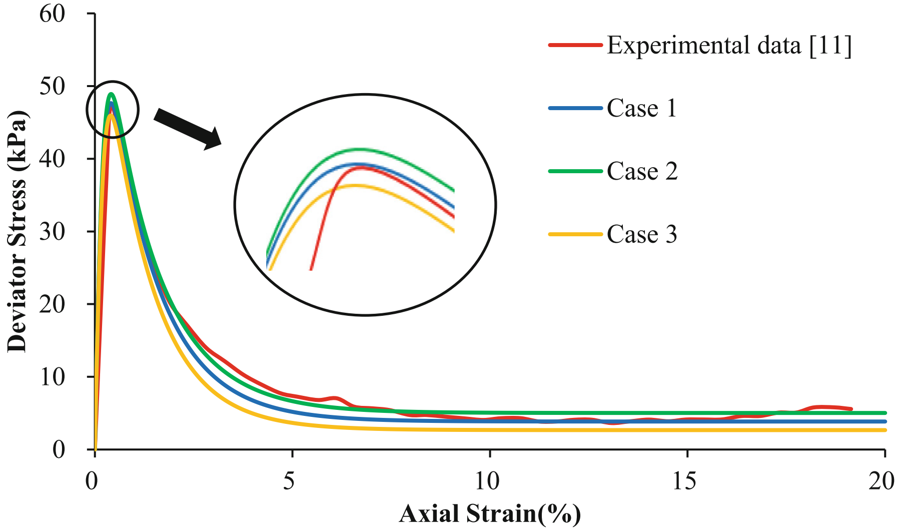


Fig. 12. Deviatoric stress vs axial strain of three simulated cases and experimental data

excess pore water pressure also increases continuously until it remains stable around 100 kPa, which indicates that with the increase of the axial strain, the effective stress of soil continuously decreases and finally approaches 0.

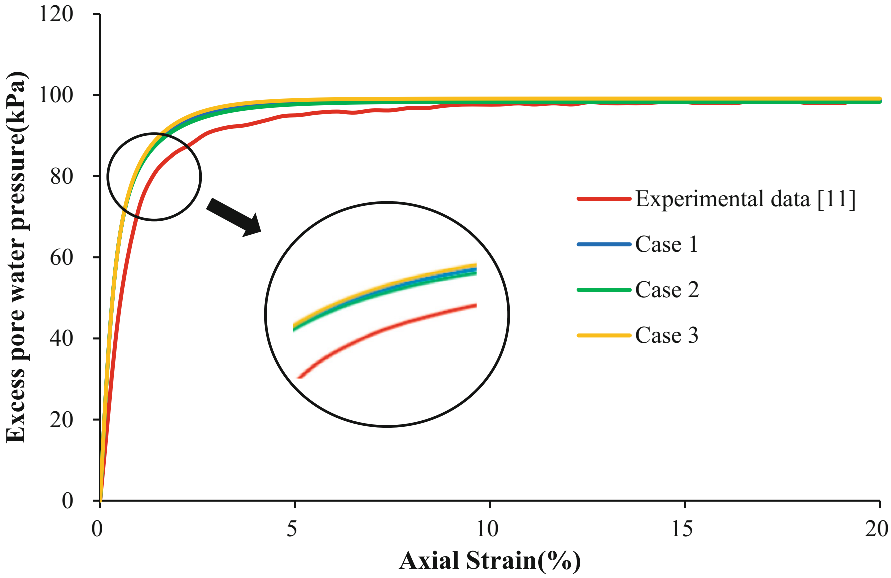


Fig.13. Excess pore water pressure vs axial strain of three simulated cases and experimental data

6 Conclusion and Perspective

In this research paper, it can be seen that the results of NorSand model have a good agreement with the experimental data. When the axial strain is small (about 0.4–0.5%), the deviatoric stress of soil reaches the maximum, after reaching the peak, the deviatoric stress starts to drop rapidly (close to 0), and then remains stable.

The uncertainties of the void ratio measurement at the critical state affect the modeling results. The maximum value of the measured void ratio at the critical state results a 3.05% difference of deviatoric stress compare to the experimental results for more conservative model (Case 2), while this difference is 3.43% corresponding to the minimum value of the measured void ratio at the critical state (less conservative model, Case 3). The 0.019 (2.14%) of void ratio increase results in 3 kPa (6.54%) of decrease in deviatoric stress. It means that there is a band for the peak of deviatoric stress to fluctuate depending on the distribution of measured void ratio at critical state. This result also suggests perspective that it is necessary to build the distribution law of liquefaction potential, or the peak of deviatoric stress, based on the distribution law of void ratio at critical state.

References

1. Castro, G.: Liquefaction and cyclic deformation of sands. J. Geotech. Eng.-ASCE **101**, 551–569 (1975)
2. Castro, G., Poulos, S.J.: Factors affecting liquefaction and cyclic mobility. J. Geotech. Eng. Div. **103**(6), 501–516 (1977)
3. Casagrande, A.: Liquefaction and Cyclic Deformation of Sands-A Critical Review. Harvard Soil Mechanics Series, no. 88. Harvard University, Cambridge (1976)
4. Kramer, S.L., Seed, H.B.: Initiation of soil liquefaction under static loading conditions. J. Geotech. Eng. **114**(4), 412–430 (1988)
5. Lu, X.B., Tan, Q.M., Wang, S.Y.: The advances of liquefaction research on saturated soils. Adv. Mech. **34**(1), 87–96 (2004)
6. Craig, R.F.: Craig's Soil Mechanics. CRC Press, Boca Raton (2004)
7. Castro, G.: Liquefaction of Sands. Harvard Soil Mechanics Series, no. 81. Harvard University, Cambridge (1969)
8. Seed, H.B.: Ground motions and soil liquefaction during earthquakes. Earthquake Engineering Research Institute (1982)
9. Tran, K.H., Imanzadeh, S., Taibi, S.: Some aspects of the cyclic behavior of quasi-saturated sand. Acad. J. Civil Eng. **36**(1), 142–145 (2018)
10. Jefferies, M., Ken, B.: Soil Liquefaction: A Critical State Approach. CRC Press, Boca Raton (2015)
11. Benahmed, N.: Comportement mécanique d'un sable sous cisaillement monotone et cyclique: application aux phénomènes de liquéfaction et mobilité cyclique. Diss. Marne-la-vallée, ENPC (2001)
12. Tran, K.H., Imanzadeh, S., Taibi, S.: Liquefaction Behavior of Dense Sand Relating to the Degree of Saturation. In: Duc Long, P., Dung, N. (eds.) Geotechnics for Sustainable Infrastructure Development, pp. 879–886. Springer, Singapore (2020). https://doi.org/10.1007/978-981-15-2184-3_114
13. Iwasaki, T.: Soil liquefaction studies in Japan: state-of-the-art. Soil Dyn. Earthq. Eng. **5**(1), 2–68 (1986)

14. Fargeix, D.: Conception et réalisation d'une presse triaxiale dynamique: application à la mesure des propriétés des sols sous sollicitations sismiques. Diss. ANRT, Université Pierre Mendes France (Grenoble II) (1986)
15. Tran, K.H., Imanzadeh, S., Taibi, S.: Liquefaction of unsaturated soils-volume change and residual shear strength. Eur. J. Environ. Civil Eng. 1–21 (2022)
16. Sternik, K.: Prediction of static liquefaction by Nor Sand constitutive model. Studia Geotechnica et Mechanica **36**(3) (2014)
17. Woudstra, L.J.: Verification, validation and application of the NorSand constitutive model in PLAXIS: single-stress point analyses of experimental lab test data and finite element analyses of a submerged landslide (2021)
18. Surarak, C., Likitlersuang, S., Wanatowski, D.: Stiffness and strength parameters for hardening soil model of soft and stiff Bangkok clays. Soils Found. **52**(4), 682–697 (2012)
19. Galavi, V.: Groundwater flow, fully coupled flow deformation and undrained analyses in PLAXIS 2D and 3D. Plaxis Report (2010)
20. Jefferies, M.G.: Nor-Sand: a simle critical state model for sand. Géotechnique **43**(1), 91–103 (1993)
21. Tran, K.H., Imanzadeh, S., Taibi, S.: Investigation of the influence of saturation degree on the cyclic behaviour of fine clear sand. Acad. J. Civil Eng. **40**(1), 66–70 (2022)
22. Tran, K.H., Imanzadeh, S., Taibi, S.: Effect of saturation degree on the behaviour of clear sand in very dense state. Acad. J. Civil Eng. **40**(1), 71–75 (2022)

Uncertainties About the Toughness Property of Raw Earth Construction Materials

Youssef Shamas[1,2](✉), H. C. Nithin[3], Vivek Sharma[3], S. D. Jeevan[3], Sachin Patil[3], Saber Imanzadeh[1,2], Armelle Jarno[2], and Said Taibi[2]

[1] Normandie Univ., INSA Rouen Normandie, Laboratoire de Mécanique de Normandie, 76801 Saint-Etienne du Rouvray, France
youssef.shamas@insa-rouen.fr

[2] Normandie Université, UNIHAVRE, Laboratoire Ondes Et Milieux Complexes, CNRS UMR 6294, Le Havre, France

[3] New Horizon College, Marathalli. Ring Road, Near Marathalli, Bangalore 560 103, India

Abstract. Silt-based construction material is an ecological and economical alternative for typical cement-based concrete and has received lately the researchers' attention more than before. Some researches were done on the raw earth material to enhance its characteristics as strength and ductility for being widely used for various materials. Yet, many other mechanical properties can be used to study the mechanical properties of raw earth materials such as strain modulus and toughness. Studies concerning the toughness of a material were rarely considered previously except for metals despite its significant role associated to the energy absorbed by the material under loading before fracturing. The purpose of this paper is to restate the normal toughness definition used in the literature for typical construction materials and presents the possibilities of the repetitions of our experimental tests showing the statistical error occurred between same tests performed comparing the stress-strain graphs for three replicates done for each formulation out of 25. This paper will focus on the uncertainties and the possibility to neglect the intruding samples to reach better results and better simulate and fit the experimental data in numerical analysis. Experimental tests has some statistical errors and the uncertainties must be minimal compared to the complications of the experiment.

Keywords: Silt-based material · Raw earth concrete · Stress-Strain curve · Area · Energy · Toughness · Statistical errors · Uncertainties

1 Introduction

Statement of uncertainty is required to develop a better understanding of the results obtained and analyze the experimental data. This is done by studying the different influencing instruments on the replication of the test by defining the different aspects of the test procedure having the greatest effects on the results so it can be controlled more closely.

Since ever, raw earth materials, consisting mainly of a compacted mixture of moist clay and sand, has been used as a building material. Pollution is well spread in whole

J. E. S. De Cursi (Ed.): UQSM 2023, LNME, pp. 16–29, 2024.
https://doi.org/10.1007/978-3-031-47036-3_2

world, and nowadays all countries are searching for new and green energies or even trying to limit the use of the grey energy.

The modern construction materials using concrete and steel are highly energy-intensive in terms of grey energy. Hence a new ecological and economical construction material is required, known as eco-geomaterials for construction, mainly raw earth-based materials with low energy consumption compared to modern materials [1, 2].

Raw earth material are natural materials available everywhere in the world and in great amounts with a low and affordable cost. They Require only 1% of energy through the production process compared to that of concrete [3]. They are recyclable and can be reversible, reducing or avoiding notable quantity of waste. Hence it needs to be studied more and developed to be a great alternative of the cement-based concrete.

Numerous studies have been established to improve the raw earth materials' properties regarding its mechanical strength, the shrinkage and swelling, the cracking and the hygro-thermal properties [4]. Stabilizers as lime, cement and/or gypsum and reinforcement as strong fibers are used to improve these properties to meet the expected performance required from a building material [5].

To improve any property of the raw earth material, several components (natural and/or manmade materials) can be added as a constituent of the mixture. Depending on the raw earth construction technique followed, adding small quantity of binders such as lime and/or cement, may enhance the material's compressive strength up to a certain level to become acceptable as a construction and building material [4, 6, 7].

Different studies on the raw earth material's ultimate compressive strength [8] and its ductility [9] with the influence of the substituents used in the mixture on these properties were previously done by Imanzadeh et al.

As much as toughness characteristic of materials is important, many studies were performed on fiber reinforced hydraulic lime mortar [10, 11] and fly ash concrete [12], but studies for toughness for raw earth concrete is rarely conducted.

Based on the Absolute toughness of the material defined from the literature as the total area under the stress-strain curve which is Total energy absorbed by the material before it fractures [13]. Different softwares are used to estimate the area under the stress-strain curves and verify the precision of these results.

In our study, we have 25 different formulations by changing the percentages by mass of the constituents; and for each formulation we have 3 replicates. Due to some uncertainties in the experiments, these 3 replicates are not exactly the same and there is small difference between the results of each from the other.

In this paper, the uncertainty from the experiments is reduced by checking and removing any intruding replica to reach better statistical data as a new approach to minimize the uncertainties.

2 Materials

The raw earth concrete is made from mixing different materials made of soil, binders (cement and lime), flax fibers and water.

2.1 Soil Material

Natural silt is chosen as a building material for being abundantly available locally.

Based on the grading size distribution curve, the Atterberg limits and according to LPC-USCS (ASTM D2487-11) standard [14], this soil is categorized as silty sand (SM). Where the detailed description of the properties of this silty sand was done by Imanzadeh et al. [8].

2.2 Binders

Cement and Lime acting as binders are used in the mixture of the raw earth concrete. According to the European standard EN459-1, the lime added is obtained from the Proviacal® DD range. It is a calcic quicklime CL 90-Q (R5, P3); it contains 90.9% available CaO and has a reactivity of $t_{60} = 3.3$ min [15]. According to the NF EN197-1[16] and NF P15-318 [17] standards, CEM I 52.5 N cement is used in the mixture of this material. Where a detailed properties of these binders is acknowledged by Eid 2017 [18].

2.3 Flax Fibers

Locally extracted flax fibers from the Normandy's region is used in the mixture of this raw earth concrete. Since the Normandy region is responsible for 55% of total production of flax in France, these flax fibers are chosen to be added to the mixture [19]. A proper description of these flax fibers was explicated by Imanzadeh et al. [9]. Fiber content in the mixture is varied in the range of 0.3%–0.45% in mass: 0.3% was considered as a low level and 0.45% as a high level of fiber content in specimens.

2.4 Incorporation of a Superplasticizer Additive

Limited water content is required in the mixture of raw earth concrete to minimize its shrinkage. Superplasticizer additives can be adjoined to the mixture acting as an alternative to preserve the consistency during the manufacturing process of the concrete. SIKA VISCOCRETE TEMPO-10, is a new generation superplasticizer based on acrylic copolymer, according to NF EN 934-2 standard, is added as an alternative [20]. This superplasticizer can be used without constraints for the construction of reinforced and pre-stressed concrete structures; because it doesn't contain chlorides or any other harming substances that might cause or promote the corrosion of the steel. The super-plasticizer contributes to deflocculating fine grains and to lubrification of the solid surfaces, decreasing the friction stresses between particles [21]. A constant amount of additive of 5 ml/m^3 has been used for each sample preparation.

The raw earth concrete samples are prepared using a potable tap water from the pipe in the laboratory.

2.5 Mixtures

Laboratory mixer of 4L capacity is used for the mixing procedure in order to obtain homogeneous samples with random distribution of the materials. Two successive phases are considered: First one for the dry mixing for all the dry materials; The second phase is the wet mixing by adding the water and additives. Additional details are mentioned by Imanzadeh et al. [9].

Limitation Considered

For the mixture of raw earth material, three constraints are considered respecting the various constraints from economical, ecological and environmental constraints.

Fundamental Constraint:

As the weight of the ingredients should sum up to 100% in weight for all the mixtures. Obtaining the first constraint:

Fiber % + Lime % + Cement% + Water% + Silt% = 100%

Economical and ecological constraints:

The main purpose of using raw earth concrete is to limit the use of unnecessary grey energy providing the best possible mechanical properties. Hence to limit the grey energy, the percentage of binders used (lime and cement) should be limited in the mixtures. Thus, Cement percentage in the mixture is limited to be maximum 16% from the sample's weight, and Lime limited to 12% maximum. Moreover, both constituents together, must not exceed the 16% of the total weight of the specimen.

Finally, the workability constraint:

Suitable workability of the mixture is insured for a better mechanical property [22] by imposing the fluidity conditions using the concrete slump test fulfilling S3 consistency condition according to the standard NF EN 206-1 [23]. Resulting with the following condition: $2.5 <- 9$ Silt – 22 lime – 9 cement + 42 water < 3.

Mixing Range

Considering the various constraints listed above, the mixing range of the constituents is defined as follows in the Table 1.

Table 1. Mixing ranges of constituents.

xi	Lower Limit (%)	Upper Limit (%)
x1: Fiber	0.3	0.45
x2: Lime	0	12
x3: Cement	4	16
x4: Water	20	25
x5: Silt	47	75

Formulations

Considering the above mixing ranges of the constituents, 25 different formulations are considered varying all constituents at the same time respecting all the constraints defined before. Table 2 is observed for each formulation with each different percentages by mass of constituents.

Table 2. List of the formulations and their constituents.

Formulation	Fiber (x1)	Lime (x2)	Cement (x3)	Water (x4)	Silt (x5)
1	0.0030	0.0000	0.1600	0.2239	0.6131
2	0.0030	0.0647	0.0400	0.2500	0.6423
3	0.0030	0.0000	0.0400	0.2290	0.7280
4	0.0045	0.0000	0.0400	0.2287	0.7268
5	0.0030	0.0000	0.0400	0.2345	0.7225
6	0.0045	0.0000	0.0400	0.2342	0.7213
7	0.0030	0.0000	0.1600	0.2295	0.6075
8	0.0045	0.0000	0.1600	0.2292	0.6063
9	0.0045	0.0719	0.0881	0.2500	0.5855
10	0.0030	0.0560	0.1040	0.2500	0.5870
11	0.0045	0.0568	0.1032	0.2500	0.5855
12	0.0045	0.0000	0.1200	0.2253	0.6502
13	0.0045	0.0000	0.1600	0.2255	0.6100
14	0.0040	0.0000	0.1600	0.2237	0.6123
15	0.0030	0.0237	0.1363	0.2326	0.6044
16	0.0045	0.0543	0.0400	0.2500	0.6512
17	0.0045	0.0438	0.0400	0.2429	0.6688
18	0.0045	0.0479	0.1121	0.2412	0.5943
19	0.0035	0.0480	0.0400	0.2500	0.6585
20	0.0035	0.0713	0.0887	0.2500	0.5865
21	0.0038	0.0000	0.1000	0.2291	0.6672
22	0.0030	0.0299	0.0841	0.2396	0.6433
23	0.0038	0.0261	0.0859	0.2409	0.6433
24	0.0038	0.0320	0.1280	0.2383	0.5980
25	0.0038	0.0301	0.0840	0.2395	0.6426

3 Experimental Method

3.1 Sample Preparation

First, the used silt in the mixture is oven-dried at 60 °C for 48 h in order to control the amount of water in the specimens.

Then, the homogeneous samples are prepared carefully in the laboratory to ensure a random distribution of the fibers. Laboratory mixer of a 4L capacity is used for the mixing procedure involving two successive phases: First phase, mixing the dry components (silt, fibers and binders) for two minutes; Second phase, wet mixing for three minutes by adding the solvents (water and additives).

Afterwards, the molds of 100 mm height and 50 mm diameter are filled associated with vibration for two minutes via a vibrating table. To ensure the achievement of the majority of the chemical reactions due to binders, a 90 days of curing time is assured by storing the obtained specimens time in controlled laboratory environment (relative humidity RH≈50% and temperature T≈22 °C). In total of 75 specimens are produced, including 3 trials for each of the 25 formulations with different mix proportions.

3.2 Unconfined Compression Strength Test

After 90 days of curing-time, an axial Unconfined Compressive Strength (UCS) test (Fig. 1) according to NF P94-420 [24], NF P94-425 [25] French standards is performed on the cured specimens with different mix proportion.

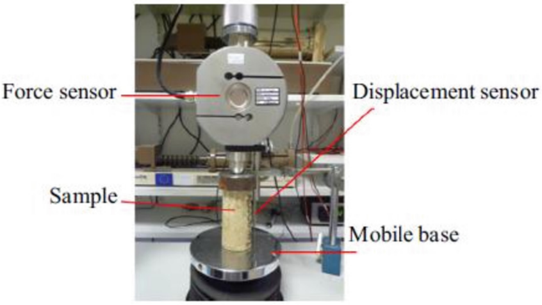


Fig. 1. Experimental Unconfined Compressive Test.

Hence, the toughness parameters are deduced from the applied axial stress versus axial strain curve obtained from UCS test.

The UCS experimental work is done in the laboratory using an apparatus of 100 kN maximum load capacity and ±0.05 mm accuracy potentiometric displacement sensor connected to a computer-controlled acquisition center. The tests are controlled under a strain rate of 0.1 mm/min without any confining stress on the sample. Additional information concerning the experimental device was stated by Imanzadeh et al. [8]. The force exerted on each specimen was recorded with its corresponding displacement from the sensor. Where the axial stress is calculated by dividing the force with the specimen's cross sectional are and its corresponding strain is calculated by dividing the displacement recorded by the initial height of the specimen.

3.3 Stress-Strain Curves

Using the obtained data from the experiment, the stress-strain curve is drawn for each specimen to study the behavior of the material and the effect of each component on it.

Imanzadeh et al. [9] have already validated that the threshold of 15% of strength loss after reaching the peak is enough to analyze the plastic behavior of the material. Hence the study of toughness is done according to this threshold. Moreover Formulations 2,4,5,15 and 19 were excluded since these specimens show gradual increase in strength without reaching maximum compressive strength.

Most of the specimen gave similar path for stress-strain curve containing three regions:

1. Region A containing both the toe region and the linear part of the curve.
2. Region A' including a non-linear path of the pre-peak region on the curve, after region A and before reaching the peak described as a plastic zone, usually associated to the non-linear phase due to micro-cracking.
3. Finally, region B involving the non-linear path after the peak until reaching the 85% stress of the ultimate compressive stress achieved. B is the area under the curve between Ultimate Compressive Strength (UCS) stress value and its 85% value defining part of the post-peak region as indicated in Fig. 2.

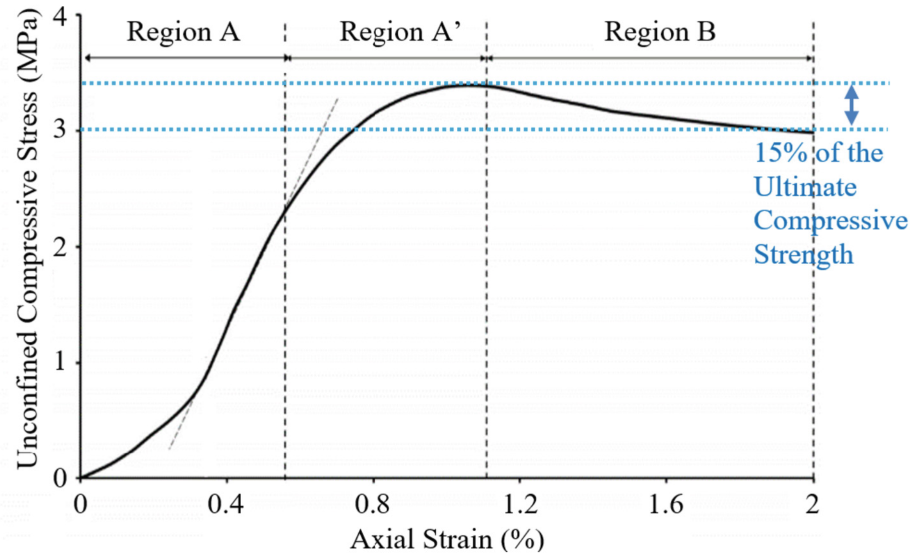


Fig. 2. Raw Earth Material's Stress-Strain Curve with Regions defined

4 Absolute Toughness

Raw earth material got the world's attention lately for its high suitability to replace normal concrete in construction for better environments. Various studies were previously achieved on raw earth material by Imanzadeh et al. [8, 9] by performing unconfined compressive test on raw earth samples where the ultimate compressive strength of the material without fibers and its ductility in the presence of fibers was already established. In this work, the focus will be on the additional mechanical properties of this material as the material's toughness for its importance in the civil engineering studies.

In general, toughness is defined as the ability of the material to withstand impacts and dynamic loads. It is defined as the energy absorbed without cracking or energy needed to slow down the crack's propagation before fracturing [13].

Toughness is the combination of strength and plasticity [26], where tough material can take hard blows without rupturing.

Absolute toughness indicates the maximum and total amount of strain energy per unit volume that the material can absorb just before it fractures. The analysis of the plastic behavior of the material using the threshold of 15% of strength loss after reaching the peak was previously validated by Imanzadeh et al. [9].

Hence, in our case, absolute toughness is calculated as the area under the stress-strain curve until reaching a value of 85% of UCS after reaching the maximum. This quantity represents the entire area under the stress-strain diagram considering the strain at threshold of 15% of strength loss after reaching the peak as the maximum studied strain.

In conclusion, tough material should be ductile and strong at the same time. Stronger: Withstand higher load value; Tougher: resist changes and maintain properties under load.

For the following parts these notations are used:

i: Number of formulations = 1, 2 … 25.
j: Number of trials = 1,2 or 3.
$X_{i,j}$: is the toughness component of the trial j of the formulation i.

In some trials UCS (the maximum in the stress-strain graph) is not reached so this index couldn't be calculated and replaced by '/' or the curve looks strange.

In some trials 85% of UCS is not reached but due to obvious decrease in the curve after reaching UCS and a small interpolation the value of deformation at the 85%UCS stress could be calculated and hence the toughness index too.

4.1 Methodology

For the unit of this value obtained it should be in Joules per cubic meter to be as defined energy per unit volume.

$T_{i,j} = (A + A' + B)_{i,j}$ which is the total area under the stress-strain curve until reaching 85% of the UCS after reaching the peak. The three different regions are presented in Fig. 2.

4.2 Results

Within the three trials there might be incoherence between one and the other two causing the existence of high standard deviation. So, these incoherent values are deleted neglecting this intruding trial considered to be deficient in order to reach an acceptable standard deviation for all values.

The toughness obtained from the areas under the stress-strain curves for each of the three trials of each formulation is presented in Table 3 with the mean value for these three trials. Moreover the standard deviation (STD) which is the probability distribution of values is also presented; also, Table 3 contains the coefficient of variation (CV) which is the percentage of the ratio of the standard deviation to the mean value, expressed as percentage permitting the comparison between distributions of values whose scales of measurement are not comparable. For more accuracy, a low CV gives more precision.

Table 3. Absolute toughness values for the 3 trials of the 25 formulations with their mean value, standard deviation and the coefficient of variation.

Formulation	$T_{i,1}$ [kJ/m^3]	$T_{i,2}$ [kJ/m^3]	$T_{i,3}$ [kJ/m^3]	T_i [kJ/m^3]	STD [kJ/m^3]	CV (%)
1	90,52	103,63	86,81	93,66	8,84	9,44
2	/	/	/	/	/	/
3	11,81	/	/	11,81	/	/
4	/	/	/	/	/	/
5	/	/	/	/	/	/
6	9,01	/	/	9,01	/	/
7	74,53	65,85	72,96	71,11	4,63	6,50
8	101,51	73,23	95,08	89.94	14,82	15,08
9	30,48	46,42	37,51	38,14	5,82	15,26
10	43,65	35,25	/	39,45	5,94	15,06
11	44,31	26,75	34,85	35,30	7,39	20,93
12	51,24	57,71	38,64	49,20	10,16	20,65
13	82,42	107,32	55,48	94,87	25,93	27,33
14	88,94	107,90	90,36	95,73	10,56	11,03
15	/	/	/	/	/	/
16	11,85	/	/	11,85	/	/
17	10,52	/	11,02	10,77	0,35	3,28
18	41,16	49,65	51,71	47,51	5,59	11,77
19	/	/	/	/	/	/
20	/	29,41	34,37	31,89	0,88	2,62
21	36,22	37,25	36,33	36,60	0,57	1,55
22	29,98	32,23	33,89	32,03	1,96	6,13
23	45,02	35,12	35,17	38,44	5,70	14,83
24	68,87	64,03	65,07	65,99	2,55	3,86
25	34,79	48,07	48,48	43,78	7,79	17,79

The results can be shown more clearly in Fig. 3 where the standard deviation of the absolute toughness is shown for each formulation.

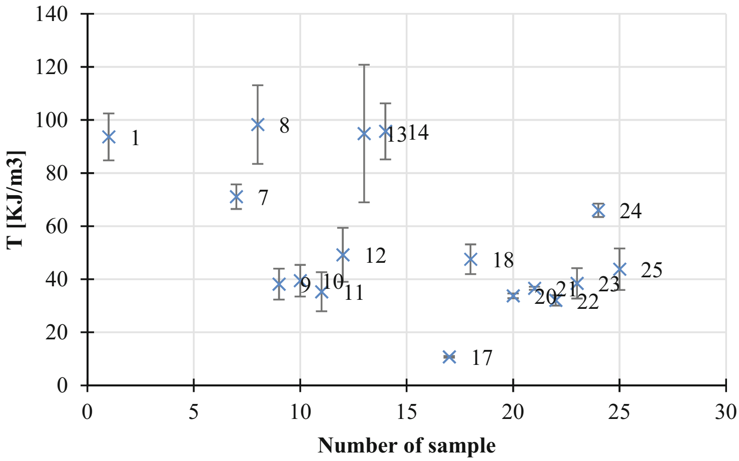


Fig. 3. Standard deviation of the mean value of the absolute toughness for each formulation

In order to minimize the uncertainties that come from the experiments, and decrease the observed statistical parameters, for each formulation, an introduing experiment is tried to be distinguished and removed. The intruding trial for each of the formulation is removed to obtain a lower CV and hence more precised results. An example on the intruding trial for formulation 1 is shown in the Fig. 3, where the second trial has a different behavior corresponding to the other two trials of the same formulation. Hence removing this false trial gave a much better standard deviation and a better coefficient of variation, leading to more precise results corresponding to the formulation. Same procedure was done on all the other formulations and the results are presented in the Table 4. In all the formulations, the intruding replicate was chosen in terms of difference in the area under the stress strain graphs compared to the other two graphs. In other words, this intruding one was found in a way to minimize the standard deviation for the toughness property of the material (Fig. 4).

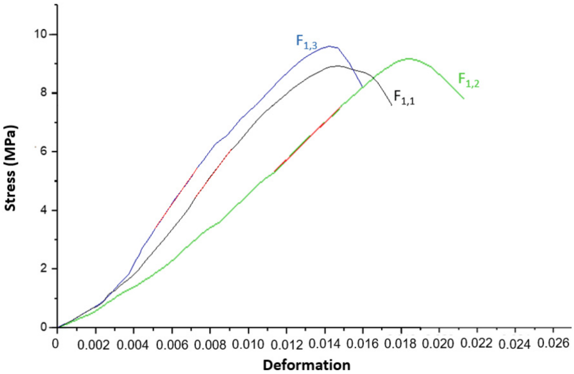


Fig. 4. Stress-Strain curve of the three trials of Formulation 1.

Hence in all the following analysis done, the adjusted mean values of each formulation from Table 4 is considered to have more precise results.

Table 4. Adjusted absolute toughness values for the 25 formulations with acceptable standard deviation and coefficient of variation.

Formulation	$T_{i,1}$ [kJ/m^3]	$T_{i,2}$ [kJ/m^3]	$T_{i,3}$ [kJ/m^3]	T_i [kJ/m^3]	STD [kJ/m^3]	CV (%)
1	90,52	/	86,81	88,67	2,62	2,96
2	/	/	/	/	/	/
3	11,81	/	/	11,81	/	/
4	/	/	/	/	/	/
5	/	/	/	/	/	/
6	9,01	/	/	9,01	/	/
7	74,53	/	72,96	73,74	1,11	1,51
8	101,51	/	95,08	98,29	4,55	4,63
9	30,48	/	37,51	34,00	1,44	4,22
10	43,65	35,25	/	39,45	5,94	15,06
11	/	26,75	34,85	30,80	3,61	11,71
12	51,24	57,71	/	54,48	2,45	4,50
13	82,42	107,32	/	94,87	17,61	18,56
14	88,94	/	90,36	89,65	1,00	1,12
15	/	/	/	/	/	/
16	11,85	/	/	11,85	/	/
17	10,52	/	11,02	10,77	0,35	3,28
18	/	49,65	51,71	50,68	1,46	2,87
19	/	/	/	/	/	/
20	/	29,41	34,37	31,89	0,88	2,77
21	36,22	37,25	36,33	36,60	0,57	1,55
22	29,98	32,23	33,89	32,03	1,17	3,66
23	/	35,12	35,17	35,14	0,04	0,10
24	68,87	64,03	65,07	65,99	2,55	3,86
25	/	48,07	48,48	48,27	0,29	0,60

This showed the high decrease in the coefficient of variation for most of the studied formulations meaning more precision with lower experimental uncertainties.

The results showed a huge improvement in terms of uncertainties and specially with reaching a coefficient of variation lower than 5% in most of the formulations and low standard deviation; except for formulations 10, 11 and 13 but still, this statistical parameter was decreased with this method (Fig. 5).

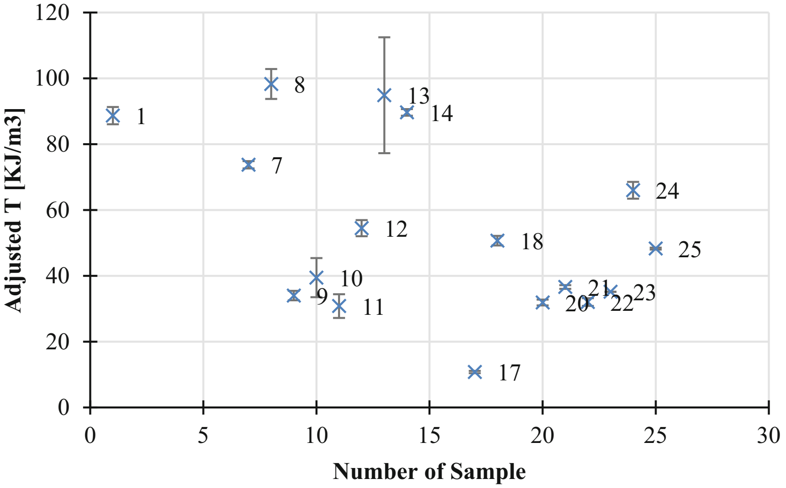


Fig. 5. Standard deviation of the mean value of the adjusted absolute toughness for each formulation

5 Conclusion

After studying the toughness of the raw earth concrete sample, the analysis of the results was improved by removing the intruding replicate. The toughness of the material is the energy of the material can handle before breaking, and it was used in this study as the energy needed for the material to reach its maximum strength and loses 15% of it. So, the absolute toughness was studied as the area under the stress-strain graph in the defined domain and it was clear how to distinguish the intruding formulation using this graph. Two statistical parameters were considered, the standard deviation and the coefficient of variation, to remove the intruding sample and improve these parameters to have more precise results.

Removing the intruding data improved the coefficient of variation to reach values lower than 5% for most of the formulations which is convenient with the recommendations.

As much as the experiments were done precisely, but a small distortion in the sample can cause a huge misleading in the tests results. That's is why replicates of the tests is recommended, but also the intruding data should be well located and removed from the analysis to reach a better precision. In this way the analyzed data can be modelized easier and better using the fundamental laws.

References

1. Přikryl, R., Török, Á., Theodoridou, M., Gomez-Heras, M., Miskovsky, K.: Geomaterials in construction and their sustainability: understanding their role in modern society. Spec. Publ. **416**(1), 1–22 (2016). https://doi.org/10.1144/SP416.21
2. Hibouche, A.: Présentée à Sols traités aux liants Performances hydro-mécaniques hydro et hygro-thermique Applications en BTP (2013)

3. Zak, P., Ashour, T., Korjenic, A., Korjenic, S., Wu, W.: The influence of natural reinforcement fibers, gypsum and cement on compressive strength of earth bricks materials. Constr. Build. Mater. **106**, 179–188 (2016). https://doi.org/10.1016/j.conbuildmat.2015.12.031
4. Ashour, T., Korjenic, A., Korjenic, S., Wu, W.: Thermal conductivity of unfired earth bricks reinforced by agricultural wastes with cement and gypsum. Energy Build. **104**, 139–146 (2015). https://doi.org/10.1016/j.enbuild.2015.07.016
5. Delgado, M.C.J., Guerrero, I.C.: Earth building in Spain. Constr. Build. Mater. **20**(9), 679–690 (2006). https://doi.org/10.1016/j.conbuildmat.2005.02.006
6. Kanema, J.M., Eid, J., Taibi, S.: Shrinkage of earth concrete amended with recycled aggregates and superplasticizer: Impact on mechanical properties and cracks. Mater. Des. **109**, 378–389 (2016). https://doi.org/10.1016/j.matdes.2016.07.025
7. Al-Mukhtar, M., Lasledj, A., Alcover, J.F.: Lime consumption of different clayey soils. Appl. Clay Sci. **95**, 133–145 (2014). https://doi.org/10.1016/j.clay.2014.03.024
8. Imanzadeh, S., Hibouche, A., Jarno, A., Taibi, S.: Formulating and optimizing the compressive strength of a raw earth concrete by mixture design. Constr. Build. Mater. **163**, 149–159 (2018). https://doi.org/10.1016/j.conbuildmat.2017.12.088
9. Imanzadeh, S., Jarno, A., Hibouche, A., Bouarar, A., Taibi, S.: Ductility analysis of vegetal-fiber reinforced raw earth concrete by mixture design. Constr. Build. Mater. **239**, 117829 (2020). https://doi.org/10.1016/j.conbuildmat.2019.117829
10. Chan, R., Bindiganavile, V.: Toughness of fibre reinforced hydraulic lime mortar. Part-1: quasi-static response. Mater. Struct. Constr. **43**(10), 1435–1444 (2010). https://doi.org/10.1617/s11527-010-9598-4
11. Chan, R., Bindiganavile, V.: Toughness of fibre reinforced hydraulic lime mortar. Part-2: dynamic response. Mater. Struct. Constr. **43**(10), 1445–1455 (2010). https://doi.org/10.1617/s11527-010-9599-3
12. Golewski, G.L.: Estimation of the optimum content of fly ash in concrete composite based on the analysis of fracture toughness tests using various measuring systems. Constr. Build. Mater. **213**, 142–155 (2019). https://doi.org/10.1016/j.conbuildmat.2019.04.071
13. Hibbeler, R.: Mechanics of Materials Tenth Edition in SI Units (2017)
14. Astm-Standard Practice for Classification of Soils for Engineering Purposes (Unified Soil Classification System)
15. Nf en 459-1 - Building Lime - Definitions, specifications and conformity criteria (2015)
16. Nf en 197-1 Composition, specifications and conformity criteria for common cements (2012)
17. Nf P15-318 - Cements à teneur en sulfures limitée pour beton précontraint (2006)
18. Eid, J.: New construction material based on raw earth: cracking mechanisms, corrosion phenomena and physico-chemical interactions. Eur. J. Environ. Civ. Eng. **22**(12), 1522–1537 (2018). https://doi.org/10.1080/19648189.2017.1373707
19. Abbar, B., et al.: Experimental investigation on removal of heavy metals (Cu2+, Pb2+, and Zn2+) from aqueous solution by flax fibres. Process. Saf. Environ. Prot. **109**, 639–647 (2017). https://doi.org/10.1016/j.psep.2017.05.012
20. Nf en iso 9342-2 - Verres étalons pour frontofocomètres pour le mesurage des lentilles de contact (2006)
21. Pierre, A., Mercier, R., Foissy, A., Lamarche, J.M.: The adsorption of cement superplasticizers on to mineral dispersions. Adsorpt. Sci. Technol. **6**(4), 219–231 (1989). https://doi.org/10.1177/026361748900600405
22. Bui, Q.B., Morel, J.C., Hans, S., Walker, P.: Effect of moisture content on the mechanical characteristics of rammed earth. Constr. Build. Mater. **54**, 163–169 (2014). https://doi.org/10.1016/j.conbuildmat.2013.12.067
23. Gitleman, L.: NF EN 206-1, Pap. Knowl. Towar. a Media Hist. Doc. (2014)
24. Nf p94-420 - Détermination de la résistance à la compression uniaxiale (2000)

25. Roches - Détermination du module de Young et du coefficient de Poisson - Nf p94-425 (2002)
26. S. P. S. Incorporated: Metal Properties: Hardness, Toughness, & Strength. https://www.polymersolutions.com/blog/defining-metal-properties/

Uncertainties of Experimental Tests on Cyclic Liquefaction Potential of Unsaturated Soils

Youssef Shamas[1,2](✉), Wenhao H. Huang[1,2], Khai Hoan Tran[2,3], Saber Imanzadeh[1,2], Armelle Jarno[2], Said Taibi[2], and Elie Rivoalen[1]

[1] Normandie Univ., INSA Rouen Normandie, Laboratoire de Mécanique de Normandie,
76801 Saint-Etienne du Rouvray, France
Youssef.shamas@insa-rouen.fr

[2] Normandie Université, UNIHAVRE, Laboratoire Ondes et Milieux Complexes,
CNRS UMR 6294, Le Havre, France

[3] Faculty of Civil Engineering and Environment,
Thai Nguyen University of Technology, Thai Nguyen City, Thai Nguyen Province, Vietnam

Abstract. Day after day, soil liquefaction took the attention of the researchers due to its huge and dangerous impact on its surroundings. Generally, soil liquefaction is related to saturated soils; however, recent studies showed that it is possible for unsaturated soils. Laboratory tests are done using dynamic triaxial test on unsaturated Hostun sand RF. As all experimental tests, uncertainties are unavoidable due to many factors affecting these tests from the sample preparation, the considered approximations (the stopping conditions of each phase of the test) and the heterogeneity of the material. Dynamic triaxial experiments were done on Hostun Sand RF samples with relative density of 50% in undrained conditions. This paper presents the possibilities of the repetitions of our experimental tests showing the statistical error (standard deviation and coefficient of variation) occurring between the same tests performed, focusing on the different parameters that might cause these uncertainties. The experimental tests carried out showed that it is impossible to be 100% repeatable and perfect, but the statistical errors and the uncertainties must be minimal compared to the complexity of the experimental test. To reduce these uncertainties, it is necessary to perform more replicated tests; however, in geotechnical field, it costs additional time and expenses.

Keywords: Cyclic Liquefaction Potential · Dynamic Triaxial Tests · Unsaturated soils · Hostun Sand RF · Statistical Errors · Uncertainties

1 Introduction

1.1 Uncertainties in Experiments

All kind of experiments are subjected to various types of uncertainties due to unavoidable errors. There are some random errors that come from human beings and their experimental skills and performance and precision in reading the experimental measurements. Plus, there is the systematic errors that come from tools utilized during the experiments

J. E. S. De Cursi (Ed.): UQSM 2023, LNME, pp. 30–41, 2024.
https://doi.org/10.1007/978-3-031-47036-3_3

[1]. If these types of errors are considered in the analysis of the received data, accuracy of the experiments increases effectively and a better correlation can be reached with better fittings [2].

Uncertainty is a term referred to a possible magnitude of error defined by Airy 1861 [3], where Moffat 1988 [4] described various types of errors in measurements for single-sample test. This kind of uncertainty for such tests is rarely considered because the mostly used one is for the multiple-samples tests. Moffat 1982 and 1985 [5, 6] has described the analysis of single-sample uncertainty.

1.2 Liquefaction and Dynamic Triaxial Tests

Liquefaction is a very dangerous phenomena that occurs when the pore pressure of the soil increases rapidly subjecting the soil particles to a high pressure. As a result, the sand particles lose their bonds and rapidly lose their strength and behave like a liquid under cyclic loading [7–10].

When soils liquefy, deformations develop rapidly and causes large-scale infrastructure collapse [11, 12].

In order to reproduce real soil consolidation conditions and perform cyclic loading on soil samples, dynamic triaxial apparatus is used in the laboratory to better understand liquefaction for partially saturated soils.

2 Materials and Test Procedure

For the experiment test listed in this paper, Hostun RF sand is used as a testing material. In the domain of liquefaction, Hostun RF sand is one of the most used materials in geotechnical tests for being considered as fine clear sand [13, 14]. This sand has a void ratio of 1.041 as maximum and 0.648 as minimum; it has 60% of it grains smaller than 400 µm and 10% smaller than 200 µm. This is clearly shown with the grain size distribution of this sand as shown in Fig. 1; where it can be seen also that this distribution lies inside the liquefiable zone of soils characterized by Iwaski, 1986 [15].

A sample with relative density of 50% corresponding to initial void ratio of 0.85 and initial water content of 19% is reconstructed using the wet tamping technique. This technique is widely considered in the literature for sample preparation for tests on liquefaction phenomena [16]. The sample with diameter of 70 mm and height of 140 mm is compacted using this method and it has initial saturation degree of 60%. A latex membrane of 0.3 mm thickness is used to prohibit the direct contact between the sample and the surrounding cell water.

The sample is well installed into the cell and the initial saturation degree is verified by the Skempton's parameter B measurement. After that, the sample is consolidated and subjected to cyclic loading and then the increase in pore pressure to this dynamic load is dissipated and then the sample was fully saturated using ramp for a long time.

Finally, the sample is removed and its final characteristics are measured and calculated.

Five experiments has been done for the purpose to study the uncertainty of the machines and all falls in the same range and the results shown in this paper is for the

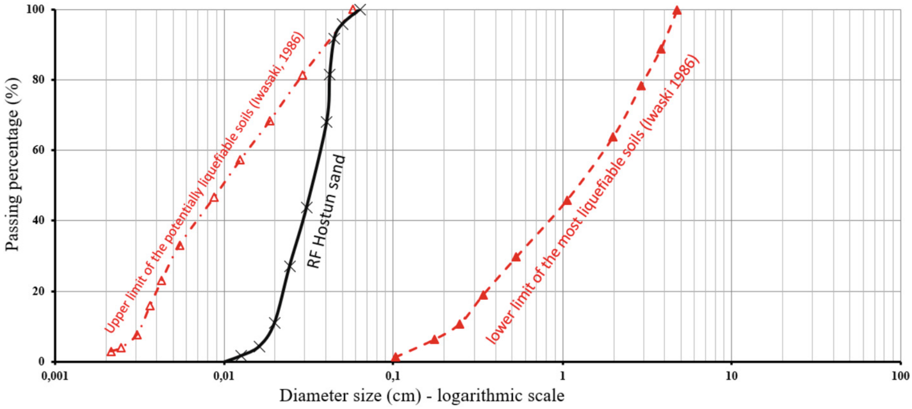


Fig. 1. Grain size distribution of Hostun RF sand

closest one to the average values. The other tests followed other paths from drained or undrained consolidated dynamic triaxial tests.

3 Tools

Different laboratory instruments are used during the test as the dynamic triaxial apparatus and its sub-devices as pressure controllers and axial displacement sensors and the Load transducers in addition to the balance used to measure the final characteristics of the sample.

3.1 Balance

The sample studied is reconstructed using sand and water with specific defined quantities to have the required relative density and needed initial saturation degree. To do so, the mass of the sand used and the amount of added water is measured using a balance of a maximum reading of 2010 g and reading error of ±0.01 g (Fig. 2).

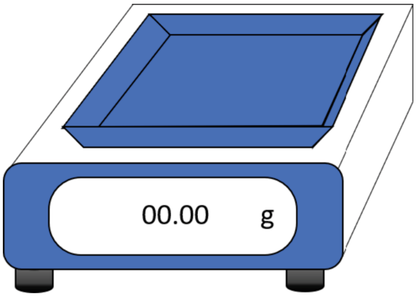


Fig. 2. Precise Balance

3.2 Load Transducer

The cell, where the sample is inside, is fixed at the base of the press and the load transducer is the one applying the cyclic loading from above. This load transducer can read a maximum of 5 kN with a total error of ±0.112%. As the test starts, the press is not used to move the sample towards the load transducer as the case in normal cases; rather than the load transducer is the one moving to apply dynamic load on the sample. So, in our tests, the load transducer's precision plays an important role in the calculations (Fig. 3).

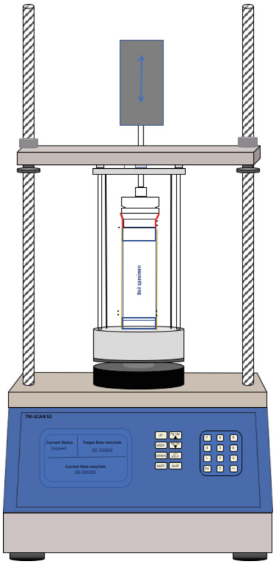


Fig. 3. Press and Load transducer

3.3 Axial Displacement Transducer

The axial displacement transducer is placed to measure the displacement of the load transducer during the dynamic load. This axial displacement transducer has a maximum reading of 25.73 mm and a non-linear error of 0.09% full scale. This error is considered in the calculations in the uncertainties of the measurements due to this transducer.

3.4 Pressure Controllers

To measure the pore water pressure inside the sample, an automatic pressure controller (APC) is used. This instrument can have a maximum reading of 1000 kPa in pressure and maximum of 200 cubic centimeters (cc) in volume of water. It has a pressure resolution of ±0.1 kPa and volume resolution of ±0.001 cc. The same system is used to control the cell pressure and measure the volume of water used in the process (Fig. 4).

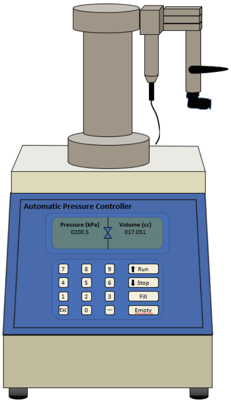


Fig. 4. Automatic pressure controller

4 Experimental Uncertainties Calculation

During the triaxial tests, various phases are followed from sample preparation to sample saturation if needed and its consolidation, then applying the cyclic loading on and after that dissipating the pressure induced due to this loading. Finally, saturation phase of the sample and then removing it to calculate its final characteristics as the final water content of the sample, considering that the sample is fully saturated after the final phase (Fig. 5).

To calculate the status and characteristics of the sample during the test, the most precise path is to follow it starting from its final state at the sample removal and going back step by step through the previous phases. This is due to lack in precision in well densifying the sample to the required density manually where the marge error will be significant compared to the calculated one due to random errors that come from the person doing the experiment.

The uncertainty of measurement in each step is calculated as using the differential error analysis [17, 18].

Considering that a final result f in function of variables x, y, z, etc.; the uncertainty of obtaining the value of is related to the uncertainty in the measurement of the variables as the following:

$$\Delta f = \left|\frac{\partial f}{\partial x}\right|\Delta x + \left|\frac{\partial f}{\partial y}\right|\Delta y + \left|\frac{\partial f}{\partial y}\right|\Delta z + \ldots \tag{1}$$

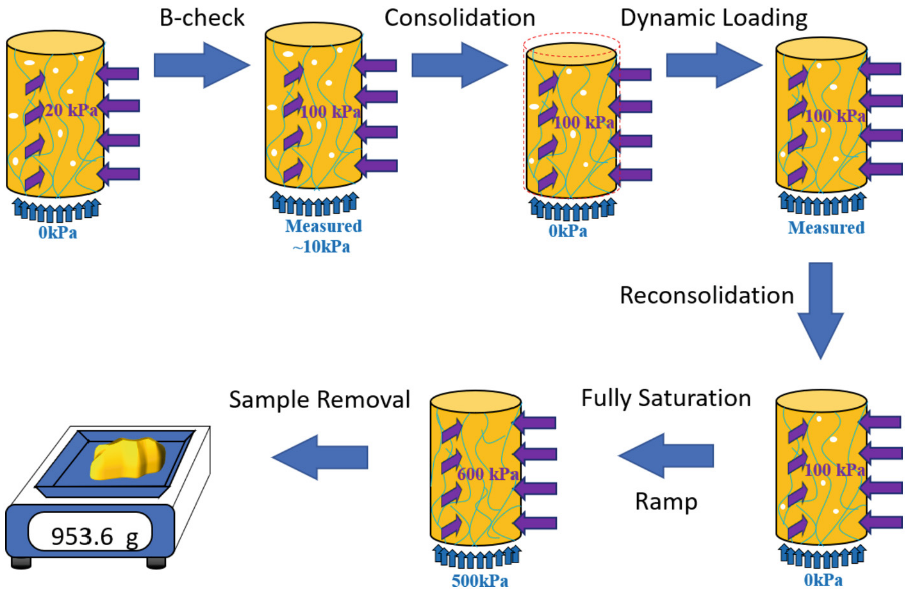


Fig. 5. Steps followed for dynamic triaxial test

5 Studied Variables

Based on the different tools defined, various variables are measured in each phase of the test and can be summarized in the Table 1. The values of the parameters measured during one of the tests are shown; where these values differ from one test to another depending on the state of the sample like its density and its saturation degree.

The percentage of the uncertainty in the measured values remains almost the same in the other tests followed.

Table 1. Variables Measured in various steps of the test

Variables	Phases				
	Saturation	Consolidation	Loading	Reconsolidation	Sample Removal
Volume of sand *Vs* (cm^3) – from balance	-	-	-	-	293.3
Volume of water in Sample from final state (cm^3) – from balance	-	-	-	-	205.0
Volume of Back Pressure Controller (cm^3)	129.5	137.1	137.1	202.5	-
Pressure from the Back-Pressure Controller (kPa)	10	0	-	500	-
Pressure from the Cell Pressure Controller (kPa)	30	100	100	600	-
Axial Displacement (mm)	-	-	±1.4, 2.1, 2.8, 3.5	-	-
Load Applied (N)	-	-	+1473.4 → −282.4	-	-

6 Results and Discussion

Starting from the last phase of our test, which is the sample removal. Considering that the error in this phase come from the balance alone. As the sample is removed, few sands get stuck on the membrane and these few grams are considered by measuring the weight of the membrane and the sand stuck on it together. After that, the membrane is cleaned well and dry it before we remeasure its weight alone. The difference in the weights gives the weight of the lost sand to have more precise calculations.

From the wet sand that is removed from the triaxial test, we measure its weight which is the weight of the sand and water measured as 953.60 g ± 0.01 g given by the balance used. This sample is entered to the oven for 24 h at 100 °C and then measured the weight of the completely dry sand and it is found to be 754.60 g ± 0.01 g.

From these measurements, the final water content is calculated using the following formula:

$$\mathrm{w}(\%) = \text{Water Content } (\%) = \frac{\mathit{Mass\ of\ water}}{\mathit{Mass\ of\ dry\ sand}}.100 = \frac{Mw}{Ms}.100. \tag{2}$$

The water content is calculated as 26.4%. To calculate the uncertainty of this measured value, Eq. 1 is used as the following.

Knowing that range error of balance is ±0.01 g, this uncertainty in measuring the mass of the sample will be ±0.01 g. From this, the uncertainty of measuring the mass of dry sand is the same. Then, the relative error of this measurement is 0.0013% of the measured value. The range error concerning the measurement of the weight of the water is the summation in errors of measurements of the wet sample (costing from sand and water) and the error from measuring the dry sand alone (Fig. 6). Then the relative error from measuring the weight of water is 0.01%. Finally using Eqs. 1 and 2 we can calculate the relative error in measuring the final water content as following:

$$\frac{\Delta w}{w} = \frac{\Delta Mw}{Mw} + \frac{\Delta Ms}{Ms} = \pm 0.0113\% \tag{3}$$

Then the error of measuring water content is:

$$\Delta w = \pm 0.0113\%.26.4\% = \pm 0.003\% \tag{4}$$

Then the final water content of the sample is 26.4% ± 0.003%.

From the final phase, final water content of the sample is calculated. Going one step before, the is the saturation phase of the sample using ramp by increasing the pore water pressure and the cell pressure at the same time. So, there is increase in effective stress on sample and the change in volume measured in this phase from automatic pressure controller is just the volume of air in sample replaced by the volume of water due to the applied pressure.

The volume of water in sample at the end of sample is calculated in previous paragraph by the balance. Plus, from the back-pressure controller, the volume of water added to the sample during this phase is measured. Hence, the initial state of the sample at the beginning of the phase of saturation can be calculated.

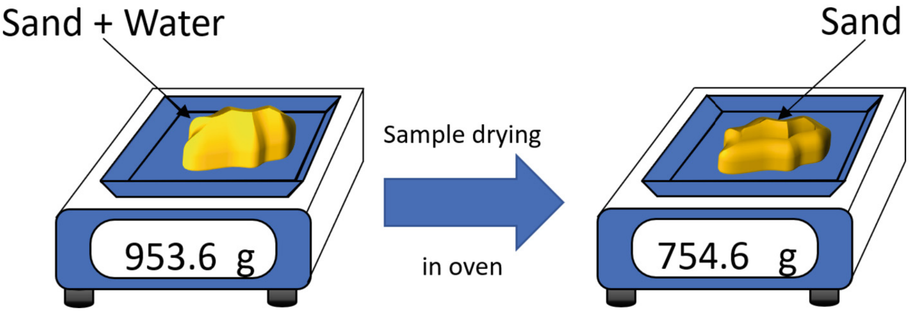


Fig. 6. Measuring of sample before and after drying it

The volume of water at the end of saturation phase is:

$$Vw_f = 205.00\,\text{cm}^3 \pm 0.02\,\text{cm}^3 \tag{5}$$

The volume measured by the APC during the saturation phase is:

$$deltaV = 77.57\,\text{cm}^3 \pm 0.001\,\text{cm}^3 \tag{6}$$

Then the volume of water in sample at the beginning of saturation phase is:

$$Vw_{i-sat} = \left(205.00\,\text{cm}^3 - 77.57\,\text{cm}^3\right) \pm \left(0.02\,\text{cm}^3 + 0.001\,\text{cm}^3\right)$$

$$Vw_{i-sat} = \left(127.43\,\text{cm}^3\right) \pm \left(0.021\,\text{cm}^3\right) = 127.43\,\text{cm}^3 \pm 0.016\% \tag{7}$$

As the value of water measured is smaller, the range error decreases too.

As a result, at the beginning of saturation ramp, the sample characteristics are presented in the following table (Table 2):

Table 2. Volumes of different parts of sample before saturating the sample

	Value (cm^3)	Error (cm^3)	Relative Error (%)
Volume of sand Vs	293.3	0.004	0.001
Volume of water Vw_{i-sat}	127.43	0.021	0.016
Volume of air Va_{i-sat}	77.57	0.001	0.001
Total volume of sample V_{i-sat}	418.24	0.026	0.006

From this table, it can be seen that the percentage error decreases due to using the APC system with higher precision than the balance used. Moreover, an error of 0.026 cm^3 is still very small.

From these characteristics, the void ratio (e) of the sample can be calculated at the initial state of saturation phase. The void ratio is calculated as:

$$e_{i-sat} = \frac{Volume\ of\ Void}{Volume\ of\ dry\ sand} = \frac{Vv}{Vs} \tag{8}$$

So, as calculated with the water content, the relative error in this void ratio is found to be 0.012% and the value is 0.69. So, the initial void ratio in ts step is 0.69 ± 0.00008 where the error is negligible.

The saturation degree (Sr) of the sample at initial state is calculated as:

$$Sr_{i-sat} = \frac{\gamma_s}{\gamma_w} . \frac{w_{i-sat}}{e_{i-sat}} \tag{9}$$

where γ_s and γ_w are 2.65 g/cm^3 and 1 g/cm^3 respectively and are constants. So, the error in the saturation degree is calculated from the errors in water content and void ratio. Then, initial saturation degree is 62.9% ± 0.021%.

The same concept is followed in each step of the phases and in each step the void ratio and the saturation degree is calculated to reach a real initial saturation degree ranging between 55.774% and 55.826% of a value of 55.8% with relative error of 0.1% only. This shows the high precision of the system used compared to that of the researcher performing the experiment. The initial saturation degree of the sample is calculated to be 55% which is relatively far from the expected value which is supposed to be 50%. The relative error of this experiment considering the saturation degree is 10% using the formula given by Derenzo, 2010 [19] which the difference between calculate value from the experiment and the expected value from sample preparation with respect to the expected one. The same relative error for the relative density of the sample, or in other words its void ratio, is found to be 3% only. Meaning that the compaction error of the sample is less important than preparing the sample where there might be error occurring when preparing the quantities of sand and water used.

Before the dynamic load phase, the volume of water in sample is calculated to be 137.05 cm^3 ± 0.023 cm^3 and the value of volume of air in the unsaturated in the sample is 85.07 cm^3 ± 0.006 cm^3. The volume of air in this phase is calculated from the change in volume of the cell measured by the APC system considering that in unsaturated sample the change in volume of sample in undrained conditions is due to the change in the volume of air in it.

The load transducer and displacement transducer have very small error compared to the value measured and no difference can be seen in the measured values even with considering the errors from the system during the cyclic loading phase.

The samples is subjected to cyclic loading with the axial strain increasing by 1% in double amplitude after each 10 cycles (Fig. 7) using the same protocol utilized in the literature [20, 21].

The pore pressure of sample increased to 12 kPa only and stabilized there where the applied confining pressure applied is 100 kPa (Fig. 8). Meaning that there is no liquefaction in the sample due to relatively highly unsaturation degree of the sample.

The stress (load applied on sample over its cross section) versus the strain (displacement applied over the initial height of the sample) is shown in Fig. 9. It shows the decrease in the slope of the graph as the double amplitude applied increases. Finally, Fig. 10 shows the stress versus mean applied pressure on sample.

These figures show that the sample didn't liquify as the pore pressure of the sand didn't increase to reach the same value as that of the cell pressure. Depending on different values, such as relative density and saturation degree, these graphs can change. As the

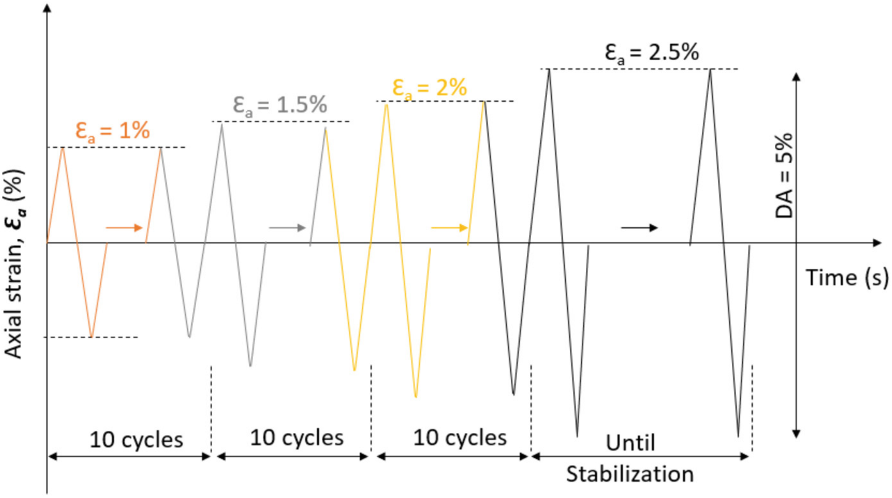


Fig. 7. Applied axial strain protocol

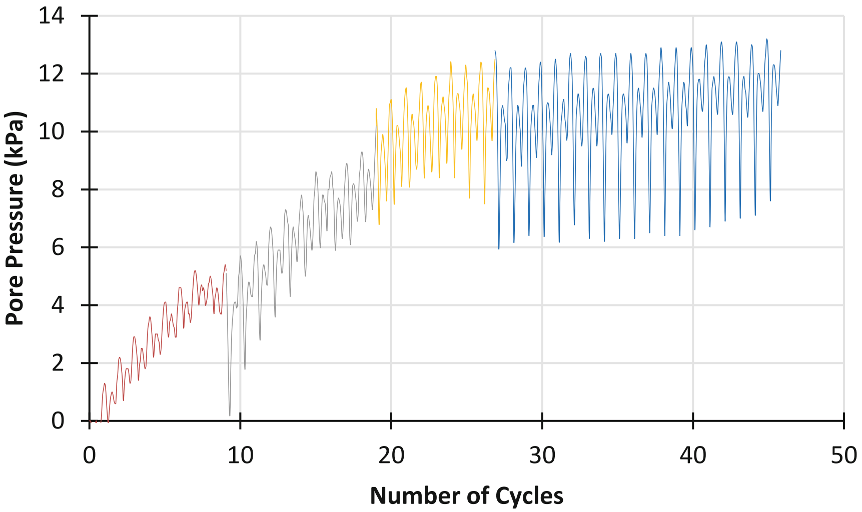


Fig. 8. Pore Pressure versus Number of cycles

relative density decreases meaning that the sample is more loose and as the saturation degree increases, the liquefaction potential of the sample increases and the change in pore pressure in Fig. 8 can increase to reach the value of 100 kPa leading to a null effective stress at liquefaction. Moreover, the stress reached at liquefaction phase in Fig. 9 can reach zero. Yet, the uncertainty in the values remains the same, whatever is the case of the test studied, as they are measured by the instruments.

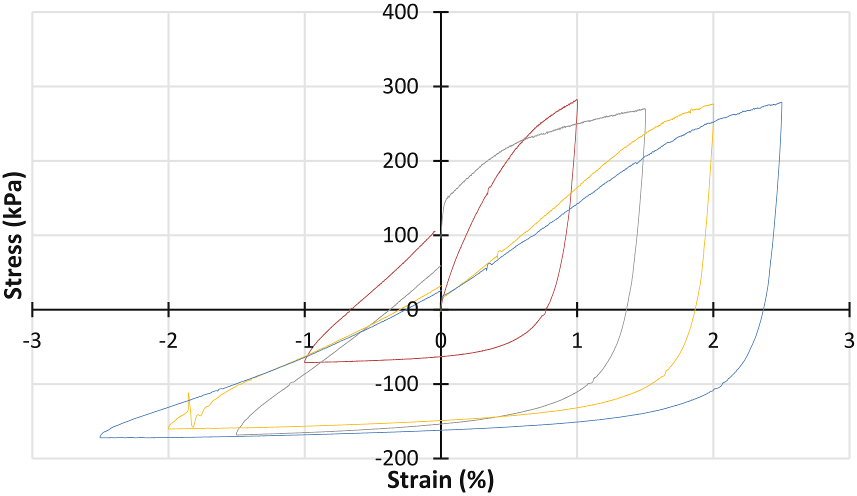


Fig. 9. Stress versus strain for the first cycle for each step of applied double amplitude

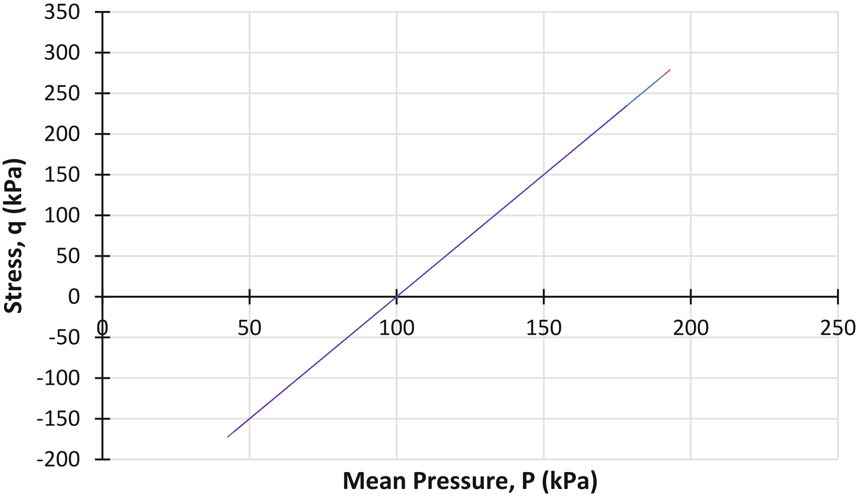


Fig. 10. Stress versus mean applied pressure

7 Conclusion

This paper studied the relative error and the error of the measurements done during a triaxial test on a sample of initial degree of saturation 55% and saturation degree of 63%. Even the sample is not liquefied, but the study of these uncertainties is essential to prove the high precision of the tools used during the test. These uncertainties can be considered later also in any modulization to obtain more precise results and a best fitting model.

Finally, in all these errors, the random errors from the human being is more important than that of the systematic errors coming from the experimental tools used.

References

1. Richard, O.: The Quality of Measurements. (2021). https://doi.org/10.1007/978-1-4614-1478-0
2. Mašín, D.: The influence of experimental and sampling uncertainties on the probability of unsatisfactory performance in geotechnical applications. Geotechnique **65**(11), 897–910 (2015). https://doi.org/10.1680/jgeot.14.P.161
3. Airy, G.B.: Theory-Of-Errors-Of-Observations.pdf (1861)
4. Hernández-Hernández, V.A., et al.: Experimental and numerical analysis of triaxial compression test for a clay soil. Chil. J. Agric. Res. **81**(3), 357–367 (2021). https://doi.org/10.4067/S0718-58392021000300357
5. Moffat, R.J.: Contributions to the theory of single-sample uncertainty analysis. J. Fluids Eng. Trans. ASME **104**(2), 250–258 (1982). https://doi.org/10.1115/1.3241818
6. Moffat, R.J.: Using uncertainty analysis in the planning of an experiment. J. Fluids Eng. Trans. ASME **107**(2), 173–178 (1985). https://doi.org/10.1115/1.3242452
7. Tran, K.H., Imanzadeh, S., Taibi, S., Souli, H., Fleureau, J.M., Pantet, A.: Some aspects of the cyclic behavior of quasi-saturated sand, vol. 1, pp. 3–6 (2018)
8. Castro, G.: Liquefaction of Sands. Harvard Univ. Harvard Soil Mech. Ser. **81** (1969)
9. Seed, H.B., Idriss, I.M.M.: Ground motions and soil liquefaction during earthquake. Earthquake Engineering Research Institute, California, p. 134 (1982)
10. Soil liquefaction during earthquakes I (1720)
11. Dunn, D.: Benefits of membership. IEEE Ind. Appl. Mag. **13**(2), 78–79 (2007). https://doi.org/10.1109/MIA.2007.322261
12. Powrie, W., Wood, D.M., Modelling, G.: Groundwater Lowering in Construction: A Practical Guide to Dewatering. Applied Geotechnics Series (2016). http://www.taylorandfrancis.com
13. Benahmed, N., Benahmed, N., Canou, J.: Propriétés de liquéfaction et structure des sables laches. 7ième Colloq. Natl. AFPS - École Cent. Paris, January 2015, p. 8 (2007)
14. Dubujet, P., Doanh, T.: Undrained instability of very loose Hostun sand in triaxial compression and extension. Part 2: theoretical analysis using an elastoplasticity model. Mech. Cohesive-Frictional Mater. **2**(1), 71–92 (1997). https://doi.org/10.1002/(SICI)1099-1484(199701)2:1%3c71::AID-CFM25%3e3.0.CO;2-9
15. Iwasaki, T.: Soil liquefaction studies in Japan: state-of-the-art. Soil Dyn. Earthq. Eng. **5**(1), 2–68 (1986). https://doi.org/10.1016/0267-7261(86)90024-2
16. Sladen, J.A., Krahn, J., Hollander, R.D., Hird, C.C., Hassona, F.: A state parameter for sands. Geotechnique **36**(1), 123–132 (1986). https://doi.org/10.1680/geot.1986.36.1.123
17. Taylor, J.R.: [John_R_Taylor]_An_Introduction_to_Error_Analysis(BookZZ.org).pdf (1997)
18. Wells, M.M.: Small genetic distances among populations of green lacewings of the genus Chrysoperla (Neuroptera: Chrysopidae). Ann. Entomol. Soc. Am. **87**(6), 737–744 (1994). https://doi.org/10.1093/aesa/87.6.737
19. Derenzo, S.E.: Experimental uncertainties. Pract. Interfacing Lab., pp. 508–509 (2010). https://doi.org/10.1017/cbo9780511615160.013
20. Unno, T., Kazama, M., Uzuoka, R., Sentos, N.: Liquefaction of unsaturated sand considering the pore air pressure and volume compressibility of the soil particle skeleton. Soils Found. **48**(1), 87–99 (2008). https://doi.org/10.3208/sandf.48.87
21. Mele, L., Lirer, S., Flora, A.: The effect of confinement in liquefaction tests carried out in a cyclic simple shear apparatus. In: E3S Web Conference, vol. 92, pp. 1–5 (2019). https://doi.org/10.1051/e3sconf/20199208002

Analysis of the Impact of Uncertainties on the Estimation of Geotechnical Engineering Properties of Soil from SPT on the Design of Aerogenerators Foundation

Giullia Carolina de Melo Mendes[1](✉), Alfran S. Moura[1], Saber Imanzadeh[2,5],
Marcos Fábio Porto de Aguiar[3], Lucas F. de Albuquerque Lima Babadopulos[4],
Said Taibi[5], and Anne Pantet[5]

[1] Department of Hydraulic and Environmental Engineering, Universidade Federal do Ceará,
Fortaleza, Brazil
giucmendes@alu.ufc.br, alfransampaio@ufc.br

[2] Normandie Univ., INSA Rouen Normandie, Laboratoire de Mécanique de Normandie,
76801 Saint-Etienne du Rouvray, France
saber.imanzadeh@insa-rouen.fr

[3] Department of Civil Construction, Ciência e Tecnologia do Ceará, Instituto Federal de
Educação, Fortaleza, Ceará, Brazil
marcosporto@ifce.edu.br

[4] Department of Structural Engineering and Civil Construction, Universidade Federal do Ceará,
Fortaleza, Brazil
babadopulos@ufc.br

[5] Normandie Université, UNIHAVRE, Laboratoire Ondes et Milieux Complexes,
CNRS UMR 6294, Le Havre, France
{Said.Taibi,anne.pantet}@univ-lehavre.fr

Abstract. In geotechnical engineering, it is common to use data from only one field test (SPT test) to predict input stiffness parameters in the study of stress *vs.* displacements behaviour of foundations. This is made from correlations available in the literature for different kinds of soils. As a result, the variation that occurs between different correlations may be significant and must be critically analysed with respect to the accuracy of the foundation design and, consequently, its safety. In this context, this paper aims to study the impact of the variations of friction angle (ϕ') and Young's modulus (E) predicted by several different correlations from field SPT measurements available in the literature. Based on the estimations, four groups of estimated results were defined with the corresponding values of ϕ' and E within such groups (for high and low values of both ϕ' and E). Such values were applied in a numerical Finite Elements Method (FEM) model of an aerogenerators foundation to calculate vertical displacements and stress fields. In the groups in which only one of the parameters was varied, it was observed that the Young's modulus has a significant influence on the displacements, while that was not the case for the friction angle in the investigated foundation, due to predominant, linear-elastic condition in the investigated foundation. The paper demonstrated the significant variation in geotechnical analysis that can occur with the use of different input correlations in geotechnical studies. These uncertainties

J. E. S. De Cursi (Ed.): UQSM 2023, LNME, pp. 42–53, 2024.
https://doi.org/10.1007/978-3-031-47036-3_4

lead either to overestimate or to underestimate the foundation design, which may affect economy and safety, thus emphasizing the need for more accurate field tests and more laboratory investigation and control.

Keywords: Foundation Behaviour · Numerical Modelling · Soil Parameters · Geotechnical Measurements

1 Introduction

The use of wind power gained prominence for being an abundant source of renewable energy in some regions of the world. It can reduce fossil fuel consumption. In this context, this is concomitant with the ONU's 7th Sustainable Development Goal: to ensure access to affordable, reliable, sustainable, and modern energy for all. Wind energy provides several socioeconomic and environmental benefits, and is one of the most cost-effective energy sources in Brazil (ABEEÓLICA, 2020).

In this context, the aerogenerators structures and soil under the foundation must be studied from the engineering point of view. The onshore wind turbines are supported in reinforced concrete foundations. Due to technological advances that cause the increase of tower's height and blades length, larger foundations of the order of hundreds of cubic meters and with high diameters are more and more common (SILVA, 2014). In addition to structural considerations, geotechnical analyses are needed to ensure proper design and the stability of the tower.

One of the biggest challenges for geotechnical studies considering Brazilian sites is related to the geotechnical investigation. There are significant differences between Brazil and France with respect to the employed tests (MILITITSKY, 2019). In Brazil, most commonly, the Standard Penetration Test (SPT), exclusively, is performed, while in France the Pressiometric Ménard Test (PMT) is conducted, and, in some cases, it is accompanied by seismic tests, as well as laboratory triaxial tests with field materials. Even in more developed countries that is not always the case, and in countries under development it is seldom the case.

Thus, analysis of the stresses and displacements is necessary to design and, for that, the geotechnical investigation is of paramount importance to determine parameters of soil such as Young's modulus, Poisson ratio, friction angle, cohesion, unit weight and dilatancy angle. These parameters can be estimated with a laboratory testing campaign. However, the extraction of undeformed samples is, usually, logistically and economically impractical. Alternatively, different authors propose the use of correlations with the required input parameters and the results obtained from the Standard Penetration Test (SPT). This test is the most common to be executed, being the most widely used in foundation projects in Brazil and in many cases, the only one to be done (CINTRA et al., 2013). It consists of penetrating the soil with a standard hammer of 65 kg forced into the hole with strokes of 75 cm of height and counting the number N_{SPT} of strokes needed to penetrate 30 cm.

The practice of determining only SPT results during geotechnical investigations to estimate values of the engineering properties, such as the friction angle (ϕ') and the Young's modulus (E), leads to strong variability, following the different available correlations and the experience of the analyst, affecting the design directly. Then, the choice of the most appropriate method the field conditions can affect the final analysis and this will be evaluated in this article. Thereby, the objective of this study is to evaluate the interference of the input parameters variation on a numerical modelling of stresses and displacements under the aerogenerator foundation and to analyse the influence of such uncertainties on its design.

2 Methodology

The research presented in this article deals with the variation of input parameters in a numerical structural modelling of wind turbines' foundation. The idea is to understand the stress and displacement behaviour when different methods to determine the resistance and deformability parameters are used, for understanding which range of variation on estimated input parameters affects significantly the results in terms of the analysed stresses and displacements.

Data from the Cacimbas wind farm (located in Trairi-Ceará-Brazil) was used. Figure 1 presents (a) the location of Ceará and (b) the location of Trairi.

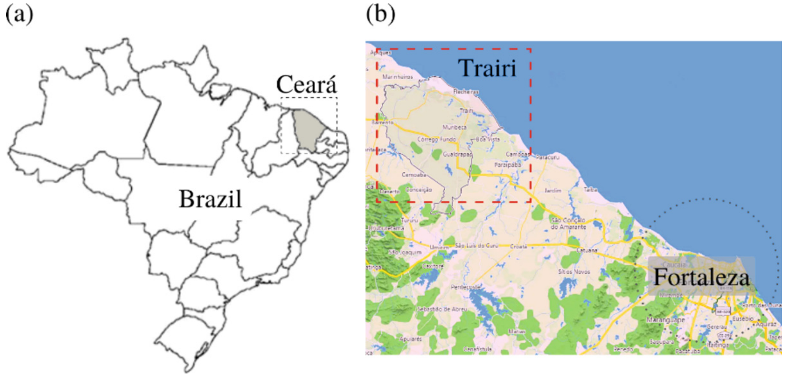


Fig. 1. Location of (a) Ceará and (b) Trairi on Brazilian map.

In this site, the geotechnical campaign was conducted, consisting of seven SPT tests ranging between 14 m and 22 m of depth. The obtained geotechnical profile is presented in Fig. 2.

SPT-03 was chosen for the investigations in this paper. The reason for this choice is because the water is at a more critical level, with most part of soil in a saturated condition. Figure 3 presents in more details the SPT profile with variation of number N as a function of the depth.

From the presented data, analysing the variation of sand compactness, the foundation was conceptually divided into five layers of soil: (i) loose sand from 0 m to 6 m; (ii) one

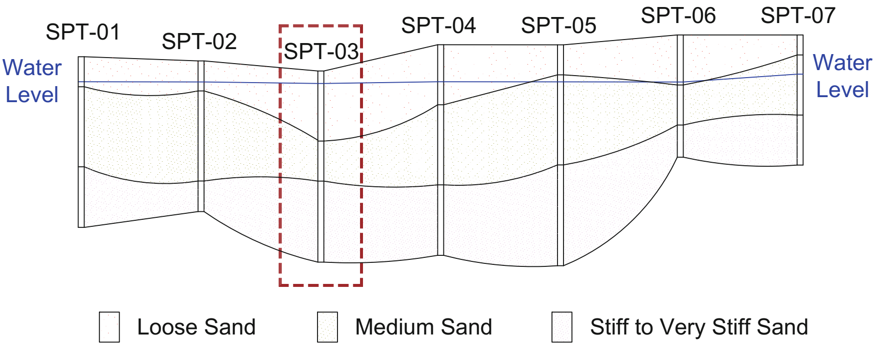


Fig. 2. Geotechnical profile of Cacimbas site.

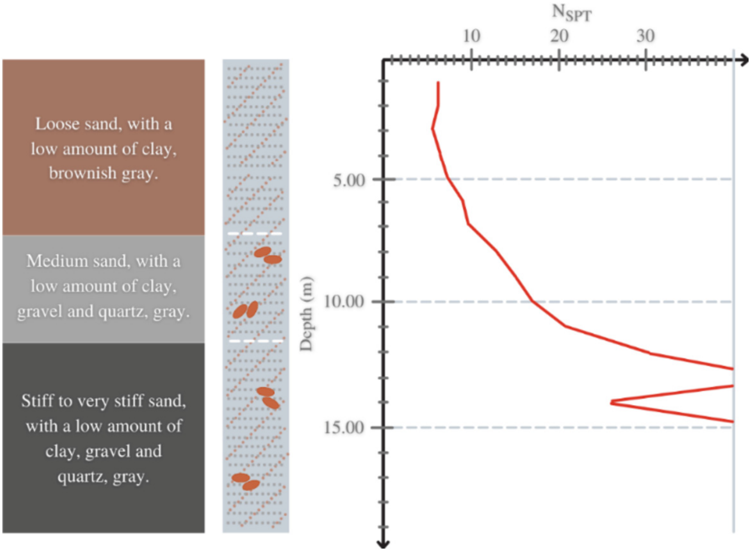


Fig. 3. Geotechnical profile of Cacimbas wind farm site.

medium sand from 6 m to 8 m; (iii) another medium sand from 8 m to 10 m; (iv) another medium sand from 10 m to 12 m; and (v) a stiff to very stiff sand from 12 m to 80 m. The 80 m depth is related to the boundary condition of the model and the intermediate layers, composed of medium sand, were divided to predict the variable parameters (ϕ’, E) with greater precision. Based on those layers, the geotechnical parameters fixed (γ, γsat, c’, Ψ and ν) were obtained with typical values from the literature for similar materials, as presented in Table 1.

Based on the presented data, the parameters ϕ’ and E were estimated using different correlations from field SPT measurements available in the literature. To determine ϕ’, the correlations used were: Kulhawy & Mayne (1990) – Propositions 1 and 2, Wolff (1989), De Mello (1971), Godoy (1983), Teixeira (1996), Meyerhof (1959) – Yoshida

Table 1. Soils parameters estimated from SPT-03 of the Cacimbas wind farm.

Parameter	Layer 1	Layer 2, 3 and 4	Layer 5	References
Moist unit weight of soil – γ	18 kN/m^3	19 kN/m^3	20 kN/m^3	Godoy (1972)
Saturated unit weight of soil – γ_{sat}	19 kN/m^3	20 kN/m^3	21 kN/m^3	Godoy (1972)
Cohesion – c'	10 kPa	10 kPa	10 kPa	Moura et al. (2014)
Dilatancy angle – Ψ	0°	5°	10°	Pinto (2006), Moura et al. (2014)
Poisson's ratio – ν	0.3	0.3	0.3	-

(1988), Muromachi (1974), Hatanaka & Uchida (1996). To determine E, the correlations used werere: Mikhejev (1961), Bowles (1996) – Propositions 1 and 2, De Mello (1971), Décourt et al. (1989), Teixeira & Godoy (1996), Trofimenkov (1974), De Freitas et al. (2012), Makwana & Gandhi (2019) and Afonso (2016). The expressions are given in Tables 2 (for friction angle) and 3 (for Young's modulus).

Table 2. Correlation methods considered for friction angle estimation.

Expression	References
$\phi' = [15.4.(N_1)_{60}]^{0.5} + 20$	Hatanaka & Uchida (1996)
$\phi' = 28 + 3.75.\sigma'^{-0.012}_{v}.\, N^{0.46}_{60}$	Meyerhof (1959) – Yoshida (1988)
$\phi' = 27.1 + 0.3.(N_1)_{60} - 0.00054.(N_1)^2_{60}$	Wolff (1989)
$\phi' = \sqrt{20.N} + 15$	Teixeira (1996)
$\phi' = 20 + 3.5.\sqrt{N}$	Muromachi (1974)
$\phi' = 54 - 27.6034.\exp\,\exp\left(-0.014.(N_1)_{60}\right)$	Kulhawy & Mayne (1990) Proposition 1
$\phi' = \mathrm{tag}^{-1}\left(\frac{N}{12.2+0.2.\sigma_v\prime}\right)^{0.34}$	Kulhawy & Mayne (1990) Proposition 2
$\phi' = 28° + 0.2.\, N$	Godoy (1983)
$\phi' = \mathrm{acrtg}\left(\frac{0.712}{1.49-\sqrt{\frac{N}{0.28.\sigma_v\prime+27}}}\right)$	De Mello (1971)

Table 3. Correlation methods considered for Young's modulus estimation.

Expression	References
$E = (15000 \text{ to } 22000).\ln N_{55}$	Mikhejev (1961)
$E = (2600 \text{ to } 2900).N_{55}$	Bowles (1996) – Proposition 1
$E = 6000.\ N_{55}$	Bowles (1996) – Proposition 2
$E = 3.(N - 3)$	De Mello (1971)
$E = 3.5.\ N_{60}$	Décourt et al. (1989)
$E = \alpha.\ K\ .\ N$	Teixeira & Godoy (1996)
$E = 43.1.(\log N_{60})$	Trofimenkov (1974)
$E = 8000.\ N_{60}^{0.8}$	De Freitas et al. (2012)
$E = 0.3925.\ N_{60} + 54.25$	Makwana & Gandhi (2019)
$E = 2.9.\ N + 2.7$	Afonso (2016)

3 Results and Discussions

3.1 Parametric Study and Numerical Modelling Conditions

The trends of the estimation results for ϕ' are presented in Fig. 4. They consider the correlations from Table 2 and the SPT results from Fig. 3.

Based on the behaviour presented by the chosen correlations, it is noted that there are two defined trends formed. The trend 1, by Kulhawy & Mayne (1990) – P1, Wolff (1989) and Godoy (1983), and the trend 2 by Meyerhof (1959) – Yoshida (1988) and Kulhawy & Mayne (1990) – P2. The other methods do not show representative behaviour, compared to the methods cited and considering the soil type. In this case, the soil is a sand and based on literature, the friction angle should have values around 28°–35°, as shown by Pinto (2006), Cintra et al. (2011) and others authors in literature, proving that De Mello (1971) and Teixeira (1996) were not representative of the range of common values. The Muromachi (1974) and Hatanaka & Uchida (1996) were excluded because they did not exhibit similar behaviour to any of the groups.

Group A was chosen as reference, because based on local experience and observations by Gonin et al. (1992), this trend is more adequate for sand in the relative density observed in the SPT test. Group B was chosen to model the situation varying the input parameters, but, in general, for local experience, these values have a low order of magnitude.

The trends of the estimation results for E are presented in Fig. 5. They consider the correlations from Table 3 and the SPT results from Fig. 3.

Based on the behaviour presented by the correlations, it is noted that there are three defined trends formed. Group A comprises De Mello (1971), Décourt et al. (1989), Teixeira & Godoy (1996), De Freitas et al. (2012) and Afonso (2016). This group is formed by methods developed by Brazilian authors and represents the highest values found. Based on this and considering the local experience and results shown by Correia (2004), this group was utilized as reference.

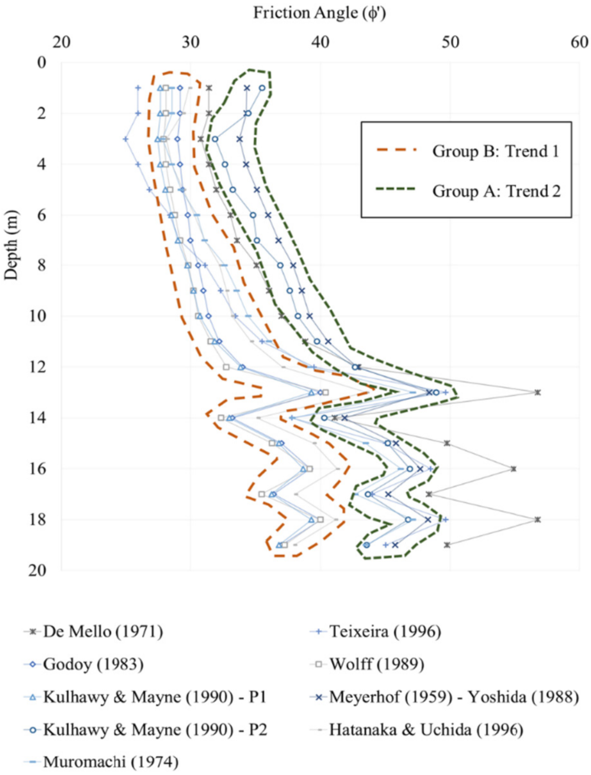


Fig. 4. Groups of behaviour trend for friction angle as a function of the depth.

Group B was formed by the behaviour trend of Mikhejev (1961) with minimum (E = 15000.ln N_{55}, cf. Table 3) and medium (E = 22000.ln N_{55}, cf. Table 3) considerations, Bowles (1996) – Proposition 1 with minimum (E = 2600.N_{55} cf. Table 3) and medium (E = 2750.N_{55}, cf. Table 3) values and Trofimenkov (1974). The variation of relative density is not significative (Correia, 2004) and based on this information, the Group B was considered for varying the input parameters in the FEM modelling.

For the numerical analyses, when it comes to the soil, the SPT-03 was considered to determine parameters and layers of soil. In the model generated, five layers were utilized within the considered depth of 80 m. Basic geotechnical parameters were based on Table 1, and soil constitutive behaviour analysed according to Mohr-Coulomb theory and input parameters based on the groups evaluated in Figs. 4 and 5.

The finite element mesh was created in a parallelepiped format with 160 m × 160 m × 80 m. In the lateral boundaries, only horizontal displacements were fixed at zero. In the bottom boundary all the displacements were zero (Fig. 6a). The mesh of Finite Element Method (FEM) comprises 10-node tetrahedral elements (Fig. 6b). A mesh sensitivity analysis was performed considering default software meshing “very coarse”, “coarse”, “fine” and “very fine” compared to “medium”. Based on obtained results, it was chosen to

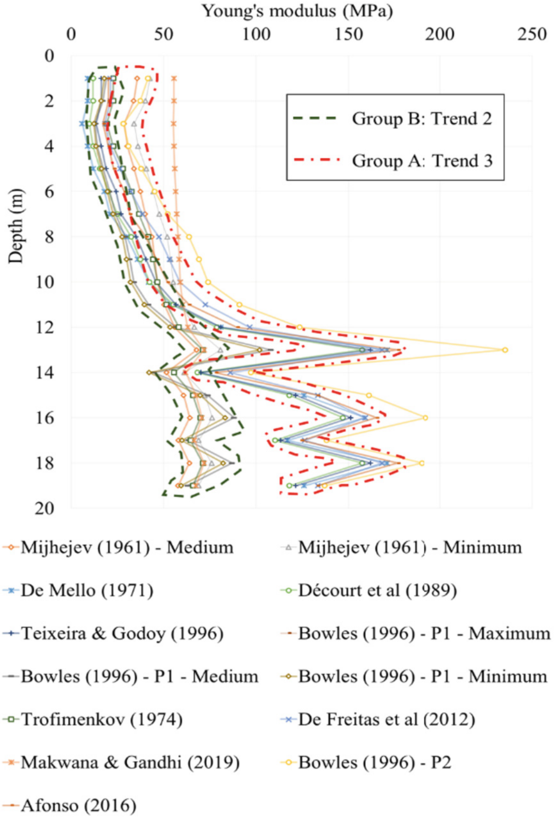


Fig. 5. Groups of behaviour trend for Young's modulus as a function of the depth.

use the medium mesh, given that the variation is not significant and this mesh adequately represents the situation in the numerical model.

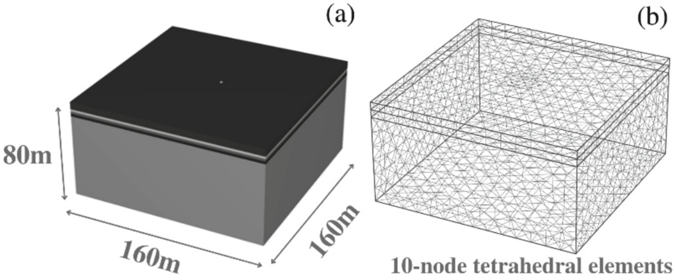


Fig. 6. Geometries of the considered numerical model of the foundation soil for (a) the soil layers-limit and (b) the finite elements mesh format. Boundary conditions consider zero-horizontal displacement on the sides and fixed condition at the bottom.

With respect to geometry and parameters of the foundation concrete structure, it is used the data provided by Imanzadeh et al. (2021) that considers a reinforced concrete with elastic behaviour; thus, the parameters were 25 kN/m^3 for unit weight, 30 GPa for Young's modulus and 0.2 for Poisson's ratio. The characteristics of the foundation was a superficial structure, circular with 19.8 m of diameter and 4.2 m of depth in which is a rigid plate element on PLAXIS 3D (Fig. 7).

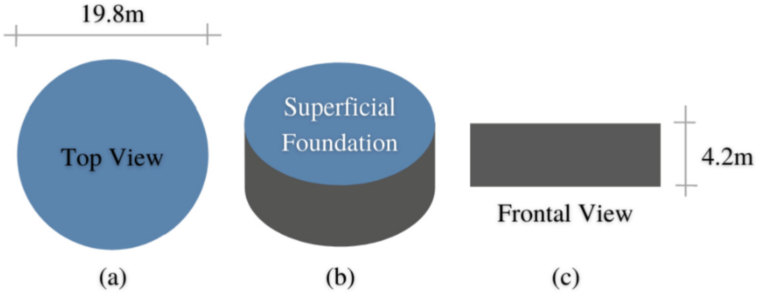


Fig. 7. Geometry of foundation: (a) top view; (b) perspective view and (c) frontal view.

The considerations related to the load take in to account a concentrated force resultant of the structure weight (vertical) and wind force (horizontal) eccentric with equivalent distance to the ratio of moment (M) to vertical force (W). This study only considers static loading, so that the dynamic effects needed to be accounted through reserve factors in the prediction of moment and wind force. The values adopted in this paper were based on data from Imanzadeh et al. (2021), with the moment (M) being equal to 109.6 MN.m, structure weight (W) equal to 17.3 MN and the wind force (Fwind) equal to 1.37 MN. The resultant load is applied in the top of foundation, disregarding the superstructure effects.

3.2 Displacements Under Foundation

The displacements under the foundation (cf. Fig. 8) were obtained considering the variation of input parameters (ϕ', E) per group (calculated from some of the correlations in Table 2 and 3, as explained in Sect. 3.1). To analyse the displacements, the −4.50 m of depth was fixed to evaluate the behaviour on horizontal position. To evaluate the behaviour as a function of the depth, four points were considered, as indicated in the Fig. 8. The points 1 and 4 indicates the end of the foundation, point 2 is the middle and the point 3 is where the load is applied.

To analyse the influence of the input parameter variation, it was considered four different groups: AA, AB, BA, BB The group AA is the group of reference, with the most adjusted results for ϕ' and E. The group AB varies the friction angle and BA varies the Young's modulus. Lastly, the group BB varies both parameters. Figure 8 shows the behaviour as a function of the depth.

As seen in Fig. 8, the variation of ϕ' (group AB compared with group AA and group BA compared with group BB) do not generate a significant difference in the curves.

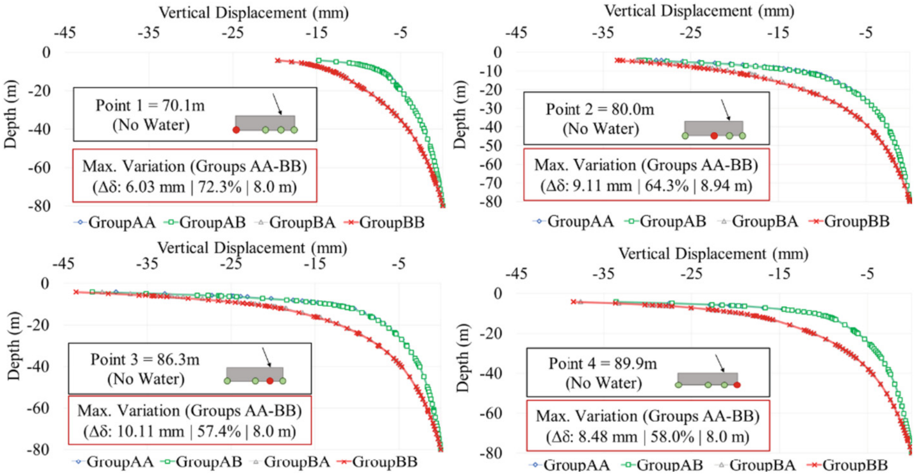


Fig. 8. Effect of input parameters variation in the displacement per point: (a) Point 1; (b) Point 2; (c) Point 3 and (d) Point 4.

The variation of E (group BA and group BB compared with group AA and group AB) generates a significant difference in the curves. Analysing the behaviour of vertical displacements, it is possible to see the highest values for point 3, followed by Points 4, 2 and 1. It happens because the load is applied in Point 3 and this affects the region of Points 4 and 2 after point 3. The Point 1 is the least affected because it is on the opposite side of the load application. The maximum variation of vertical displacements occurred at 8.00 m of depth, except to Point 2 (8.94 m of depth). The values vary from 6.03 mm to 10.11 mm when the Mohr-Coulomb Theory (LAMBE & WHITMAN, 1991) was utilized to predict foundation behaviour using PLAXIS 3D.

Observing the behaviour of the defined groups, two trends are evident, one for groups AA and AB and other for groups BA and BB. This means that the variation of friction angle does not affect the results while the Young's modulus causes a variation in the results.

The fact that the friction angle does not generate any significant variation in the results is associated with the fact that most of the analysed model behaved in elastic conditions, so that this parameter did not affect the generated displacements. For the same reason, the Young's modulus affects the results because this modulus exactly represents the change in soil elasticity.

In order to design the structure accurately and in favour of safety, it is necessary that a more complete geotechnical investigation campaign is carried out, so that these parameters can be obtained from laboratory tests and not by correlations. Errors can be on the order of 72%. In the case where only the SPT test is performed, the professional's local experience and specific methods for the region provide a direction, but they may not be enough to confirm the generated predictions, which highlights the importance of carrying out geotechnical tests.

4 Conclusions

This paper showed the importance of evaluating the input parameters for displacements results in wind farm foundations. To understand this, the SPT was utilized to provide friction angle and Young's modulus estimations through correlations methods with N_{SPT} number. The variation obtained by each correlation method reinforces the importance to use methods within the limitations in which it was developed, as a place of estimation and types of soils, as well as considering the local experience. Professional experience and literature records can guide estimates based on other case studies, but do not guarantee adequate accuracy.

The effect of the input parameters variation was evaluated, observing a discrepancy in the results for the Young's modulus variation, while there is no significant variation for the change in the friction angle. Errors in the estimation of E may cause prediction errors of the displacements and stresses of the order of 72%. The variation that occurred for E and not for ϕ' is expected due to the elastic condition of the soil in proper foundation designs. This fact demonstrates the importance of carrying out tests to obtain the parameters, mainly E, so that the input parameters have good accuracy and consequently give representative outputs.

The paper demonstrated that there is significant variation in geotechnical analysis that can occur with the use of different input correlations in geotechnical studies. These uncertainties lead either to overestimate or to underestimate the foundation design, and may affect economy and safety, thus emphasizing the need for more accurate field tests, to reduce them as much as possible.

References

ABEEÓLICA - ASSOCIAÇÃO BRASILEIRA DE ENERGIA EÓLICA. Energia Eólica: os Bons Ventos do Brasil. Infovento. Brasil, 15 June (2020). http://abeeolica.org.br/wp-content/uploads/2020/06/InfoventoPT_16.pdf. Accessed 2 Feb 2023

Afonso, A.F.G.: Correlações entre resultados de ensaios in situ de penetração dinâmico DP com o ensaio Standard Penetration Test. Instituto Politécnico de Bragança. Escola Superior de Tecnologia e Gestão (2016)

Bowles, J.E.: Foundation Analysis and Design, 5th edn., p. 624. The McGraw-Hill Companies, New York, USA (1996)

Cintra, J.C.A., Aoki, N., Tsuha, C.H.C., Giacheti, H.L.: Fundações: Ensaios estáticos e dinâmicos. Oficina de Textos, São Paulo (2013)

Cintra, J.C.A., Aoki, N., Albiero, J.H.: Fundações diretas: projeto geotécnico. Oficina de Textos, São Paulo (2011)

Correia, A.G.: Características de deformabilidade dos solos que interessam à funcionalidade das estruturas. Geotecnia **100**, 103–122 (2004). ISSN 0379-9522

Décourt, L.: The standard penetration tests. State-of-the-art-report. In: Proceedings of the XIIICSMFE, Rio de Janeiro, vol. 4, pp. 2405–2416 (1989)

De Freitas, A.C., Pachecho, M., Danziger, B.R.: Estimating young moduli in sands from the normalized N60 blow count. Soils Rocks, São Paulo **35**(1), 89–98 (2012)

De Mello, V.F.B.: The standard penetration test: state-of-the-art-report. In: 4th Pan-American Conference on Soil Mechanics and Foundation Engineering, Puerto Rico, vol. 1, pp. 1–86 (1971)

Godoy, N.S.: Fundações: Notas de aula, Curso de Graduação, São Carlos (SP): Escola de Engenharia de São Carlos – USP (1972)

Godoy, N.S.: Estimativa da capacidade de carga de estacas à partir de resultados de penetrômetro estático. Palestra. São Carlos (SP): Escola de Engenharia de São Carlos – USP (1983)

Gonin, H., Vandangeon, P., Lafeuillade, M.: Etude sur les corrélations entre le standard penetration test et le pressiometre. Rev. Fran. Geotech. **58**, 67–78 (1992)

Hatanaka, M., Uchida, A.: Empirical correlation between penetration resistance and effective friction of sandy soil. Soils Found. **36** (1996)

Imanzadeh, S., Pantet, A., Taibi, S., Ouahbi, T.: Soil behavior under onshore wind turbine foundation-case study in Brazil using new French EOLIFT technology. Adv. Civil Eng. Tech. **4**(4). ACET.000594 (2021)

Kulhawy, F.H., Mayne, P.W.: Estimating soil properties for foundation design. EPRI Report EL-6800, Electric Power Research Institute, USA, p. 306 (1990)

Lambe, T.W., Whitman, R.V.: Soil Mechanics. Series in Soil Engineering, Soil Engineering Series, p. 553 (1991). ISBN 0471022616, 9780471022619

Makwana, S., Gandhi, N.: Comparison between soil modulus based on standard penetration test and pressuremeter test- a case study of under ground Ahmedabad metro. In: Indian Geotechnical Society. IGS Proceedings (2019)

Meyerhof, G.G.: Discussion on research on determining the density of sands by spoon penetration. In: 4th International Conference on Soil Mechanics and Foundation Engineering, London (1959)

Mikhejev, V.V.: Foundation Design in the USSR. 5th edn. ICSMFE, pp. 753–757 (1961)

Milititsky, J.: Fundações de Aerogeradores - Desafios de Projeto e Execução. 9º Seminário de Engenharia de Fundações Especiais e Geotecnia 3ª Feira da Indústria de Fundações e Geotecnia SEFE 9 – 4 a 6 de junho (2019)

Moura, A.S., Cunha, R.P., Almeida, M.C.F.: Contribuição ao projeto de fundações superficiais de aerogeradores assentes nas areias de dunas do litoral cearense. Geotecnia n. º 130 – março 14, pp. 101–129 (2014)

Muromachi, T.: Experimental study on application of static cone penetrometer to subsurface investigation of weak cohesive soil. In: Proceedings European Conference on Penetration Testing, Stockolm, vol. 2 (1974)

Pinto, C.S.: Curso Básico de Mecânica dos Solos em 16 Aulas. 3° edição. Oficina de Textos, São Paulo (2006)

Silva, M.D.: Tipificação de fundações de torres eolicas em parques industriais, para diversos tipos de solos. 136 p. Dissertação (Mestrado em Engenharia Civil) – Instituto Superior de Engenharia de Lisboa (2014)

Teixeira, A.H.: Projeto e execução de fundações. Seminário de Engenharia de Fundações Especiais e Geotecnia, SEFE, São Paulo, vol. 1, pp. 33–50 (1996)

Teixeira, A.H., Godoy, N.S.: Análise, projeto e execução de fundações rasas. In: Hachich, W., et al. (eds.) Fundações: teoria e prática. 2 edn., pp. 227–264. Editora PINI, São Paulo (1996)

Trofimenkov, J.G.: Penetration testing in eastern Europe. In: Proceedings of the European Symposium on Penetration Testing, Stockholm, 5–7 June, vol. 2.1, pp. 24–28. Published by National Swedish Building Research (1974)

Wolff, T.F.: Pile capacity prediction using parameter functions **23**, 96–107 (1989)

Yoshida, I.: Empirical formulas of SPT blow counts for gravelly soils, 1st ISOPT, Orlando, United States (1988)

Uncertainties on the Unconfined Compressive Strength of Raw and Textured Concrete

Afshin Zarei[1], Samid Zarei[1], Saber Imanzadeh[2,3](✉), Nasre-dine Ahfir[3], Said Taibi[3], and Teddy Lesueur[4]

[1] Construction Company of Gohar Sang Gareh Dag, Tabriz, Iran
[2] Laboratoire de Mécanique de Normandie, Normandie Univ., INSA Rouen Normandie, 76801 Saint-Etienne du Rouvray, France
saber.imanzadeh@insa-rouen.fr
[3] Normandie Université, UNIHAVRE, Laboratoire Ondes et Milieux Complexes, CNRS UMR 6294, Le Havre, France
[4] GCCD Department, Normandie Université, UNIHAVRE, Institut Universitaire de Technologie (IUT), Le Havre, France

Abstract. For many years, starting in the 1960s and 1970s, concrete was known only as a utilitarian material providing mechanical strength and durability for new construction (civil works; bridges, tunnels, and administrative and residential buildings). Today, environmental considerations complement the design of structures. Aesthetics has been treated differently, it has been too often forgotten in the construction of large complexes for many years, to become the "unloved" grey material. Indeed, raw concrete surfaces tend to be porous and have a relatively uninteresting appearance. However, Auguste Perret, the French architect of reinforced concrete was the initiator of the use of concrete as a stone in the design of building facades. The present study concerns the uncertainties on the Unconfined Compressive Strength (UCS) of raw and textured concrete. For do this, the nine raw concrete samples and the nine texture concrete samples were prepared. Thereafter, the Unconfined Compressive Strength test was carried out to measure the Unconfined Compressive Strength values. The corresponding uncertainties are evaluated to quantify the uncertainties on the Unconfined Compressive Strength for raw and texture concrete samples. Finally, the compression of the Uncertainties on the Unconfined Compressive Strength of raw and textured concrete samples was done.

Keywords: raw and textured concrete · bush hammering · Unconfined Compressive Strength · Statistical errors · Uncertainties

1 Introduction

Despite very frequent use of reinforced concrete, urban constructions built with this material are sometimes unpopular, in particular by the gray and smooth aspect of the external surfaces. Working on its surface conditions is an alternative that allows to vary its appearance. The French Architect, Auguste Perret, (1874–1954) well known

J. E. S. De Cursi (Ed.): UQSM 2023, LNME, pp. 54–65, 2024.
https://doi.org/10.1007/978-3-031-47036-3_5

for its use of this material construction in urban architecture had a high regard for this construction material, did not hide it and treated it like stone. The reconstruction of the center of Le Havre (1945–1960) is a major project, and the last for him, now registered at UNESCO World Heritage Sites [1–3]. It is an architectural complex of 150 ha, which is designed and built in reinforced concrete with administrative and religious buildings (town hall, schools, churches…) but also shops and housing. With Perret, concrete presents various surfaces and colors on the same façade (Fig. 1). At that time, the additives were not very developed and the visible surfaces were either marked by the wooden forms, leaving a wood imprint, or bush hammered to expose aggregate concrete and thus to give texture and colour. Depending on the case, the material is simply coated to be protected, washed to bring out the grain of the gravel, bush-hammered to obtain an irregular surface. Bush-hammering is used on the stones worked tool to texturize stone observed on many monumental masonry constructions, as castles, churches, triumphal arch and other prestigious buildings.

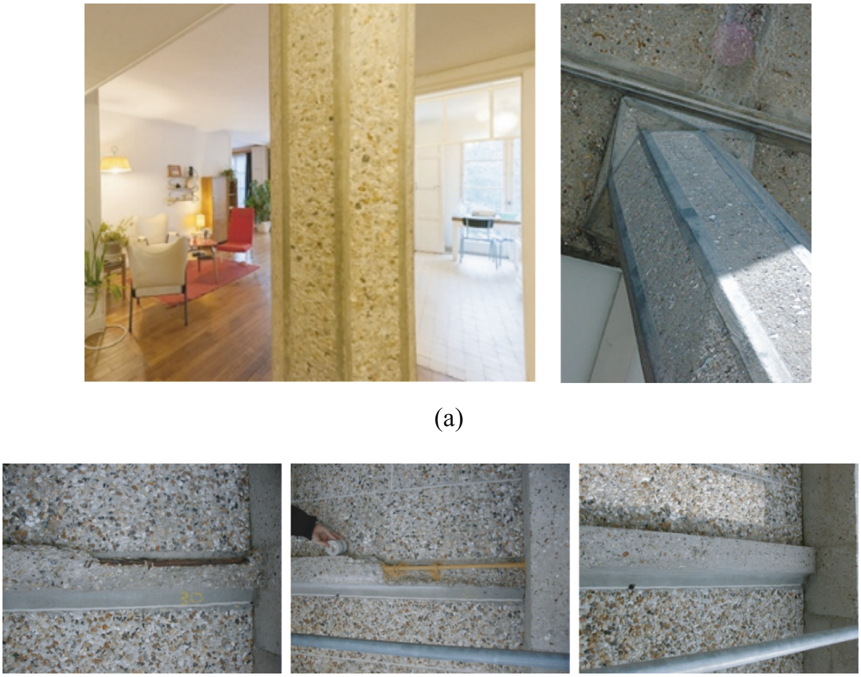
After, many architects followed this regard and developed architectural style like “Brutalism” which spread all over the world after the Second World War. It reached the culminating point in the 1960s and fade away in the late 1970s. One of the main features characterizing the brutalist trend in architecture was the exposure of building materials and their textures as well explained by Niebrzydowski, 2019, [4]. Bush-hammered concrete surfaces were introduced to brutalist architecture at the turn of the 1950s and 1960s, e.g. the town hall in Asker in Norway (1961–1963) designed by Kjell Lund and Nils Slaato.

Since the construction, the technical department of town planning of Le Havre are in charge of monitoring the evolution of buildings to preserve this heritage. If some local damages appear on exposed surfaces of the buildings housing in Le Havre of 70 years old, they are linked to corrosion effects. Regular and local repair work is carried out. It is wished to restore the same appearance (Fig. 1b). Even local damages, the structural integrity is kept like have shown many studies on Saint Joseph Church [5].

Today, for vertical surfaces, bush hammering is no longer done manually, but with an electric rollers bush hammering, it’s feasible to prepare medium to large surfaces. Figure 2 presents a textured concrete for vertical surfaces using the bush hammering performed by the construction company of Gohar Sang Gareh Dag at Tabriz city in Iran. However, the works inspection, to allow the technique, requires to verify that the repeated impact of the hammers has no effect on the structural qualities of the concrete.

Tooled finishes involve mechanically tooling or hammering the off-form finish to produce a rough texture. Common methods include bush hammering, point tooling, abrasive blasting and hammered-nib [6]. Courard 1998 [7] studied the geometric characterization of the surface of the concrete. In his study, it was established that the difference between sandblasting and polishing is more coming from the waviness than the roughness of the surface.

Some researchers have studied the influence of concrete-rock bonds and roughness on the shear behavior of concrete-rock interfaces. Badika et al., 2022 [8] recently performed the direct shear tests of three types of roughness: smooth interfaces, bush-hammered



(a)

(b)

Fig. 1. (a) details of treated surface with hammering technique (inside and outside) and (b) local repair to restore the appearance (Saint joseph church 2003 built in 1965)

interfaces, and natural granite interfaces under three levels of normal stress. It was concluded that bush-hammered interfaces generate stronger concrete-granite bonds compared to the naturally rough concrete-granite interface. Secondly the micro-roughness generated by the bush-hammering process does not significantly affect the friction angle nor captures the entire complexity of the natural roughness.

To our knowledge, a little research has been carried out on the unconfined compressive strength of textured concrete. However, verifying that the impact of the hammers has no effects on the structural qualities of the concrete is important. The present study concerns the uncertainties on the Unconfined Compressive Strength (UCS) of raw and textured concrete samples. This study was carried out by the construction company of Gohar Sang Gareh Dag with the collaboration of the two French laboratories: Laboratory of Mechanics of Normandy (INSA Rouen Normandie) and the Laboratory of Waves and Complex Media (University Le Havre Normandie). The corresponding uncertainties are evaluated experimentally to quantify the uncertainties on the Unconfined Compressive Strength for raw and textured concrete samples. Finally, the comparison of the uncertainties on the Unconfined Compressive Strength of raw and textured concrete samples was done.

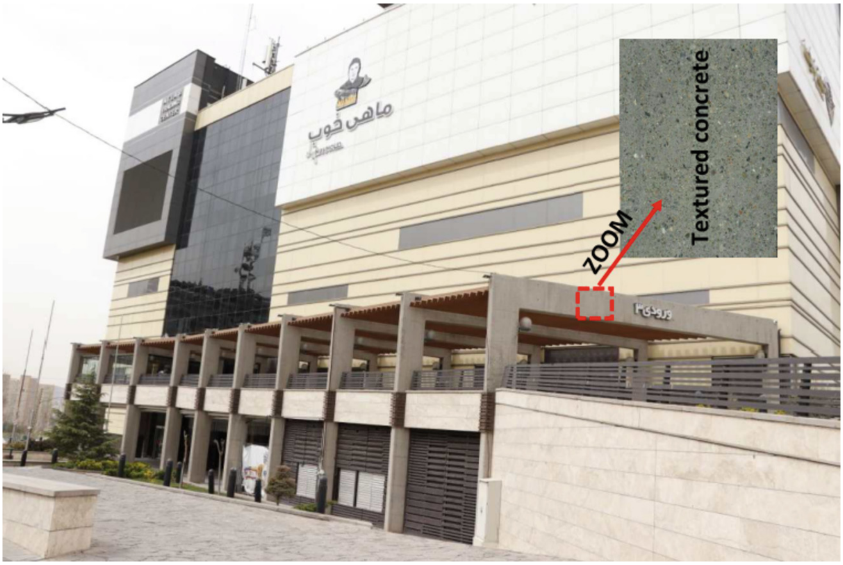


Fig. 2. Textured concrete performed by the construction company of Gohar Sang Gareh Dag (Tabriz, Iran)

2 Materials and Experimental Methods

2.1 Materials

The used concrete formulation was presented in the Table 1. The amount of the cement, water, gravel and sand are respectively 345, 216, 983 and 726 kg/m^3.

A high-performance cement for cold weather called the « Rapide ULTRACEM 52,5R (CEM I)» from the brand Calcia was used. The number 52,5 corresponds to the resistance within 28 days, and the letter R indicates a fast resistance from 2 days. This cement is reserved for highway structures and Civil Engineering works. Its reactivity reduces the effects of the cold and enables the development of short-term resistances. Moreover, it's very high mechanical resistances allow fast formworks removal. This type of cement is mainly suitable for armed or prestressed concrete and contains at least 95% of clinker and no more than 5% of minor components.

Table 1. Concrete formulation

Constituent	Kg/m^3
Cement	345
Water	216
Gravel	983
Sand	726

2.2 Experimental Methods

2.2.1 Sample Preparation

The preparation procedure was established with care in order to obtain homogeneous samples. Samples were prepared in a laboratory mixer. The experiment involves conducting various tests on samples of raw and textured concrete. If the tests are carried out in sufficient number (from 9 samples), it is possible to study the uncertainties on the UCS values [9]. For this purpose, the results of the nine samples of raw concrete and nine samples of textured concrete are analyzed. Thereafter, the comparison of the obtained results is performed.

The preparation of concrete is an essential step in building construction and infrastructure. To prepare concrete, it is necessary to mix aggregates such as sand and gravel, cement, and water, as well as other possible additives such as admixtures or pigments. Most concrete mixes are prepared using special machines called concrete mixers. Concrete mixers are equipped with a rotating drum that mixes the concrete ingredients until they are well blended. Concrete mixers can be manually or automatically fed, depending on the amount of concrete to be produced. In large constructions, concrete batching plants are often used, which are specialized facilities for producing large quantities of concrete efficiently and regularly. Once the concrete has been properly mixed (using a vibrating needle, for example), it is placed in cylindrical mold of 11 * 22 cm, ensuring that the mold is well filled and the surface is smooth and level. Then, the samples were stored for 90 days of curing-time in controlled laboratory environment. After allowing the sample to dry for 90 days of curing-time, it is carefully demolded and left to dry completely before cleaning the sample surface to remove any impurities as shown in Fig. 3. Once the sample has been cleaned, the Unconfined Compressive Strength test (UCS) can be carried out to evaluate the quality of the concrete. Proper preparation of concrete samples ensures accurate and reliable results during UCS test.

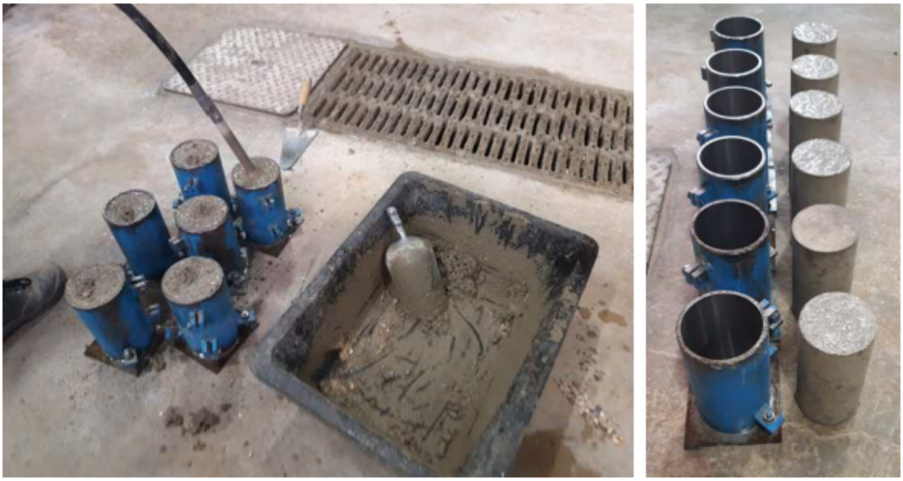


Fig. 3. Sample preparation

Figure 4 presents the prepared raw and textured concrete cylindrical samples using the bush hammering.

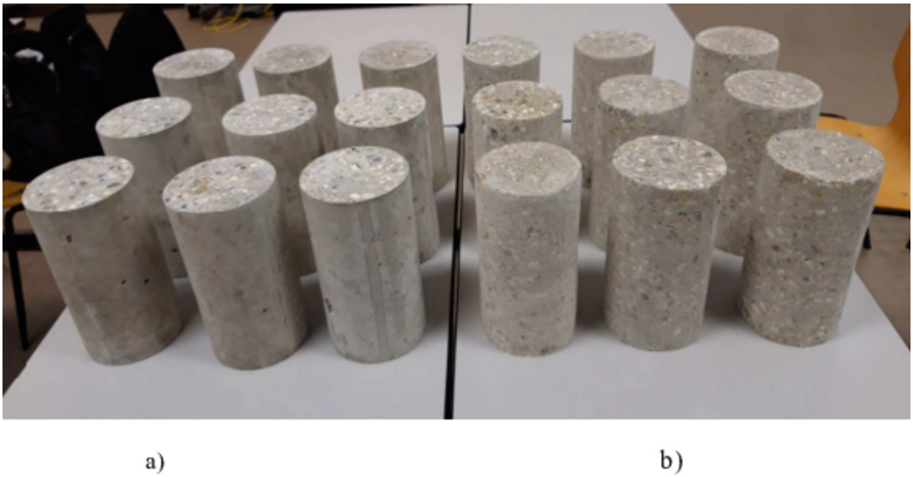


Fig. 4. a) raw concrete samples, b) textured concrete samples

2.2.2 Unconfined Compressive Strength Test (UCS)

The Unconfined Compressive Strength test is a laboratory test used to derive the Unconfirmed Compressive Strength (UCS) of a concrete sample. The Unconfirmed Compressive Strength (UCS) test stands for the maximum axial compressive stress that a sample can bear under zero confining stress. Due to the fact that stress is applied along the longitudinal axis, the Unconfined Compressive Strength test is also known as Uniaxial Compression test. Laboratory prepared samples were used to estimate unconfined compressive strength values.

During the test, from the axial load, axial displacement is commonly measured to measure the sample's UCS value. The Fig. 5 shows the apparatus used to carry out the Unconfined Compressive Strength test. The two plates shall be carefully cleaned before the sample is placed in the testing chamber. The load should be continuously applied at a rate of 5.7 kN/s. The samples were sheared on unconfined compressive strength path according to NF P94-420 and NF P94-425 French standards [10, 11]. Then, experimental stress and strain values are estimated for each sample. The experimental stress-strain curve was analyzed through its maximum value (UCS_{max}) which are presented in the following section.

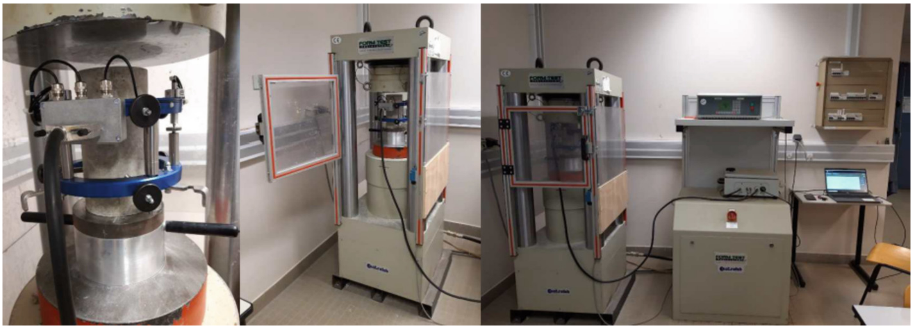


Fig. 5. Unconfined Compressive Strength test (UCS)

3 Results and Discussion

The UCS tests were carried out on the 9 raw concrete samples and the 9 textured concrete samples. The UCS test allows to plot the experimental unconfined compressive stress versus strain. Then, calculate the value of unconfined compressive strength for each sample. The Fig. 6 shows the experimental unconfined compressive stress versus strain for raw and textured concrete samples. For the raw concrete samples for the strain value of less than 0.1% the curves are very close but for the strain value more than 0.1% one can see a little dispersion in the curves. Furthermore, for the latter, the curves are nonlinear compared to the those with the strain values less than 0.1%.

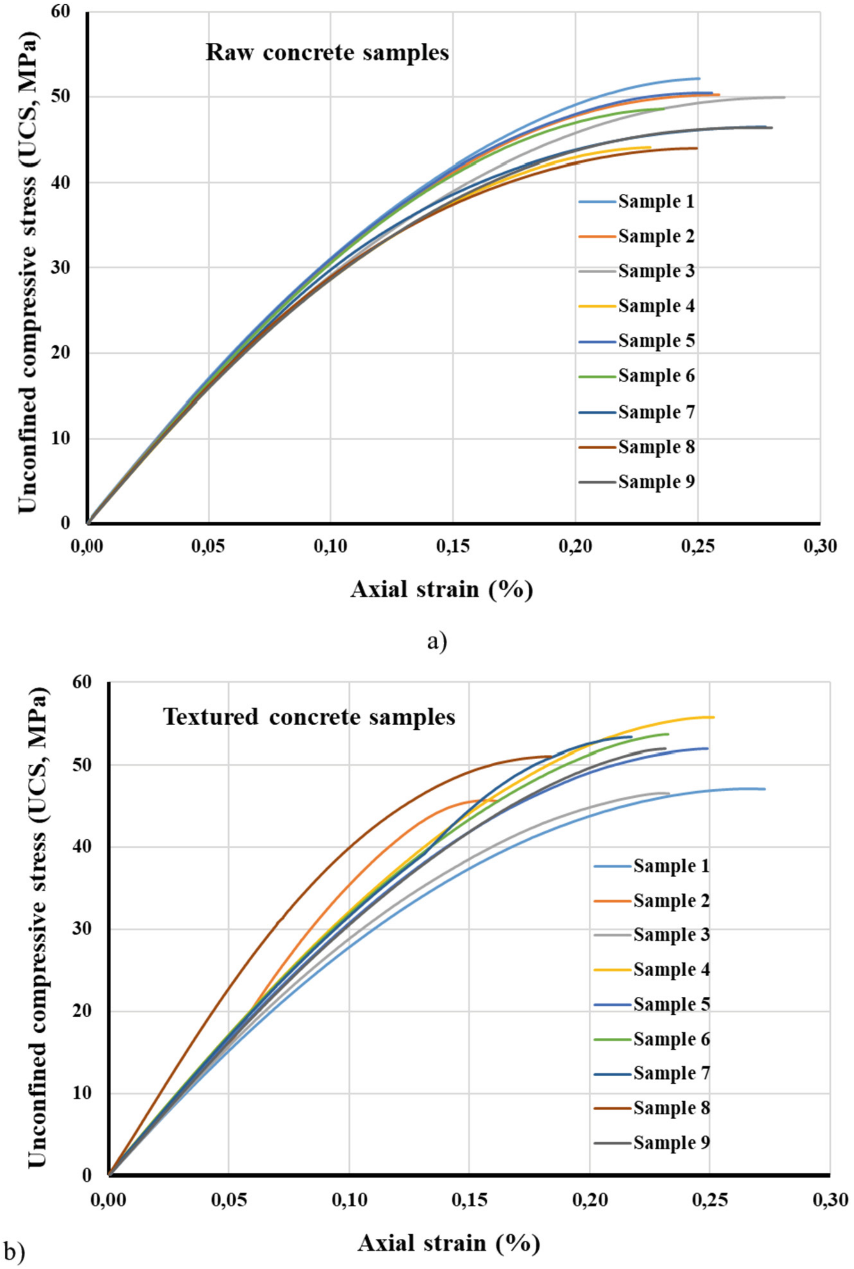
In the same way, for the textured concrete samples for the strain value of almost less than 0.07% the curves are very close except for the curve of sample 8. For the strain value more than 0.07% one can see the dispersion in the curves are important compared to those one from raw concrete samples for the value of the strain smaller than 0.07%. Furthermore, for the latter, the curves are nonlinear compared to the those with the strain values less than 0.07%.

The experimental stress-strain curves for textured concrete samples are more nonlinear than those ones from raw concrete samples. In the same way, the curves dispersions are more important for textured concrete samples compared to those ones from raw concrete samples.

The Table 2 presents the measured unconfined compressive strength values (UCS_{max}) for the both raw and textured concrete samples. The unconfined compressive strength values vary respectively from 43.96 to 52.16 MPa and from 45.65 to 55.75 MPa for the raw and textured concrete samples.

The unconfined compressive strength values versus sample number, for the raw and textured concrete samples, are presented in Fig. 7. One can observe that the dispersion of UCS_{max} values for the raw and textured concrete samples. Concerning the comparing the UCS_{max} values for the raw and textured concrete samples, the difference is very important for the sample 4 which is around 12 MPa. This difference for the sample 5 is equal to zero, while the average difference of UCS_{max} values for the other samples is around 5 MPa.

The different types of uncertainties can explain the dispersion and the differences in the UCS_{max} values. In fact, the uncertainties can usually be divided into two groups:



a)

b)

Fig. 6. Experimental stress-strain curves for a) raw concrete samples and b) textured concrete samples

aleatory or active uncertainty and epistemic or passive uncertainty [12, 13]. The first group is irreducible and due to the natural variability of random phenomena and it is

Table 2. Measured unconfined compressive strength values (UCS_{max}, MPa) for raw and textured concrete samples

Samples	UCS_{max} (raw concrete)	UCS_{max} (textured concrete)
1	52.16	47.12
2	50.23	45.65
3	49.95	46.59
4	44.02	55.74
5	50.42	50.42
6	48.56	53.63
7	46.46	53.30
8	43.96	51.08
9	46.36	51.92

intrinsic to the material. The second group is due to a lack of knowledge but can be reduced by obtaining additional information [14]. This epistemic or passive uncertainty is divided into three categories: measurement uncertainties, model uncertainties, and statistical uncertainties.

Even though the samples have the same concrete formulation and preparation method, concrete is a heterogeneous material. Therefore, one cannot have exactly the same quantity and distribution of aggregates, water, sand, and cement in all of prepared samples. This active uncertainty related to the material is therefore inevitable, and the knowledge of the engineer can only estimate it at best. On the other hand, epistemic or passive uncertainty can be reduced. Indeed, the experiment was carried out on nine raw and nine textured samples. One could reduce the statistical uncertainties related to the obtained results by increasing the number of studied samples. Furthermore, the measuring equipments have a degree of uncertainty regardless of precision and accuracy. These numerous uncertainties explain the differences in the UCS_{max} values.

The Table 3 presents for the raw and textured concrete samples, mean, standard deviation and coefficient of variation values. One can see that from the results the standard deviation and coefficient of variation of both raw and textured concrete samples are close and the standard deviation remains less than 4%, so there isn't any risk on the values. This also means that the data scatter is low and the precision of the UCS_{max} mean value is good. One can note that the textured concrete samples have almost close mean UCS_{max} values as the raw concrete samples. Indeed, on the 9 textured concrete samples, the mean UCS_{max} value is about 5% higher than raw concrete samples. Then, one can conclude that, the impact of the hammers has very small effects on the structural qualities of the concrete samples and it can be negligible.

Indeed, one can note that a clear effect on the stress-strain curves before the failure for the textured concrete samples which remains limited (Fig. 6b). In this study, the authors are interested in measuring the maximum Unconfined Compressive Stress values (UCS_{max}) which remains comparable for the raw and textured concrete samples. It would

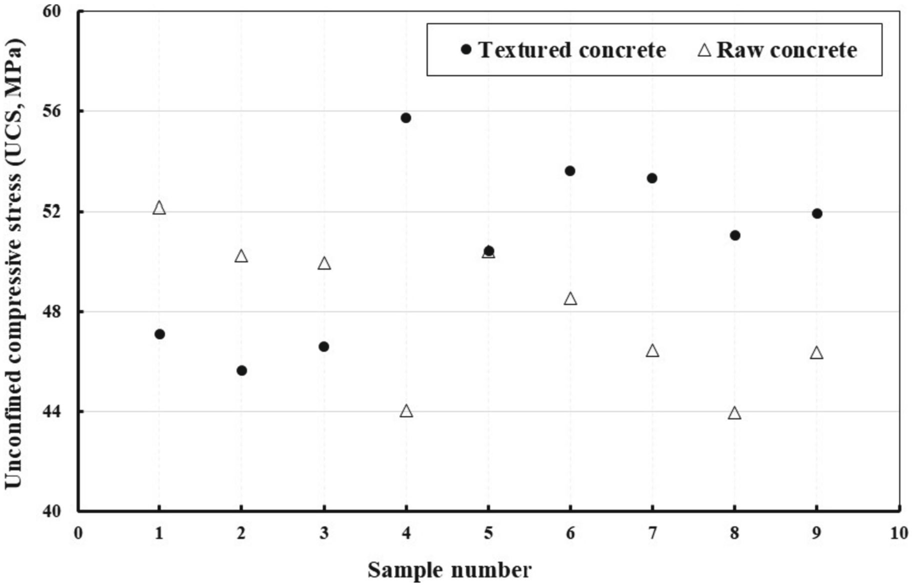


Fig. 7. Unconfined compressive strength values versus sample number, for the raw and textured concrete samples

Table 3. Statistical analysis of UCS_{max} values (MPa) for raw and textured concrete samples

Statistical parameters	UCS_{max} (raw concrete)	UCS_{max} (textured concrete)
Mean	48.01	50.61
Standard deviation	2.95	3.49
Coefficient of variation	6.14	6.90

be interesting to deepen this point on the behavior before failure by multiplying the number of tested samples not only by Unconfined Compressive Stress test but also with other characterization techniques.

4 Conclusions

In this research paper, it has been noted that the bush hammering used to valorize the aesthetic appearance of reinforced concrete, increased granular and color texture, has not effects on the structure integrity. The treated surfaces observed in Le Havre are 70 old year and local damages were identified since 20 years during restoration of Saint Joseph church and defects are quickly cured. However, the degraded zones can be easily repaired with the good formulation to respect the initial texture (grains and color). Indeed, during this restoration works, a chart of different formulations has been established. Nowadays, it is difficult to distinguish repaired zones from the other. Of course, the energy of the hammer must be controlled, just to peel off the superficial film after the formworks are removed and when defects are detected, it is necessary to cure effectively.

To complete the observations, two series of samples (raw and textured concrete) are compared on the base of the Unconfined Compressive Strength (UCS). If the tests are carried out in sufficient number (from 9 samples), it is possible to study the uncertainties on the UCS values. Thereafter, the stress-strain curves were plotted to calculate UCS_{max}. The statistical analysis was performed on UCS_{max} values. It was demonstrated that the mean, standard deviation and coefficient of variation of both raw and textured concrete samples are close and the standard deviation remains less than 4%, so there isn't any risk on the values. These results point out that, the impact of the hammers has very small effects on the structural qualities of the concrete and it can be negligible.

Acknowledgements. The authors would like to thank the professor Anne PANTET from University Le Havre Normandie for her careful reading of the paper and her constructive comments leading to improvements in this paper.

References

1. Pantet, A., Eleta-Defilippis R., Valtier, I., Chevé, M., Bonneau Contremoulins, V.: The reconstruction of Le Havre: cross-disciplinary view disciplinary. Fib, Fédération Internationale du Béton, Maastricht, juillet (2017)
2. Pantet, A., Eleta-Defilippis R., Valtier, I., Chevé, M.: Le chantier de la reconstruction de la ville du Havre Troisième Congrès Fran-cophone d'Histoire de la Construction, Nantes - Editions Picard, France (2017)
3. Pantet, A., Valtier, I., Eleta-Defilippis, R., Chevé, M., Bonneau- Contremoulins, V.: Le Havre, la première ville architecturée en béton armé. $36^{\text{ème}}$ Rencontres de l'AUGC, ENISE/LTDS, Saint Étienne, France, 19 au 22 juin 2018
4. Niebrzydowski, W.: From "as found" to bush-hammered concrete – material and texture in brutalist architecture. IOP Conf. Ser. Mater. Sci. Eng. **471**, 072016 (2019). https://doi.org/10.1088/1757-899X/471/7/072016
5. Kanéma, J.M., Pantet, A., Jamet, C.: Maintenance and repair of concrete structures of Le Havre, the city rebuilt by Perret, inscribed to the World Heritage list. Fib, Fédération Internationale du Béton, Maastricht, juillet 2017
6. Guide to Off-form Concrete Finishes, Cement Concrete & Aggregates Australia (CCAA T57) (2006). ISBN 1-877023-17-5
7. Courard, L.: Parametric definition of sandblasting and polished concrete surfaces, IXICPIC98-Bologna-Italy (1998)
8. Badika, M., El Merabi, B., Capdevielle, S., Dufour, F., Saletti, D., Briffaut, M.: Influence of concrete–rock bonds and roughness on the shear behavior of concrete–rock interfaces under low normal loading, experimental and numerical analysis. Appl. Sci. **12**, 5643 (2022). https://doi.org/10.3390/app12115643,2022
9. Ricotier, D., Canet, V.: Dimensionnement des structures en béton selon l'Eurocode 2, De la descente de charges aux plans de ferraillage, Le Moniteur (2021)
10. Standard AFNOR: NF P94-420, Détermination de la résistance à la compression uniaxiale (2000)
11. Standard AFNOR: NF P94-425, Méthodes d'essai pour roches - Détermination du module d'Young et du coefficient de Poisson (2002)

12. Lacasse S., Nadim F.: Uncertainties in characterizing soil properties in uncertainty in the geologic environment: from theory to practice. In: Shackleford, C.D., Nelson, P.P., Roth M.J.S. (eds.) ASCE Geotechnical Special Publication, no. 58, pp. 49–75. ASCE, New York (1996)
13. Uzielli, M., Nadim, F., Lacasse, S., Kaynia, A.M.: A conceptual framework for quantitative estimation of physical vulnerability to landslides. Eng. Geol. **102**, 251–256 (2008)
14. Sallak, M., Aguirre, F., Schon, W.: Incertitudes aléatoires et épistémiques, comment les distinguer et les manipuler dans les études de fiabilité? pp. 1–8 (2013)

Isogeometric Optimization of Structural Shapes for Robustness Based on Biomimetic Principles

Chunmei Liu(✉), Eduardo Souza de Cursi, and Renata Troian

Laboratory of Mechanics of Normandy (LMN), INSA Rouen, 685 Avenue de l'Université, Saint-Etienne-du-Rouvray, 76800 Normandy, France
chunmei.liu@insa-rouen.fr

Abstract. New challenges in shape optimization design under uncertainties lead to inspiration from nature. In this paper, we choose trees as the inspiration resource and apply the axiom of uniform strains, a governing principle of tree design, to avoid material overloading or underutilizing. The hypothesis of the uniform strains is formulated as the mean and standard deviation of strains which are defined as the optimization objectives. Then we use the isogeometric analysis (IGA) method to establish the numerical models. To take the geometric uncertainties into account, the coordinates of control points are defined as design variables. In the optimization process, the non-dominated sorting genetic algorithm II (NSGA-II) is applied to update design variables to figure out the optimal geometry. The Pareto front is obtained after iterative computation. The results based on bio-inspired criteria show that structural resistance can be increased significantly. This research provides new criteria for structural robust design under uncertainties.

Keywords: Bio-inspired · shape optimization · robust design · IGA · NSGA-II · uncertainties

1 Introduction

Shape optimization is a crucial component of many engineering applications, including aerospace, transportation vehicles, and architecture [1–4]. By optimizing the structural geometry, one can ensure that the design functions effectively, adapts to changes, and performs optimally under various deterministic parameters. However, in the real-world scenario, structures may face numerous uncertainties during production or service, such as environmental changes, material imperfections, equipment wear, and failure [5,6]. These uncertain factors can lead to unexpected changes or performance failures, potentially compromising the structure's normal operation. Effectively and accurately quantifying these uncertainties and incorporating them into the design procedures poses a significant challenge for researchers and engineers alike.

Numerous researchers have proposed various methods in the literature to address shape optimization design under uncertainties. These methods can

J. E. S. De Cursi (Ed.): UQSM 2023, LNME, pp. 66–77, 2024.
https://doi.org/10.1007/978-3-031-47036-3_6

be broadly categorized into two important categories: Robust Optimization Design [3,6–8] and Reliability-based Design Optimization [9–11]. Shape optimization design under uncertainties has found widespread applications in engineering. However, further development is necessary to minimize cost/benefit ratios and enhance competitiveness, particularly when dealing with complex structures.

Numerous scholars have made efforts to seek effective and simple methods of shape optimization design by turning to nature for inspiration. Natural structures such as bone, tooth, and tree exhibit lightweight and acceptable resistance, superior stiffness and toughness, allowing them to survive in diverse environments. In the work of Sun [12], a bio-inspired hull shape for autonomous underwater vehicles (AUVs) is proposed by studying and modeling the body shape of humpback whales. The paper aims to develop an innovative shape design for AUVs with minimal drag and energy consumption. However, they pointed out that the shape optimization has not been considered yet. In the paper [13], a new bearing inspired by a lamb's elbow is proposed. They have found the best clearance specification to reduce contact pressure through finite element simulations, and the optimal shape is obtained using the design of experiment method. In the article [14], a new bio-inspired method of shape optimization based on biological growth using ABAQUS was first proposed. Later, [15] pointed out that a novel method for shape optimization based on a tree's biological growth coupling boundary element method was simple and efficient. These biological structures self-optimize their shape using their unique approach to reduce stress peaks, withstand external loads, and produce a well-distributed strain map internally, leading to lightweight structures [16].

In this paper, we are focusing on the robust design optimization of the L-shape under geometric uncertainties inspired by the tree's biological growth. To describe the L-shape accurately we used the Non-Uniform Rational B-splines (NURBS) curve. To model and analyze the L-shape we applied the isogeometric analysis (IGA) method. The IGA method, first developed in 2005 [17], could implement geometric models and analysis models in the same framework. Owing to plenty of advantages of the IGA, it has been applied for shape optimizations with a relatively high computation speed [4,18–20].

Assumption that the material properties of the trunk are constant, the axiom uniform strain has been verified and applied it as a design rule in structural design [21–23]. However, they haven't given any mathematical formulation to express it. Therefore in this paper, we propose an innovative shape design optimization criteria based on the tree's biological growth called the axiom uniform strain considering the tree's material is various heterogeneous. In this paper, we focus on robust shape optimization with regard to geometrical uncertainties. A Latin Hypercube sampling feeds into the geometrical design variables. NSGA-II executes the optimization and carries out the Pareto front, which could give the optimal robust design solution.

The research has been organized in the following way. First, the bio-inspired design criteria is proposed. The robustness shape optimization design method is formulated by using bio-inspired design criteria based on the IGA method. Finally, the numerical example of the L-shape optimization design is presented to prove that it is an efficient method for practical application.

2 Biomimetics Design Criteria

The branch-trunk joint now widespread in nature and its shape has mostly been optimized during evolution. Müller et al. [24] studied the strain distribution by measuring a loaded branch-trunk joint with the 3D Electronic Speckle Pattern Interferometry (ESPI) experiences. In his research, he pointed out that the branch-trunk joint carried out a uniform distribution of strain by combining its natural structural shape, material properties, and fiber orientation. This example illustrates tree is an efficient structure that we could learn from. Mattheck proposed using the Soft-Kill-Option (SKO) for shape design could obtain an optimized shape achieving the axiom of uniform strain in his work [25].

Our inspiration is also from the structure of the branch-trunk joint. We propose using the axiom of uniform strain as the shape optimization design criteria. Thus in the process, the optimization considers minimizing both the mean and standard deviation of the strain as the criteria, especially in the weak area. It can be denoted ad Eq. (1):

$$\min \begin{cases} \bar{\mu} = \frac{\sum_{i=1}^{N} \varepsilon_{\text{von}}}{N}, \\ \bar{S} = \sqrt{\frac{\sum_{i=1}^{N} (\varepsilon_{\text{von}} - \bar{\mu})^2}{N}}. \end{cases} \tag{1}$$

ε_{von} is the strain of the i-element with $i = 1, 2, ..., N$. $\bar{\mu}$ is refers to the mean strain value of the optimization domain area and $\bar{S}$ is the standard deviation of strain.

3 Isogeometrical Model

The IGA method is suitable for shape optimization, as the geometric design model can be described precisely and can be modified easily without remeshing. It couples tightly design models and analysis models. In our study, the IGA method is under consideration. The detailed descriptions of establishing the geometric model and the analysis model can be found in references [17–20, 26]. The initial shape could be built using NURBS basis functions with giving the information, i.e. the position of the control point P_i and the weight of the control point ω_i, knot vector and so on. The most important parameters are P_i and ω_i, which is the essence of shape optimization with the IGA method. In our study, we just focus on changing the shape by moving the position of the control point.

4 Robust Optimization

Shape optimization problem considering the uncertainties based on bio-inspired design criteria using the IGA method consists of:

(1) Design variables. In our research, we mainly focus on the geometrical properties of the structure to reduce the high concentrated strain under uncertain conditions. After giving some specifications about the geometry of the structure with bounding conditions and the equations of calculations about the equivalent Von Mises strain ε_{von}. The design variables can be defined as $X = [X_1, X_2, ..., X_i]$ for adapting the coordinates of the control points in the horizontal direction and the vertical direction respectively to change the geometry.

(2) Objective functions. The axiom of uniform strain is employed as the bio-inspired criteria to reduce the high concentrated equivalent Von Mises strain for the purpose of improving the structural resistance in the weak area. The high concentrated equivalent Von Mises strains are to be minimized in the weak area, and thus the objective functions can be formulated as Eq. (2):

$$\min \begin{cases} \mu(\bar{\mu}) = \frac{\sum_{k=1}^{M} \bar{\mu}}{M}, \\ S(\bar{\mu}) = \sqrt{\frac{\sum_{k=1}^{M} (\bar{\mu} - \mu(\bar{\mu}))^2}{M}}, \\ \mu(\bar{S}) = \frac{\sum_{k=1}^{M} \bar{S}}{M}, \\ S(\bar{S}) = \sqrt{\frac{\sum_{k=1}^{M} (\bar{S} - \mu(\bar{S}))^2}{M}}, \end{cases} \tag{2}$$

where M is the number of the Latin hypercube samplings.

(3) Constraint conditions. Constraints are set on the optimization problem according to the specific engineering problem.

5 Robustness Problem

Two important characteristics that have a significant influence on the equivalent Von Mises strain consist of geometry and material properties. In this study, we only consider the uncertainties of geometry, considering that the output is commonly subject to such problems during the manufacturing process. Therefore, we apply normal distribution for geometry design parameters, i.e. the standard deviation σ around its mean value μ. In other words, it is expressed as a normal random variable $N(\mu, \sigma^2)$.

6 Optimization Method

To improve the accuracy and the efficiency of the robust optimization, the initial model is established to optimize the geometry shape by using the IGA method. Figure 1 illustrates the flowchart of the robust optimization problem. Firstly, input the geometrical data, material properties, and boundary conditions. Define the design variables and their values of the stochastic parameters during which the Latin hypercube sampling is adopted to obtain sampling points. Secondly, using the IGA method to analyze the structure model of each case was designed based on the samplings data, and the response values were obtained through numerical simulation. Then the bio-inspired criteria we proposed in Eq. (2) are applied as the objective function. Finally, robust optimization uses the non-dominated sorting genetic algorithm-II (NSGA-II) to select and complete the optimal design.

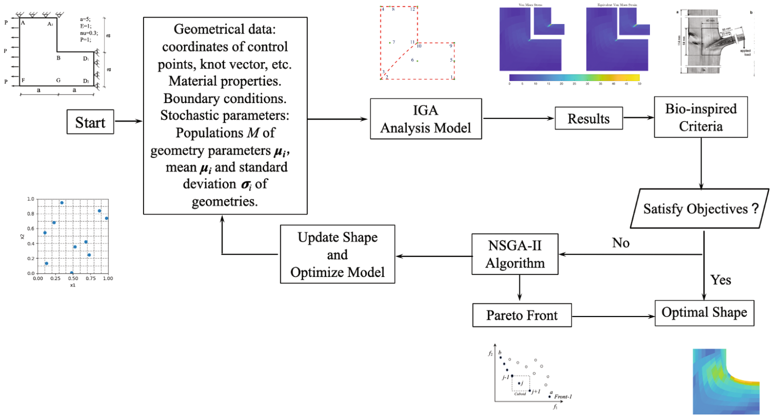

Start
Geometrical data: coordinates of control points, knot vector, etc. Material properties. Boundary conditions. Stochastic parameters: Populations M of geometry parameters μi, mean μi and standard deviation σi of geometries.
IGA Analysis Model
Results
Bio-inspired Criteria
Satisfy Objectives ?
No
Yes
NSGA-II Algorithm
Update Shape and Optimize Model
Pareto Front
Optimal Shape

Fig. 1. The flowchart of the robust optimization.

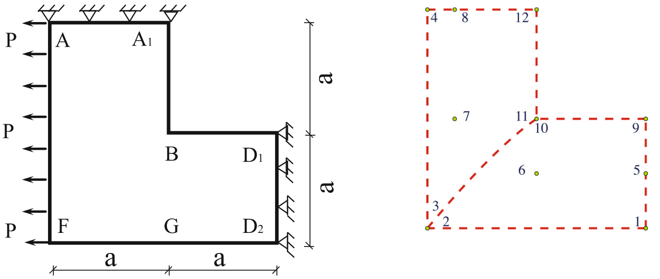
7 2D L-Shape Structure Example

7.1 Problem Statements

In this research, our optimization problem is related to the 2D L-shape structure design described in Fig. 3(a). It is a structure broadly used as a structural shape design to define the best solutions based on some specifications. For this case, the parameters Young's modulus $E = 1$, Poisson's ratio $\nu = 0.3$, $a = 5$, and the external force $P = 1$ proposed (Fig. 2).

7.2 Model Establishment

After specifying the geometry of the L-shape and the boundary conditions, employing NURBS surfaces contributing the 4 × 3 control point net to govern the whole design domain as shown in Fig. 3(b). The detailed coordinates and weight of the twelve control points in total are listed in Table 1. NURBS surfaces to ensure the C^1 continuity for all the control points and the coarsest mesh is defined by the open knot vectors $\Xi \times H$, where $\Xi = \{0, 0, 0, 0.5, 1, 1, 1\}$ and $H = \{0, 0, 0, 1, 1, 1\}$. The whole L-shape domain was discretized using h-refinement which enables elements to split uniformly around the corner.



(a) Geometry of the L-shape

(b) Control points of the L-shape

Fig. 2. 2D L-shape structure design.

Table 1. Control points coordinates for the initial shape of the L-shape domain.

Number	X	Y	ω[1]
1	2a [2]	0	1
2	0	0	1
3	0	0	1
4	0	2a	1
5	2a	0.5a	1
6	a	0.5a	1
7	0.25a	a	1
8	0.25a	2a	1
9	2a	a	1
10	a+X_1 [3]	a	1
11	a	A+X_2 [3]	1
12	a	2a	1

[1] ω is the weight.

[2] a is the fixed length of the horizontal/vertical.

[3] X_1 and X_2 are design variables to move specific control points 10 and 11 respectively.

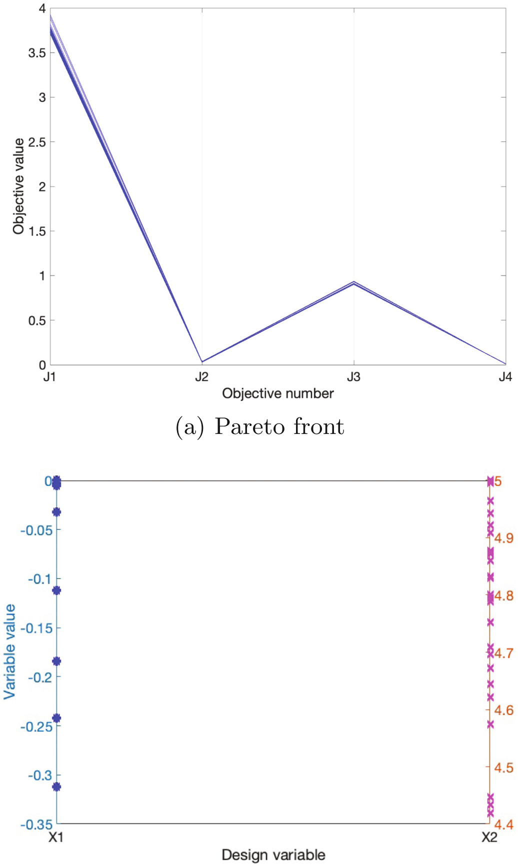
7.3 Optimization Problem

We focus on the geometrical properties of the L-shape corner B to reduce the high concentrated strain under uncertainties. The design variables can be defined as $X = [X_1, X_2]$ for adapting the coordinates of the control points 10 and 11 respectively to change the shape of the corner B. More specifically, X_1 enables horizontal direction movement of control point 10 with a range from -3 to 0 while X_2 implements vertical direction movement of control point 11 with a range from 0 to 5. Geometric uncertainties are expressed as normal random variables $X_1 \sim N_1(nominal_1, 0.1)$, $X_2 \sim N_2(nominal_2, 0.1)$, where $nominal_i$ is a value of design variables X_i. The total number $M = 100$ of the normal random samplings could be executed with the Latin hypercube procedure [27]. Therefore, the mathematical formulation of the L-shape optimization problem can be expressed as Eq. (3). The NSGA-II used as the optimization algorithm is executed together using the IGA analysis model, with a sleeted population size of 30, the maximum simulation generation number set at 100. It takes 19163.2 s to calculate in total with Macbook Pro 2021, Matlab version R2021b.

$$
\begin{aligned}
\min : & \begin{cases} J_1 : \mu(\bar{\mu}) = \frac{\sum_{k=1}^{M} \bar{\mu}}{M}, \\ J_2 : S(\bar{\mu}) = \sqrt{\frac{\sum_{k=1}^{M} (\bar{\mu} - \mu(\bar{\mu}))^2}{M}}, \\ J_3 : \mu(\bar{S}) = \frac{\sum_{k=1}^{M} \bar{S}}{M}, \\ J_4 : S(\bar{S}) = \sqrt{\frac{\sum_{k=1}^{M} (\bar{S} - \mu(\bar{S}))^2}{M}}, \end{cases} \\
\text{s.t} : & \begin{cases} -3 \leq X_1 \leq 0, \\ 0 \leq X_2 \leq 5. \end{cases}
\end{aligned} \tag{3}
$$

7.4 Results and Discussions

The results are presented in this section. Figure 3 shows the last generation results of the L-shape design optimization process executed by the NSGA-II algorithm directly. We can see in Fig. 3(a) that we had found the solutions (or Pareto sets) enable all objectives to converge. And these non-dominated solutions only appear for a few ranges as shown in Fig. 3(b). From the Pareto set data given in Appendix, most of the design variable X_1 is equal to 0 while the range of the design variable X_2 is from 4.447 to 5. Hence, several different solutions to the L-shape design optimization problem under the geometric uncertainties seem to have emerged which needs us to take care of making a decision.



(a) Pareto front

(b) Pareto set

Fig. 3. The results of Pareto front and set for the L-shape design.

One of the optimal L-shape for the non-dominated solution is presented in Fig. 4, where $X_1 = 0$ and $X_1 = 5$. We can see that the optimal shape of the L-shape design looks like the branch-trunk joint maintaining a uniform strain distribution even considering the uncertainties. The maximum equivalent Von Mises Strain is 7.90. Compared with the initial shape where the maximum equivalent Von Mises Strain reaches 45.69, it has been improved on a large scale.

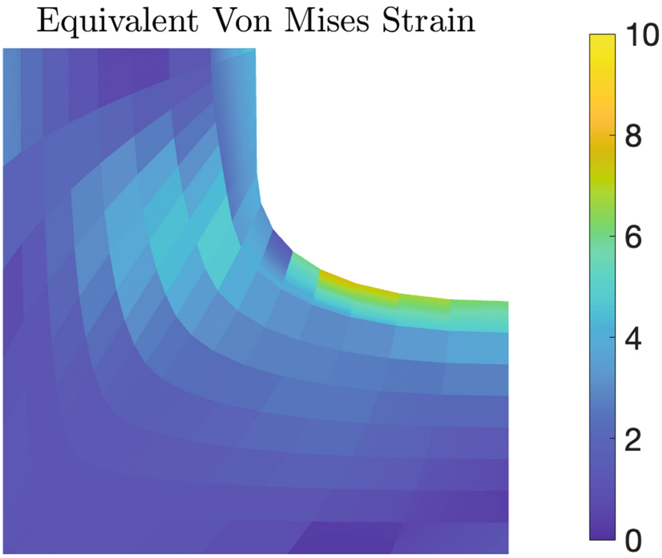


Fig. 4. The optimal shape of L-shape with $X_1 = 0$ and $X_1 = 5$.

8 Conclusions

The tree is our brilliant and experienced engineer, which has survived the merciless trials of evolution. This is the source of our inspiration. We are trying to find it out and apply it to instruct us in optimizing the structural shape design.

In our study, shape optimization considering uncertainties is going through an evolution from complex FEM-based optimization to IGA-based using a bio-inspired method inspired from the branch-trunk joint, which is hoped to be an efficient method for practical application.

Using the bio-inspired shape optimization for structural robust design based on the IGA method not only can be achieved in a relatively simple way but also improve the structural resistance. Thus, this paper provides a new structural design principle and a new method to do shape optimization considering the uncertainties.

Appendix

See Table 2.

Table 2. The non-domination solutions of the L-shape design optimization.

Number	X_1	X_2	J_1	J_2	J_3	J_4
1	0	5	3.71147	0.0377116	0.901097	0.0115308
2	−0.31244	4.43324	3.9284	0.0376476	0.938071	0.00788164
3	0	5	3.7108	0.0304324	0.901587	0.0133521
4	−0.0318365	5	3.71909	0.0273385	0.905746	0.0135637
5	0	4.87747	3.73778	0.0367894	0.899441	0.0107874
6	−0.18407	4.447	3.88236	0.0338068	0.929603	0.00816209
7	−0.111876	4.62095	3.82316	0.0319539	0.915325	0.00818323
8	−0.241927	4.41902	3.9073	0.0366396	0.934421	0.00805071
9	0	4.5738	3.80156	0.0295998	0.910463	0.00826809
10	0	4.64428	3.78736	0.032156	0.905971	0.00847748
11	−0.00550185	5	3.71261	0.0281521	0.902407	0.0134078
12	0	4.75196	3.76433	0.0318756	0.901469	0.0090477
13	0	4.70828	3.77362	0.0299529	0.903051	0.00881749
14	0	4.78827	3.75655	0.0351174	0.900379	0.00970629
15	0	4.83184	3.74729	0.0308973	0.900139	0.0102957
16	0	4.79461	3.7554	0.0333627	0.900513	0.00941333
17	0	4.96375	3.71896	0.0305739	0.900785	0.012412
18	0	4.90835	3.7313	0.0358567	0.899602	0.0110063
19	0	4.6719	3.78127	0.0302468	0.90453	0.0085032
20	−0.00330848	5	3.71193	0.0291382	0.902028	0.0132548
21	0	4.86015	3.7415	0.0286517	0.900238	0.0107393
22	0	4.99569	3.71241	0.0331029	0.901271	0.0127761
23	0	4.92196	3.72834	0.034686	0.899881	0.0111289
24	0	4.82914	3.74846	0.0368639	0.899723	0.0104729
25	0	4.69613	3.77598	0.0319809	0.903447	0.00863407
26	0	4.87257	3.73904	0.0313682	0.900083	0.0114176
27	0	5	3.71143	0.0327453	0.901293	0.0130035
28	0	4.80051	3.75406	0.0354783	0.900124	0.00995685
29	0	5	3.71156	0.034229	0.901313	0.0119798
30	0	4.94221	3.72377	0.0326062	0.900311	0.0119512

References

1. Kou, J., et al.: Aeroacoustic airfoil shape optimization enhanced by autoencoders. Exp. Syst. Appl. (2023). https://doi.org/10.1016/j.eswa.2023.119513
2. He, Z., Liu, T., Liu, H.: Improved particle swarm optimization algorithms for aerodynamic shape optimization of high-speed train. Adv. Eng. Softw. **173**, 103242 (2022). https://doi.org/10.1016/j.advengsoft.2022.103242
3. Yu, M., Liu, J., Huo, W., Zhang, J.: Shape optimization of the streamlined train head for reducing aerodynamic resistance and noise. Appl. Sci. **12**(19), 10146 (2022). https://doi.org/10.3390/app121910146
4. Hirschler, T., Bouclier, R., Duval, A., Elguedj, T., Morlier, J.: Isogeometric sizing and shape optimization of thin structures with a solid-shell approach. Struct. Multidiscip. Optim. **59**, 767–785 (2019). https://doi.org/10.1007/s00158-018-2100-6
5. Zang, T.A.: Needs and opportunities for uncertainty-based multidisciplinary design methods for aerospace vehicles. National Aeronautics and Space Administration, Langley Research Center (2002)
6. Moustapha, M., Galimshina, A., Habert, G., Sudret, B.: Multi-objective robust optimization using adaptive surrogate models for problems with mixed continuous-categorical parameters. Struct. Multidiscip. Optim. **65**(12), 1–22 (2022). https://doi.org/10.1007/s00158-022-03457-w
7. Troian, R., Shimoyama, K., Gillot, F., Besset, S.: Methodology for the design of the geometry of a cavity and its absorption coefficients as random design variables under vibroacoustic criteria. J. Comput. Acoust. **24**(02), 1650006 (2016). https://doi.org/10.1142/S0218396X16500065
8. Wang, R., Luo, Y.: Uncertainty-based comprehensive optimization design for the thermal protection system of hypersonic wing structure. Appl. Sci. **12**(21), 10734 (2022). https://doi.org/10.3390/app122110734
9. Allen, M., Maute, K.: Reliability-based shape optimization of structures undergoing fluid-structure interaction phenomena. Comput. Methods Appl. Mech. Eng. **194**(30–33), 3472–3495 (2005). https://doi.org/10.1016/j.cma.2004.12.028
10. Kim, D.W., Kwak, B.M.: Reliability-based shape optimization of two-dimensional elastic problems using BEM. Comput. Struct. **60**(5), 743–750 (1996). https://doi.org/10.1016/0045-7949(95)00433-5
11. Enevoldsen, I., Sørensen, J.D., Sigurdsson, G.: Reliability-based shape optimization using stochastic finite element methods. In: Der Kiureghian, A., Thoft-Christensen, P. (eds.) Reliability and Optimization of Structural Systems 1990: Proceedings of the 3rd IFIP WG 7.5 Conference Berkeley, California, USA, 26–28 March 1990, pp. 75–88. Springer, Heidelberg (1991). https://doi.org/10.1007/978-3-642-84362-4_8
12. Sun, T., et al.: Design and optimization of a bio-inspired hull shape for AUV by surrogate model technology. Eng. Appl. Comput. Fluid Mechanics **15**(1), 1057–1074 (2021). https://doi.org/10.1080/19942060.2021.1940287
13. Sysaykeo, D., Mermoz, E., Thouveny, T.: Clearance and design optimization of bio-inspired bearings under off-center load. CIRP Ann. **69**(1), 121–124 (2020). https://doi.org/10.1016/j.cirp.2020.03.006
14. Mattheck, C., Burkhardt, S.: A new method of structural shape optimization based on biological growth. Int. J. Fatigue **12**(3), 185–190 (1990). https://doi.org/10.1016/0142-1123(90)90094-U
15. Cai, R., Cai, S., Yang, X., Lu, F.: A novel method of structural shape optimization coupling BEM with an optimization method based on biological growth. Struct. Optim. **15**, 296–300 (1998). https://doi.org/10.1007/bf01203545

16. Mattheck, C.: Teacher tree: the evolution of notch shape optimization from complex to simple. Eng. Fract. Mech. **73**(12), 1732–1742 (2006). https://doi.org/10.1016/j.engfracmech.2006.02.007
17. Hughes, T.J., Cottrell, J.A., Bazilevs, Y.: Isogeometric analysis: CAD, finite elements, NURBS, exact geometry and mesh refinement. Comput. Methods Appl. Mech. Eng. **194**(39–41), 4135–4195 (2005). https://doi.org/10.1016/j.cma.2004.10.008
18. Wall, W.A., Frenzel, M.A., Cyron, C.: Isogeometric structural shape optimization. Comput. Methods Appl. Mech. Eng. **197**(33–40), 2976–2988 (2008). https://doi.org/10.1016/j.cma.2008.01.025
19. Hassani, B., Tavakkoli, S.M., Moghadam, N.: Application of isogeometric analysis in structural shape optimization. Scientia Iranica **18**(4), 846–852 (2011). https://doi.org/10.1016/j.scient.2011.07.014
20. Wang, Y., Wang, Z., Xia, Z., Poh, L.H.: Structural design optimization using isogeometric analysis: a comprehensive review. Comput. Model. Eng. Sci. **117**(3), 455–507 (2018). https://doi.org/10.31614/cmes.2018.04603
21. Dean, T., Long, J.N.: Validity of constant-stress and elastic-instability principles of stem formation in Pinus contorta and Trifolium pratense. Ann. Bot. **58**(6), 833–840 (1986). https://doi.org/10.1093/oxfordjournals.aob.a087265
22. Mattheck, C.: Engineering components grow like trees. Materialwiss. Werkstofftech. **21**(4), 143–168 (1990). https://doi.org/10.1002/mawe.19900210403
23. Mattheck, C., Bethge, K.: The structural optimization of trees. Naturwissenschaften **85**, 1–10 (1998). https://doi.org/10.1007/s001140050443
24. Müller, U., Gindl, W., Jeronimidis, G.: Biomechanics of a branch-stem junction in softwood. Trees **20**, 643–648 (2006). https://doi.org/10.1007/s00468-006-0079-x
25. Mattheck, C.: Design in Nature: Learning from Trees. Springer, Heidelberg (1998). https://doi.org/10.1007/978-3-642-58747-4
26. Nguyen, V.P., Anitescu, C., Bordas, S.P., Rabczuk, T.: Isogeometric analysis: an overview and computer implementation aspects. Math. Comput. Simul. **117**, 89–116 (2015). https://doi.org/10.1016/j.matcom.2015.05.008
27. Helton, J.C., Davis, F.J.: Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. Reliab. Eng. Syst. Saf. **81**(1), 23–69 (2003). https://doi.org/10.1016/S0951-8320(03)00058-9

Uncertainties About the Water Vapor Permeability of Raw Earth Building Material

Ichrak Hamrouni[1,2](✉), Habib Jalili[1], Tariq Ouahbi[1], Saïd Taibi[1], Mehrez Jamei[2], Hatem Zenzri[2], and Joanna Eid[3]

[1] Normandie Université, UNIHAVRE, LOMC, CNRS UMR 6294, Le Havre, France
ichrak.hamrouni@univ-lehavre.fr
[2] National Engineering School of Tunis, LGC, University Tunis Elmanar, Le Belvedere, 1002 Tunis, Tunisia
[3] AI Environnement, R&D Department, Fontenay-Sous-Bois, France

Abstract. In order to reduce the energy impact of building materials and, more generally, the environmental impact, raw earth can be an alternative to the conventional building materials like cement and fired earth bricks. Since earliest times, earthen building materials are becoming one of the most known construction techniques in the world thanks to its important benefits. This material has the capacity to play a significant role in regulating moisture and heat in buildings. It is distinguished by its low thermal conductivity, which renders it an effective thermal insulator. Additionally, it exhibits a remarkable ability to facilitate the diffusion of water vapor.

In this study, we are interested in the hygric characterization of raw earth material. Generally, in the experiment studies, one test is not representative and is not sufficient to analyze the results and the observed phenomenas. For this reason, a repetitive hygric experimental tests were carried out in this study and the results showing the uncertainties of the measurements will be presented and analyzed to determine the parameters that could cause these uncertainties.

Keywords: Raw earth · flax fibers · water vapor permeability · uncertainties

1 Introduction

The choice of an eco-material based on raw earth as a construction material must meet precise specifications concerning mechanical resistance, durability, rigidity, and performance in transfers by determination of hydro-mechanical properties. Several researches have been carried out on the hygrothermal properties of raw earth materials. These materials are capable of ensuring hygro-thermal comfort in buildings, due to breathability and insulation in terms of vapor and heat transfers respectively. In this study, we are interested in the hygric characterization of a composite material raw earth-flax fibers. Repetitive tests of measuring the water vapor permeability of raw earth alone, flax fibers alone and composite raw earth-flax fibers were carried out. The results showing the uncertainties of the measurements will be presented and analyzed to determine the parameters that could cause these uncertainties.

J. E. S. De Cursi (Ed.): UQSM 2023, LNME, pp. 78–85, 2024.
https://doi.org/10.1007/978-3-031-47036-3_7

2 Materials and Methods

2.1 Materials

In the current research, the studied materials are raw earth material and flax fibers. The raw earth is a naturally occurring silt sourced from the Normandy region in France. It is classified as a sandy-silt SL (SM) material according to the **LPC-USCS classification**, as indicated in Table 1. The flax fibers are locally extracted, then processed by cutting and drying before using for the experiment. These fibers are defined by their thickness, which ranges from 40 to 80 μm, a length of 7 cm, and a density of 1400 kg/m^3.

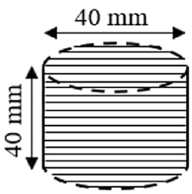
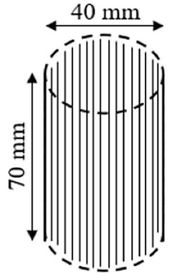
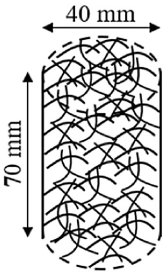
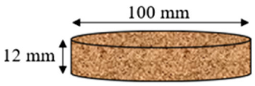
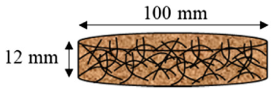
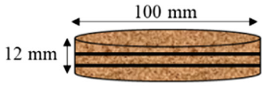
Table 1. Grading size table of silt GO [1]

Limon GO	
Teneur en fines (< 80 μm)	35%
Argile (<2 μm)	0%
Limon (de 2 μm à 60 μm)	25%
Sable (de 0,06 mm à 2 mm)	67%
Gravier (>2 mm)	8%
Limites d'Atterberg	
Limite de liquidité	$W_L = 22\%$
Indice de plasticité	IP $\approx$ 6
Classification LPC-USCS sable-limoneux SL (SM)	

The water vapor permeability of raw earth alone, fibers alone and the earth-fiber composite was measured using the cup method. The fibers were studied with different arrangements (longitudinal, transverse and random) and densities. The composite raw earth-flax fibers are prepared by compaction using a static press in a proctor mold.

The remarkable factors that were taken into account during the water vapour permeability tests were the density of the silt material with different percentages of mixed fibres, as well as the different arrangements of the latter (longitudinal, transverse and random) as shown in Table 2.

Table 2. Different cases studied for the measure of water vapor permeability [2]

		
Flax fibers in transversal arrangement	Flax fibers in longitudinal arrangement	Flax fibers in random arrangement
		
Raw earth alone	Raw earth-random flax Fibers	Raw earth-transversal flax fibers

2.2 Methods

To measure the water vapor permeability of fibers alone silt alone and composite raw-earth, two experimental setups have been developed in the laboratory based on the "Cup method" following to standard **ISO 12572:2001** [3] (Fig. 1 and Fig. 2).

Using the Fick law, we can determine the water vapor permeability δ_v [kg/ (m. s. Pa)] of the studied materials:

$$\boldsymbol{\delta_v = e\frac{j_v}{\Delta p_v} = e\frac{j_v}{p_{v2} - p_{v1}}} \tag{1}$$

e: sample's thickness [m]

j_v: water vapor flux density [kg/m^2. s].

pv: water vapor pressure [Pa].

Δp_v is the gradient of the water vapor pressure [Pa].

3 Results

3.1 Flax Fibers

In order to evaluate the anisotropy of water vapor permeability of flax fibers regard their orientation, we conducted tests involving the transfer of vapor through fiber bundles oriented in different directions: random, transverse, and longitudinal. Figure 3 illustrates the water vapor permeability evolution of flax fibers as a function of density for the three different orientations. The results show that the water vapor permeability decreases as fiber density increases, regardless of the fiber arrangement.

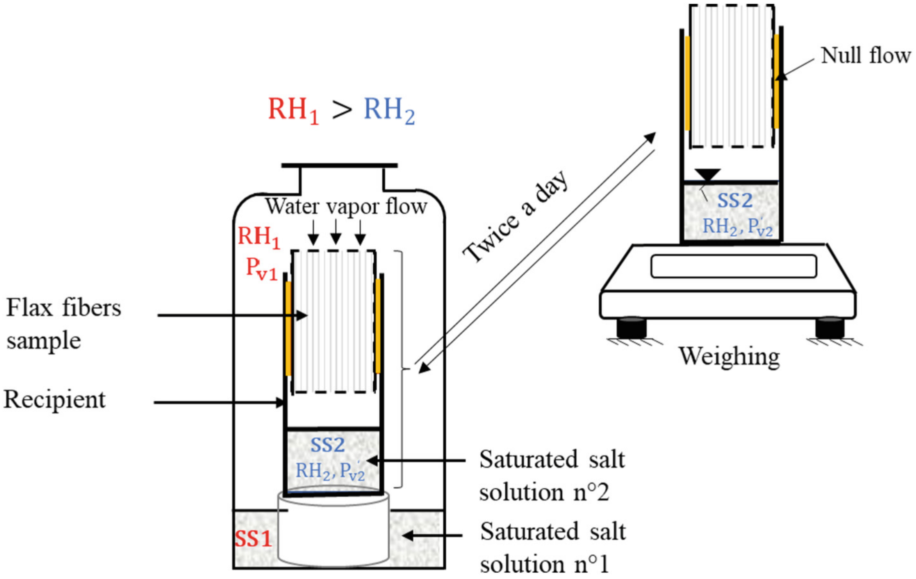


Fig. 1. Water vapor permeability measuring device for raw earth-flax fibers samples [2]

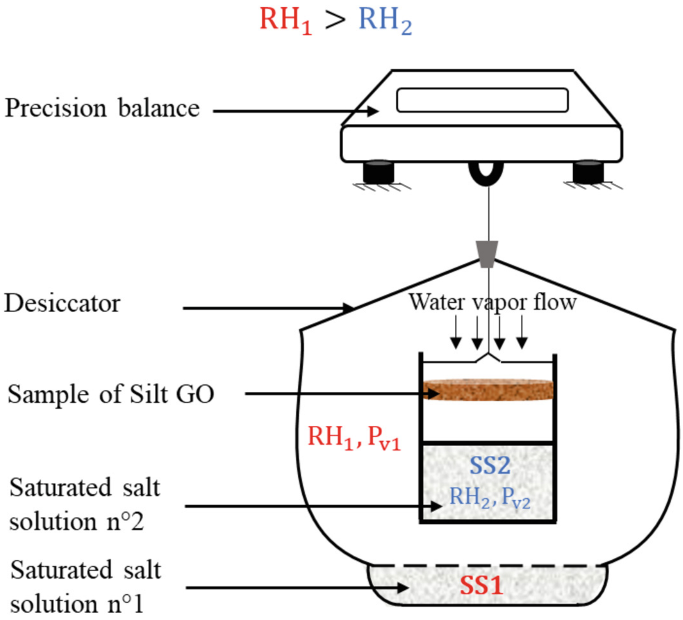


Fig. 2. Water vapor permeability measuring device for silt-flax fibers [2]

For longitudinal fibers, the permeability ranges from 1.11×10^{-10} to 9.93×10^{-11} kg/(m. s. Pa) when the density goes from 100 to 300 kg/m^3. In the case of transverse fibers, this decrease is more pronounced, with permeability decreasing from 6.83 × 10^{-11} to 4.78 × 10^{-11} kg/ (m. s. Pa). Additionally, our results demonstrate that fibers oriented in the same direction as the vapor flow exhibit higher permeability to water vapor.

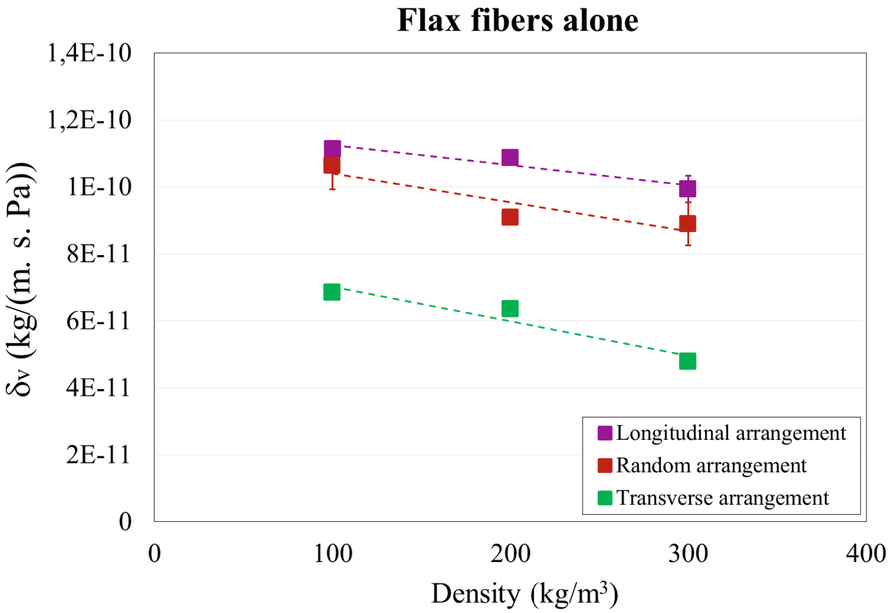


Fig. 3. Flax fibers water vapor permeability evolution as a function of their arrangement and densities [2]

3.2 Raw Earth-Flax Fibers

Figure 4 illustrates the water vapor permeability evolution as a function of the percentage of added fibers of the composite material raw earth-flax fibers with randomly and transversely arrangements. A significant increase in vapor permeability with the increasing percentage of fibers, regardless of their orientation was observed.

A particularly significant change occurs when 1% of fibers is added to the soil, compared to the soil alone without fibers. At this point, the permeability increases from 9.38 × 10^{-12} to 1.10 × 10^{-11} kg/ (m. s. Pa) in the case of transversely oriented fibers and to 1.30 × 10^{-11} kg/ (m. s. Pa) with randomly oriented fibers. Then, this growth softens with the increase of percentage of fibers from 1% to 5% with a slope of 2.65 × 10^{-13} kg/ (m. s. Pa) for the two types of fibers' arrangement.

Furthermore, for a given percentage of fibers, the permeability of silt GO reinforced with a random arrangement of fibers is higher than that with a transverse arrangement. These results align with the those observed in Fig. 3, where the water vapor permeability obtained with a random distribution of fibers alone was higher than that in the case of a transverse arrangement.

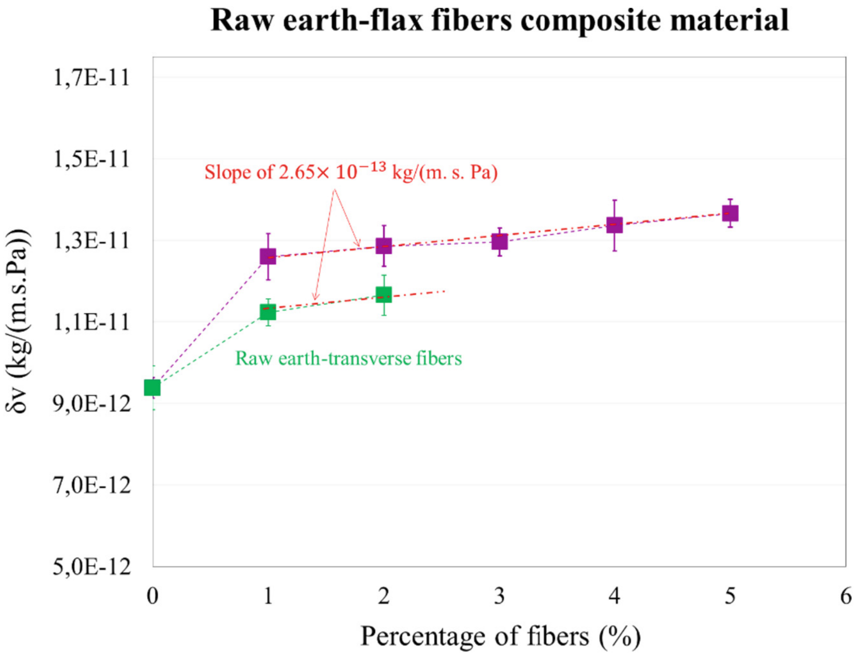


Fig. 4. Raw earth -flax fibers water vapor permeability evolution as a function of fibers' percentage [2]

4 Uncertainties

For experimental studies, one single test is not representative and not sufficient to analyze the results and the observed phenomena. For this reason, in this study, repetitive experimental tests of permeability measurement were realized. The results showed uncertainties of measurements which were calculated as follows:

Standard Deviation (STD):

$$STD = \sqrt{\frac{\sum_{\mathrm{i}=1}^{N}\left(\delta_{vi} - \overline{\delta_{vi}}\right)^2}{\mathrm{N}}} \tag{2}$$

δ_{vi}: water vapor permeability for the test i [kg/ (m. s. Pa)].
$\overline{\delta_{vi}}$: mean value of water vapor permeability [kg/ (m. s. Pa)].
N: number of tests.

Coefficient of Variation (CV):

$$CV = \frac{STD}{\overline{\delta_{vi}}} \times 100 \tag{3}$$

4.1 Flax Fibers

The uncertainties obtained for the case of the tests for the flax fibers are illustrated in the Table 3. We can observe from the results that the maximum CV is equal to 7. According to Costa et al., 2002 [4], if the CV is lower than 7%, that's show a high experimental precision. So, we can conclude from the results of the water vapor permeability of flax fibers in different arrangements show acceptable uncertainties.

Table 3. Uncertainties about the water vapor permeability of flax fibers in different arrangements

	Density	$\overline{\delta_{vi}}$ [kg/ (m. s. Pa)]	*STD* [kg/ (m. s. Pa)]	*CV* [%]
Longitudinal Arrangement	100	1.11. 10^{-10}	1.2. 10^{-12}	1.08
	200	1.09. 10^{-10}	3.54. 10^{-13}	0.33
	300	9.93. 10^{-11}	4.03. 10^{-12}	4.06
Random Arrangement	100	1.06. 10^{-10}	7.02. 10^{-12}	6.6
	200	9.08. 10^{-11}	1,31. 10^{-12}	1.45
	300	8.89. 10^{-11}	6.49. 10^{-12}	7
Transversal Arrangement	100	6.83. 10^{-11}	3.13. 10^{-13}	4.59
	200	6.34. 10^{-11}	1.58. 10^{-12}	2.48
	300	4.78. 10^{-11}	2.57. 10^{-13}	5.37

4.2 Raw Earth - Flax Fibers

The uncertainties obtained for the case of the tests for the raw earth-flax fibers are illustrated in the Table 4. We can observe that the CV for these experiments does not exceed 5% so according to Costa et al., 2002 would be classified as of high precision.

Table 4. Uncertainties about the water vapor permeability of raw earth – flax fibers composites

	Percentage of fibers	$\overline{\delta_{vi}}$ [kg/ (m. s. Pa)]	*STD* [kg/ (m. s. Pa)]	*CV* [%]
Raw earth – random flax fibers	0	9.38. 10^{-12}	2.55.10^{-13}	2.71
	1	1.26. 10^{-11}	5.7.10^{-13}	4.53
	2	1.29. 10^{-11}	5.10^{-13}	3.89
	3	1.30. 10^{-11}	3.44.10^{-13}	2.65
	4	1.34. 10^{-11}	6.24.10^{-13}	4.67
	5	1.37. 10^{-11}	3.42.10^{-13}	2.51
Raw earth – transversal flax fibers	1	1.12. 10^{-11}	3.26. 10^{-13}	2.9
	2	1.17. 10^{-11}	4.91. 10^{-13}	4.22

5 Conclusion

This paper focus on the hygric characterization of a raw earth-fibers composite material. The water vapor permeability of a natural silt, flax fibers and silt reinforced with flax fibers was determined using the "cup method". The results highlight the effect of fibers' orientation on their ability to let pass water vapor, more the fibers are oriented in the same direction as the vapor flow, more they are permeable to water vapor. Besides, the addition of flax fibers to the raw earth material improves the water vapor permeability of the raw earth from 9.38. 10^{-12} to 1.37. 10^{-11} kg/ (m. s. Pa) in the case of random fibers and from 9.38. 10^{-12} to 1.17. 10^{-11} kg/ (m. s. Pa) in the case of transversal fibers.

As in this study, repetitive tests of measuring water vapor permeability of different materials were carried out, the results show uncertainties. The maximum observed coefficient of variation (CV) in the case of fibers alone is about 7% and about 5% in the case of composite raw earth-fibers. These coefficients show a high precision of the experimental tests.

References

1. Hibouche, A.: Sols traités aux liants Performances hydro-mécaniques et hygro-thermiques Applications en BTP [Soils treated with binders Hydro-mechanical and hygro-thermal performance Construction applications]. Ph.D. thesis. University of Le Havre (2013)
2. Hamrouni, I., Ouahbi, T., El Hajjar, A., Taibi, S., Jamei, M., Zenzri, H.: Water vapor permeability of flax fibers reinforced raw earth: experimental and micro-macro modeling. Eur. J. Environ. Civil Eng. (2022). https://doi.org/10.1080/19648189.2022.2123857
3. ISO 12572-27: Performance hygrothermique des matériaux et produits pour le bâtiment-Détermination des propriétés de transmission de la vapeur d'eau (2001)
4. Costa, N.H.D.A.D., Seraphin, J-C., Zimmermann, F-J-P.: A new method of variation coefficient classification for upland rice crop. Pesquisa Agropecuária Brasileira, Brasília **37**(3), 243–249 (2002). https://doi.org/10.1590/S0100-204X2002000300003

Dynamic analysis of a building equipped with a Tuned Mass Damper subjected to artificial seismic excitations considering uncertainties in the parameters of the structure and of the excitation

João Victor Restelatto(✉), Letícia Fleck Fadel Miguel,
and Sergio Pastor Ontiveros-Pérez

Department of Mechanical Engineering, Federal University of Rio Grande Do Sul,
Porto Alegre, Brazil
joao.restelatto@hotmail.com, letffm@ufrgs.br,
sergio.ontiveros.perez@outlook.com

Abstract. The present work aims to analyze the effectiveness of a passive vibration control device in a structure subjected to random vibrations. The structure is a ten-story building equipped with a Tuned Mass Damper (TMD) at the top and it is subjected to artificial seismic excitations generated by the Kanai-Tajimi spectrum. The uncertainties present in both the systems and excitation parameters are taken into account. Thus, mass, stiffness and damping of the structure and the TMD, as well as peak ground acceleration (PGA), ground frequency and ground damping ratio are assumed as random variables, and the problem is solved via Monte Carlo Simulation. The study uses Newmark's numerical integration method to obtain the results of displacement, velocity, acceleration and maximum interstory drift values of the structure. The results obtained during the study demonstrate that the variance decreased and the dynamic response of the structure in terms of interstory drift is considerably reduced by about 55% after installing the TMD at the top of the building.

Keywords: Seismic Excitation · Tuned Mass Damper · Vibration Control

1 Introduction

The installation of vibration control mechanisms in structures subjected to the effects of dynamic forces aims to reduce the magnitudes of displacement and interstory drift, providing a greater level of comfort and safety to users. These devices have different varieties and operating principles, with their specific characteristics and their application is determined by the type of structure designed and the possible events it will be subjected to [1].

These devices can be classified into three types, the active ones that need an external energy power to be activated, the semi-actives that require a small source of external

J. E. S. De Cursi (Ed.): UQSM 2023, LNME, pp. 86–96, 2024.
https://doi.org/10.1007/978-3-031-47036-3_8

energy for operation and use the movement of the structure to develop control forces, and the passive ones that stand out for consuming a minimum amount of power compared to the others [2]. The operation of passive mechanisms consists of energy dissipation using the oscillation of the main structure, with a secondary system that reduces the unwanted kinetic energy of the system, some examples of passive systems are friction dampers, viscoelastic dampers and tuned mass dampers [3].

The device model implanted in this work for vibration control is the tuned mass damper (TMD), whose operating principle consists of a mass, a spring and a viscous damper system coupled to the main structure [4]. This mechanism has a positive point which is the low waste of energy as it is a passive system, being a cheap and effective device. The TMD is dimensioned mainly to tune with the first vibration mode of the structure [5]. In this work, the TMD is able to reduce the vibrations of the building against seismic events.

Second [6], one of the most famous applications of the tuned mass damper is in the Taipei 101 skyscraper, in Taiwan, whose implementation is a milestone of modern engineering in the field of mechanical vibrations, serving as a basis for studies of new scientific projects. The application of this damping system in the building took into account that the city of Taipei is often hit by earthquakes and the TMD contributes to reducing the recurring vibrations of these unwanted events.

It is of great importance to note that the uncertainties in the structural properties of mass, stiffness and damping together with the parameters of the random dynamic excitation that take into account the peak ground acceleration, ground frequency and ground damping can change the optimal solution of the stipulated problem. Thus, it is extremely important to take into account uncertainties in the procedure in order to avoid damage to the structure, reduced performance and failures [7].

Within this context, the main objective of the present work is to propose a methodology to verify the effectiveness of installing a tuned mass damper at the top of a ten-story structure subject to random earthquakes, considering uncertainties and with the aim of reducing the maximum interstory drift of the building.

2 Methodology

In this chapter, the main concepts necessary for carrying out the case study are discussed. Therefore, as the focus is on the area of mechanical vibrations, the analysis of equations of motion as well as basic notions of vibrating elements are essential foundations for the work. Other factors such as methods of numerical integration and analysis of seismic events are also theorized throughout the chapter. In addition, in the area of probability and statistics, the input of random variables, among other fundamentals are necessary concepts for the work.

2.1 Equation of Motion

According to [8], the equation of motion for a vibrating system with the action of an external force and with n degrees of freedom (DOF) is formulated to provide, from numerical calculations, the amplitudes of displacement, velocity and acceleration of the system. To obtain the response of the equation of motion in the case of a structure subjected to an earthquake, it must be considered that the force vector is a base acceleration. The description of this formula is shown by Eq. (1).

$$[M]\ddot{x}(t) + [C]\dot{x}(t) + [K]x(t) = -[M]\ddot{x}_g(t) \tag{1}$$

where [M] is the mass matrix, [C] is the damping matrix and [K] is the stiffness matrix, all with dimensions (*n* x *n*). The term ẍ(t) represents the acceleration vector as a function of time, ẋ(t) represents the velocity vector as a function of time, x(t) represents the displacement vector as a function of time and ẍg(t) is the acceleration vector of the soil as a function of time, generated by the Kanai-Tajimi spectrum, which multiplied by the matrix [M] gives the forces of the seismic event.

In the numerical resolution of a shear building system considering the vectors in the transverse orientation of the structure, with *n* degrees of freedom, there is a need to formulate matrices containing the vibratory parameters of the structure for the calculation. The sizing of the matrices follows a pattern linked to the number of DOFs in the structure, forming square matrices of dimensions (*n* x *n*) [9]. The mass matrix, stiffness matrix and damping matrix are cited below (see Eqs. 2, 3 and 4).

$$[M] = \begin{bmatrix} m_1 & 0 & \cdots & 0 \\ 0 & m_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & m_n \end{bmatrix} \tag{2}$$

where [M] is the mass matrix, m_1 is mass 1, m_2 is mass 2, and m_n is the mass of the last inertia element of the system.

$$[K] = \begin{bmatrix} k_1 + k_2 & -k_2 & \cdots & 0 & 0 \\ -k_2 & k_2 + k_3 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & k_{n-1} + k_n & -k_n \\ 0 & 0 & \cdots & -k_n & k_n \end{bmatrix} \tag{3}$$

where [K] is the stiffness matrix, k_1 is the spring 1, k_2 is the spring 2, and k_n is the stiffness of the last element of the system.

$$[C] = \begin{bmatrix} c_1 + c_2 & -c_2 & \cdots & 0 & 0 \\ -c_2 & c_2 + c_3 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & c_{n-1} + c_n & -c_n \\ 0 & 0 & \cdots & -c_n & c_n \end{bmatrix} \tag{4}$$

where [C] is the damping matrix, c_1 is the damping 1, c_2 is the damping 2, and c_n is the damping of the last element of the system.

To solve Eq. (1), a program developed in Matlab based on Newmark's method of numerical integration was used, which is an implicit method used to solve displacement, velocity and acceleration equations in the time domain [10].

2.2 Random Parameters of the Structure

In this work, the parametric probabilistic approach was used to obtain the uncertainties of the system parameters, in order to take into account, the uncertainties in the mass, stiffness and damping parameters of the main structure and the TMD installed at the top. These three stochastic variables are modeled as uncorrelated random variables with Lognormal distribution, as they cannot assume negative values, due to physical aspects, with known mean and coefficient of variation. Thus, in each run of the routine, the structure presents different parameters. Since the system's response depends on these random variables, it becomes random [11].

The structure to be studied was defined as a ten-story building equipped with a single tuned mass damper at the top. The building is simplified for the case study, simulating it as a shear building with ten tranversal DOFs, the mass is concentrated in each slab and the columns are springs of a certain stiffness, with viscous damping, the demonstration of this structure with the TMD installed is illustrated in Fig. 1:

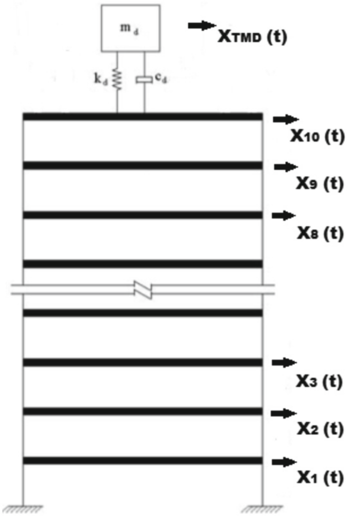


Fig. 1. Ten-degree-of-freedom building with a single tuned mass damper at the top, (adapted from [15]).

The work was analyzed for the case situation where a single TMD is installed at the top of the structure, considering it has a mass of 3% of the total mass of the building. The mass, stiffness and damping parameters for each of the ten floors of the structure, with the TMD parameters and their respective mean values and coefficient of variation of 5%, representing the uncertainties of the application, are described in Table 1.

Considering the mean values of the parameters in Table 1 and assuming coefficients of variation equal to zero for the random variables, the first three natural frequencies of the present structure are 1.01, 3.00, and 4.94 Hz.

Table 1. Mean value and coefficient of variation of structure and TMD random parameters.

Random Variable	Mean Value	Coefficient of Variation (%)
Mass per story	360 ton	5
Stiffness per story	650 MN/m	5
Damping per story	6.2 MNs/m	5
Mass of the TMD	108 ton	5
Stiffness of the TMD	3.865 MN/m	5
Damping of the TMD	0.181 MNs/m	5

2.3 Random Seismic Excitations

According to [11], to solve Eq. (1) it is necessary to define the seismic loading. In this study, the seismic load is defined as a force of a one-dimensional seismic event that is simulated through the Kanai-Tajimi spectrum [12, 13] with a power spectral density function given by Eq. (5):

$$S(\omega) = S_0\left[\frac{\omega_g^4 + 4\omega_g^2\xi_g^2\omega^2}{\left(\omega^2-\omega_g^2\right)^2 + 4\omega_g^2\xi_g^2\omega^2}\right], S_0 = \frac{0.03\xi_g}{\pi\omega_g\left(4\xi_g^2+1\right)} \quad (5)$$

where S$_0$ is the constant spectral density, related to the peak ground acceleration (PGA), ω_g is the ground frequency and ξ_g is the ground damping. Uncertainties in the ground and in the seismic excitation itself, which cannot be predicted, interfere with the optimal solution of the Kanai-Tajimi spectrum, therefore the excitation must be considered taking into account uncertainties. In the case of the seismic excitation, two levels of uncertainty can be observed, first the uncertainty of the random phase angle of the Tanai-Kajimi formula and finally the uncertainties of the soil of the dynamic excitation, in this case, the frequency of the soil, the soil damping and the PGA are assumed as independent Lognormal variables with known mean and coefficient of variation, as well as structure

Table 2. Mean value and coefficient of variation of seismic random parameters.

Random Variable	Mean Value	Coefficient of Variation (%)
PGA	0.35 g	10
ω_g	1 Hz	10
ξ_g	0.3	10

Note in Table 2 that the ground frequency value is very close to the natural frequency of the building, around 1 Hz. When the frequencies approach, the resonance phenomenon occurs, where the displacement amplitudes tend to infinity, so this is the worst possible situation for this structure [10]. In Fig. 2 the spectral intensity of the earthquake is plotted, where the density of frequencies around 1 Hz is confirmed parameters, so the excitation has uncertainty over uncertainty [5]. The random variables of the problem, their mean values and their respective coefficient of variation of 10%, representing the uncertainties of the excitation, are described in Table 2.

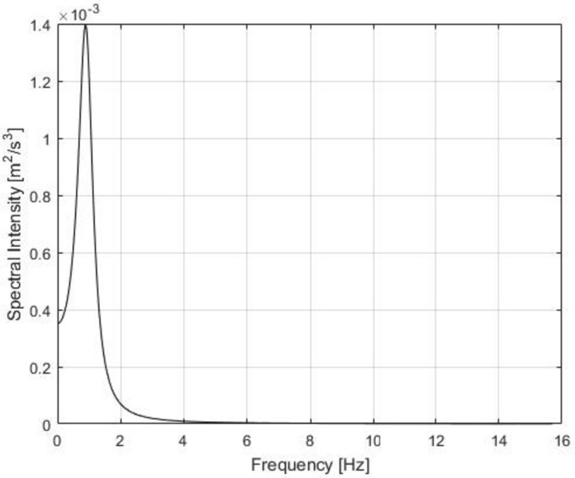


Fig. 2. The Kanai-Tajimi spectral intensity in relation to the frequency for the soil scenario of Table 2.

In this study, the earthquake was generated for a duration of 20 s with an integration step of 0.02 s. Following the mean values mentioned in Table 2 this graph was generated and plotted in Matlab, and the demonstration of the earthquake is illustrated in Fig. 3.

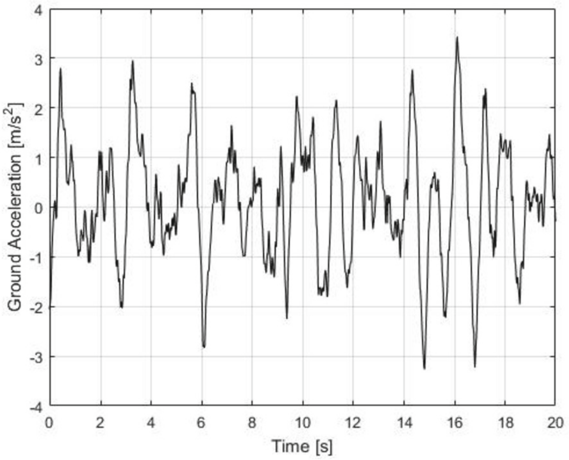


Fig. 3. Accelerogram of the artificial seismic event generated with Kanai-Tajimi spectrum of the mean values in Table 2.

2.4 Monte Carlo Simulation and Latin Hypercube Sampling

The Monte Carlo Simulation (MCS) is a numerical method based on sampling random values with a large number of observations in order to acquire statistical results for probabilistic situations. The simulation starts from an average value with a specific variation to obtain the observations, where first it is assumed that all random variables are unrelated to each other, adjusting the problem in terms of these variables, and characterizing the probabilities as a probability density function to finally generate their determined values [14].

Thus, to generate the random variables for MCS was used Latin Hypercube Sampling (LHS). The LHS reduces the computational cost and provides an efficient way to generate variables from their distributions, taking samples from equal intervals, and selecting different values of a random variable where the domain of the random variable is divided into *n* intervals of equal probability. A value from each range is chosen at random with respect to the probability density in the interval, the choice must be made in a random way with respect to the density in each interval, so the selection must reflect the height of the density across the range [5].

3 Results and Discussions

Completing the studies of this work and considering the case of the proposed ten-story-building with a TMD at the top, subjected to random artificial earthquakes, the Matlab computational routine was executed providing, from the application of the Newmark method, the values of displacement, velocity and acceleration for each instant of time. For this situation, the results obtained of maximum interstory drift are the ones desired for the vibration analysis of the problem, being able to obtain them from the difference between the displacement values of each floor. Therefore, the values of the structural model with TMD can be compared with the model without vibration control, analyzing the effectiveness of installing this device. Considering that each story of the building is a DOF, these are the values of maximum drift, described in Table 3.

Analyzing Table 3, it is notable that there is a considerable reduction in the maximum drift values per floor, reaching values around 55%, the expected reduction values are due to the installation of the tuned mass damper in the structure. The maximum interstory drift value has occurred on the second floor of the building, so this will be used for general analysis as it is the most critical situation in this case. Then it becomes necessary to define the number of samples to be used in the course of the study, so a graph of convergence of the mean of the maximum interstory drift values is drawn up to define the ideal number to be used for the work, it is shown in Fig. 4.

Analyzing Fig. 4, it was decided to use the amount of 300 samples, as this is where the line tends to stabilize, providing greater credibility of results. Based on this definition, a probability density function was created in relation to the maximum drift for the number set of samples, in the situation without vibration control and in the case with TMD installed, the results can be compared and analyzed in Fig. 5.

Figure 5 demonstrates the frequency diagrams in unit area histograms constructed for the uncontrolled structure case (red histogram) and for the controlled structure case (blue histogram) for maximum drift. Looking at the graph, it can be seen that the blue

Table 3. Values of maximum inter-story drift in uncontrolled structure and structure with TMD.

Story Number	Uncontrolled Structure (m)	Structure with TMD (m)	Reduction (%)
1	0.0862	0.0400	53.59
2	0.0878	0.0401	54.32
3	0.0793	0.0357	54.98
4	0.0815	0.0364	55.33
5	0.0686	0.0304	55.68
6	0.0665	0.0295	55.63
7	0.0480	0.0215	55.20
8	0.0397	0.0184	53.65
9	0.0262	0.0131	50.00
10	0.0138	0.0090	34.78

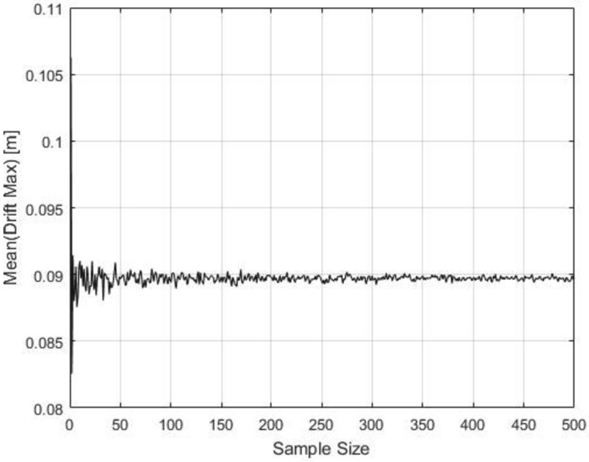


Fig. 4. Convergence of the mean of maximum interstory drift.

curve is slender compared to the red curve, this is due to the variance has been reduced after installing the TMD, thus showing the effectiveness of the proposed methodology.

For a better demonstration of the vibration control event, an analysis graph was plotted for the maximum interstory drift values obtained on the second floor of the building in relation to the time of action of the seismic forces, which is equivalent to 20 s. Thus, the data can be analyzed in another way, and the demonstration of the reduction of amplitudes can be better visualized, as illustrated in Fig. 6.

Looking at the graph in Fig. 6, it is possible to visualize a notable reduction of the amplitudes of the structure, the reduction occurs mainly after 12 s of earthquake action, where the amplitudes of the structure without control (red curve) increase, but those

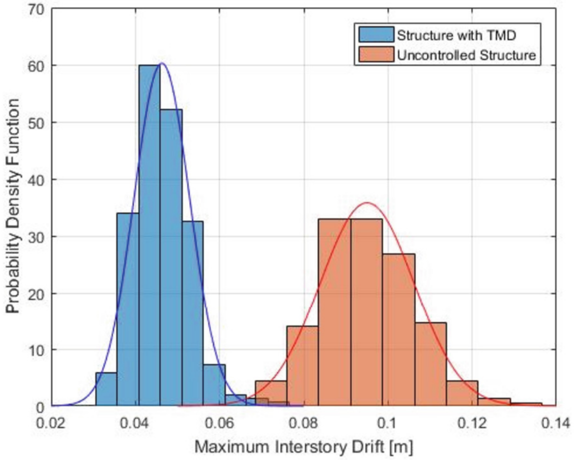


Fig. 5. Probability density function of maximum interstory drift for uncontrolled structure (red curve) and structure with TMD (blue curve). (Color figure online)

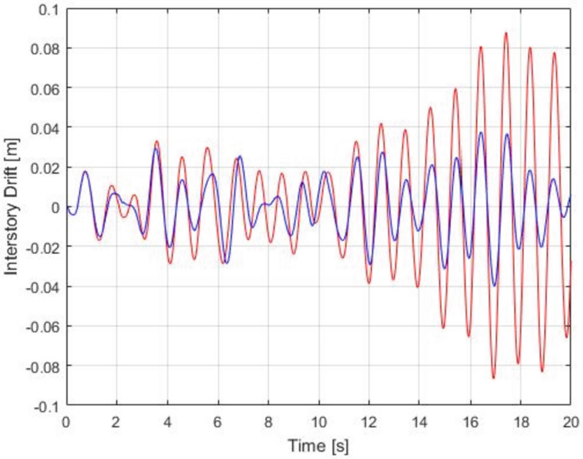


Fig. 6. Interstory drift on the second floor for uncontrolled structure (red curve) and structure with TMD (blue curve). (Color figure online)

of the structure with vibration control (blue curve) are reduced. The values refer to the second floor of the structure because, as already mentioned, it is the worst point for interstory drift in this case.

4 Conclusions

One of the positive points of a TMD is that it is a passive energy dissipation device that do not need mechanisms for activation, so it become cheaper compared to active ones, this is one of the advantages of the tuned mass damper, together with its effectiveness and reliability.

With the results obtained in the numerical simulation, a technical analysis was carried out based on theoretical references in order to prove the values that define the effectiveness of the tuned mass damper. The responses of the comparison between the structure with TMD installed and the structure without a vibration control device, as seen in Fig. 6, demonstrate that over time the seismic event was active, there was a visible drop in interstory drift amplitudes. In percentage values, this vibration control device was effective in reducing interstory drift by about 55% in its most critical case, as shown in Table 3, thus being able to be considered a passive control system applicable to the case study of this work.

In this study, the installation of the vibration control device also contributed to reducing the variance arising from uncertainties in the system, visible in the probability density function in Fig. 5. Uncertainties of the structure, the TMD and the seismic force that are common in real cases, due to human actions and of nature, uncertainties are always present and must be considered because they affect the final results of the study.

Thus, it can be concluded from the values obtained the effectiveness of the proposed methodology and the installation of the vibration control mechanism, which showed that a passive energy dissipation device was effective in reducing the interstory drift of this building, bringing safety in critical earthquake situations that may pose a danger to individuals of the place.

Acknowledgments. The authors are grateful for the financial support from CNPq and CAPES.

References

1. Moutinho, C.M.R: Controlo de Vibrações em Estruturas de Engenharia Civil. Doutorado em Engenharia Civil, Faculdade de Engenharia da Universidade do Porto, Porto, Portugal (2007)
2. Kronbauer, F: Uso de Amortecedores de Massa Sintonizados para Redução de Vibrações em Estruturas Submetidas a Eventos Sísmicos. Monografia (Trabalho de Conclusão de Curso em Engenharia Mecânica) – Departamento de Engenharia Mecânica, Universidade Federal do Rio Grande do Sul, Porto Alegre (2013)
3. Jr Steffen, V., Rade, D.A., Inman, D.J.: Using passive techniques for vibration damping in mechanical systems. J. Braz. Soc. Mech. Sci. **22**, 411–412 (2000)
4. Murudi, M.M. Mane, S.M: Seismic effectiveness of tuned mass damper (TMD) for different ground motion parameters. In: 13th World Conference on Earthquake Engineering. Vancouver, Canada (2004)
5. Vellar, L.S., Ontiveros-Pérez, S.P., Miguel, L.F.F., Miguel, L.F.F.: Robust optimum design of multiple tuned mass dampers for vibration control in buildings subjected to seismic excitation. Shock Vibr. **2019**, 1–9 (2019). https://doi.org/10.1155/2019/9273714

6. Tuan, A.Y1., Shang, G.: Q2: Vibration control in a 101-storey building using a tuned mass damper. 1-Department of Civil Engineering, Tamkang University, Tamsui, Taiwan 251, R.O.C. 2-Department of Civil and Architectural Engineering, City University of Hong Kong, Hong Kong (2014)
7. Lucchini, A., Greco, R., Marano, G.C., Monti, G.: Robust design of tuned mass damper systems for seismic protection of multistory buildings. J. Struct. Eng. **140**(8), A4014009 (2014). https://doi.org/10.1061/(ASCE)ST.1943-541X.0000918
8. Rossato, L.V., Miguel, L.F.F. Fadel Miguel, L.F.: Estimativa de Razões de Massa Ideal de Amortecedor de Massa Sintonizada para Controle de Vibrações em Estruturas. Revista Interdisciplinar de Pesquisa em Engenharia. CILAMSE, Brasília, Brasil (2016)
9. Kelly, S.G.: Vibrações Mecânicas Teorias e Aplicações. Cengage, São Paulo – SP (2017)
10. Rao, S.S.: Mechanical Vibrations, 4th edn. Pearson Prentice Hall, Singapore (2004)
11. Miguel, L.F.F., Fadel Miguel, L.F., Lopez, R.H.: Failure probability minimization of building through passive friction dampers. Department of Mechanical Engineering, Federal University of Rio Grande do Sul, Porto Alegre. Departament of Civil Engeneering. Federal University of Santa Catarina, Florianópolis (2016)
12. Kanai, K.: An empirical formula for the spectrum of strong earthquake motions. Bull. Earthq. Res. Inst. Univ. Tokyo **39**, 85–95 (1961)
13. Tajimi, H.: A statistical method of determining the maximum response of a building structure during an earthquake. In: Proceedings of 2nd World Conference in Earthquake Engineering. World Conference in Earthquake Engineering (WCEE), Tokyo, Japan, pp. 781–797 (1960)
14. Haldar, A. Mahadevan, S: Probability, Reliability, and Statistical Methods in Engineering Design. John Wiley & Sons, Inc., Hoboken (2000)
15. Mohebbi, M., Shakeri, K., Ghanbarpour, Y., Majzoub, H.: Designing optimal multiple tuned mass dampers using genetic algorithms (GAs) for mitigating the seismic response of structures. J. Vibr. Control **19**(4), 605–625 (2012). https://doi.org/10.1177/1077546311434520

Reliability-Based Design Optimization of Steel Frames Using Genetic Algorithms

Laís De Bortoli Lecchi[1](✉), Francisco de Assis das Neves[1], Ricardo Azoubel da Mota Silveira[1], Walnório Graça Ferreira[2], and Eduardo Souza de Cursi[3]

[1] Universidade Federal de Ouro Preto, Rua Diogo de Vasconcelos, 122., Ouro Preto, Brazil
lais.de_bortoli_lecchi@insa-rouen.fr

[2] Universidade Federal do Espírito Santo, Avenida Fernando Ferrari, 514., Vitória, Brazil

[3] Institut National des Sciences Appliquées de Rouen, Avenue de l'Université, 685., Saint-Étienne-du-Rouvray, France

Abstract. In the design of structures, there are uncertainties of different origin often associated with the properties of materials, geometry and applied loads. With the Reliability-Based Design Optimization (RBDO) method, it is possible to consider design constraints in terms of failure probabilities or target reliability indices, for a structure subject to performance constraints as limit state functions (LSF), in a classical optimization problem. In this way, RBDO analysis takes design variables uncertainties and its effects directly. This work intents to present a RBDO application in a steel frame, with an usual double-loop approach, considering the first and second order structural analysis, with optimization by Genetic Algorithms (GA). Target reliability indices are defined and assessed by FORM (First Order Reliability Method), while GA searches the optimal solution between 18 W-shapes from AISC database (2017), which represents the mininum material mass required for satisfy the constraints. In some cases, it is shown that considering second-order effects can result in lighter frames, as the calculated reliability index can get higher.

Keywords: RBDO · Genetic Algorithms · FORM · Structural Optimization · Second Order Analysis

1 Introduction

The design of structures is directly associated to satisfy conflicting requirements, such as cost, safety and durability. Unquestionably, uncertainties are naturally present in all the variables and steps that compose the design. Thus, optimizing a structure through a deterministic approach often results in poor reliable configurations.

In this way, the Reliability-Based Design Optimization (RBDO) approach is able to find the best compromise between cost and reliability assurance, by

J. E. S. De Cursi (Ed.): UQSM 2023, LNME, pp. 97–111, 2024.
https://doi.org/10.1007/978-3-031-47036-3_9

including the uncertainties [3]. The challenge of RBDO is the computational effort employed in evaluate reliability constraints, making it difficult to deal with more complex problems [8].

Besides the usual double-loop approach, other techniques have been proposed to speed up the RBDO process, such as SORA (Sequential Optimization and Reliability Analysis) [10], SLA (Single-Loop Approach) [17] and SAP (Sequential Approximate Programming) [9].

Truong and Kim [25] point out that for deterministic design optimization (DDO), there are many articles and research material in the literature, but in the case of RBDO, few studies related to steel frames have been performed. In addition, geometric and material nonlinear behaviors often are not considered [23,26].

This paper intents to present a RBDO double-loop approach of a steel frame, considering first and second order structural analyses. FORM is used to obtain the reliability index β and Genetic Algorithms perform the optimization. Three target reliability indices are established and design variables can also assume different values. Reliability indices found by second order analysis tend to be higher than the first order ones. However, the minimum mass found by second order approach is not always smaller than first order case.

1.1 Softwares

CS-ASA. The *CS-ASA - Computational System for Advanced Structural Analysis* is a finite element method based program [24], able to perform statics and dynamics analysis of steel structures [6] and statics analysis of composites steel-concrete structures [16], considering geometric imperfections, material nonlinearity and semi-rigid connections.

MATLAB. MATLAB® manages all the analysis stages, as calling the structural analysis program *CS-ASA*; the reliability loop (FORM algorithm) and the optimization loop ('ga' function in MATLAB®).

1.2 Second Order Analysis of Structures

The *CS-ASA* presents 3 options for second order analysis formulations. The one used in this work, called SOF-2 [24], was developed by Yang and Kuo, in 1994 [27]. The typical frame finite element adopted can be seen in Fig. 1 and its implementation passes by some simplifying assumptions, such as: cross sections remain flat after deformation and are compacts; lateral or torsional buckling are not allowed; small deformations are assumed, but large rotations/displacements are allowed; axial shortening due to curvature is neglected.

Achieving the condition of structural equilibrium consists of resolving a balance between applied external forces and internal forces of the structure [4]. Such task can be expressed in an equation, as Eq. 1 and, for the second order analysis case, it depends on displacements ($\mathbf{U}$) and internal forces in the members ($\mathbf{P}$).

$$\mathbf{F_{int}}(\mathbf{U},\mathbf{P}) \simeq \mathbf{F_{ext}} \tag{1}$$

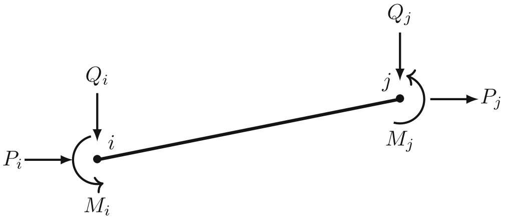


Fig. 1. Adopted finite element.

$\mathbf{F_{int}}$ is the vector of internal forces; $\mathbf{F_{ext}}$ is the vector of external forces, which can be expressed as the product between a load parameter λ and a reference external force vector $\mathbf{F_r}$, which defines the direction of the acting external forces [4]. So, $\mathbf{F_{ext}} = \lambda\mathbf{F_r}$.

The numerical strategy to solve Eq. 1 is an incremental-iterative approach, based on Newton-Raphson method. Thus, it is more convenient to expand Eq. 1, defining the elastic and geometrical matrices involved in the process:

$$\mathbf{F^t_{int}} + \sum_e^n [\mathbf{k^e_i} + \mathbf{k^e_g}]\mathbf{\Delta u^e} \simeq \mathbf{F^t_{ext}} + \varDelta\lambda\mathbf{F_r} \tag{2}$$

where the superscript "t" defines the last equilibrium configuration; $\mathbf{\Delta u^e}$ is the incremental nodal displacement vector of the element "e"; $\mathbf{k^e_i}$ is the element's elastic linear stiffness matrix, defined in Eq. 3; $\mathbf{k^e_g}$ is the element's geometric stiffness matrix, defined in Eq. 4; $\varDelta\lambda$ is the load parameter increment.

$$\mathbf{k^e_i} = \int_{L^e} \mathbf{N^T D N} dx \tag{3}$$

$$\mathbf{k^e_g} = \mathbf{P} \int_{L^e} [\mathbf{N_u^T N_u} + \mathbf{N_v^T N_v}] dx \tag{4}$$

L^e is the finite element length; $\mathbf{N}$ refers to the interpolation functions vector; $\mathbf{D}$ represents the material constitutive relationship matrix; $\mathbf{P}$ is the axial force acting on the finite element. The interpolation function vectors $\mathbf{N_u}$ and $\mathbf{N_v}$ are associated with the axial and lateral displacements, respectively.

2 Structural Optimization

2.1 Optimization Problem

The optimization problem consists of maximizing or minimizing one or more objective functions, within specific design conditions previously established [21]. We may formulate it as follows:

$$Find \;\; \mathbf{X} = \left\{ x_1 \;\; x_2 \;\; ... \;\; x_n \right\}, \;\; that \;\; minimizes/maximizes \;\; f(\mathbf{X}),$$

subject to:

$$c_i(\mathbf{X}) \leq 0, \quad i = 1, 2, ..., m, \tag{5}$$

$$d_j(\mathbf{X}) = 0, \quad j = 1, 2, ..., p, \tag{6}$$

$$x_k^{low} \leq x_k \leq x_k^{up}, k = 1, 2, ..., n, \tag{7}$$

where:

- $\mathbf{X}$ is the n-dimensional vector containing the design variables to be optimized;
- $f(\mathbf{X})$ is the objective function of the problem, which in structural optimization, can represent the weight, volume or manufacturing cost, for example;
- $c_i(\mathbf{X})$ and $d_j(\mathbf{X})$ are inequality and equality constraints, respectively, known as *behavior constraints*, related to the performance and limit states of the structural system under study;
- x_k^{low} and x_k^{up} are the lower and upper bounds that design variables can assume, known as *lateral constraints*, related to feasible physical limits [22];
- i, j, k, m, n and p are arbitrary values.

Figure 2 represents a hypothetical two-dimensional problem, in which the feasible region was obtained by applying two behavior constraints c_a and c_b, as well as the lateral constraints x_1^{low}, x_1^{up}, x_2^{low}, x_2^{up}.

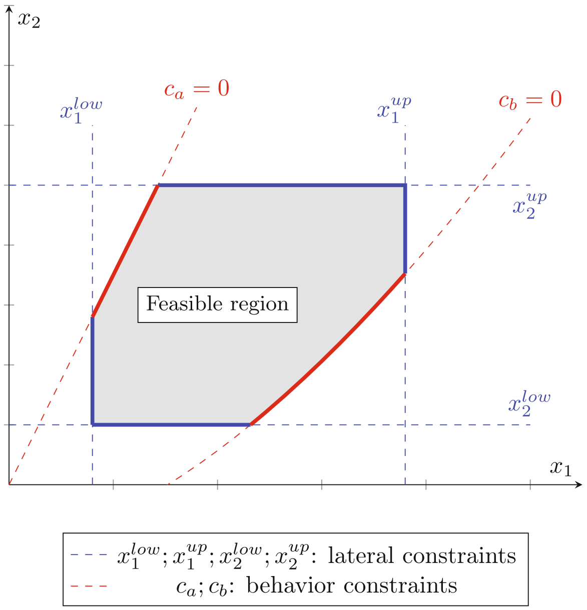

x_2
$c_a = 0$
x_1^{low}
x_1^{up}
$c_b = 0$
x_2^{up}
Feasible region
x_2^{low}
x_1
$x_1^{low}; x_1^{up}; x_2^{low}; x_2^{up}$: lateral constraints
$c_a; c_b$: behavior constraints

Fig. 2. Constraints surfaces for a hypothetical two-dimensional problem

2.2 Genetic Algorithms

In 1975, Holland [15] proposed a new optimization method, based on principles of nature, such as genetics and natural selection in the reproduction of species: the Genetic Algorithms (GA). The GA's make part of a set of so-called modern optimization methodologies [22].

As stochastic and gradient-free method, GA has good applicability in problems like multi-objective optimization; problems containing mixed continuous and discrete variables; also for discontinuous or non-differentiable functions, as well as for non-convex design spaces. The basic terminology relevant to genetic algorithms is cited below:

- **Objective function:** the function to be optimized;
- **Penalty function:** mathematical expression applied to the fitness value of an individual, calculated based on the violation of constraints;
- **Fitness function:** mathematical expression given by the sum of the objective and penalty functions, which indicates how fitted to the problem an individual solution can be;
- **Individual:** is the variables vector. It is also called *chromosome* and, its entries, *genes*. Vector **X** below represents this structure:

$$\mathbf{X} = \begin{bmatrix} x_1 & x_2 & x_3 & \dots & x_n \end{bmatrix};$$

- **Population:** is the matrix of individuals. The user must specify a value p, for the population size. Therefore, the population matrix will have dimension $p \times n$ and n is the number of variables in the problem;
- **Generation:** each generation represents an iteration, in which a new population matrix will be created, by applying the genetic operators, known as: **selection**, **elitism**, **crossover** and **mutation**;
- **Diversity:** is measured by the distance between individuals in a population. Greater the diversity of a population is, greater is the scan of the design space;
- **Parents and Children:** The GA's, through the selection process, use the individuals with the best fitness value of the current generation, called parents, to create those of the next iteration (children).

The flowchart in Fig. 3 outlines the running of genetic algorithms.

3 Reliability Analysis and RBDO Methodology

3.1 First Order Reliability Method - FORM

The reliability indices calculated in this work are obtained by applying FORM, that is an approximation of the limit state function, by a tangent hyper-surface at the design point [11], where the distance from the origin to this point is what we call the reliability index β [13]. Once we have β, it is possible calculate the failure probability p_f^{FORM}, which is the cumulative standard normal distribution function (Φ) value at $-\beta$ (Eq. 8).

$$p_f^{FORM} = \Phi(-\beta) \tag{8}$$

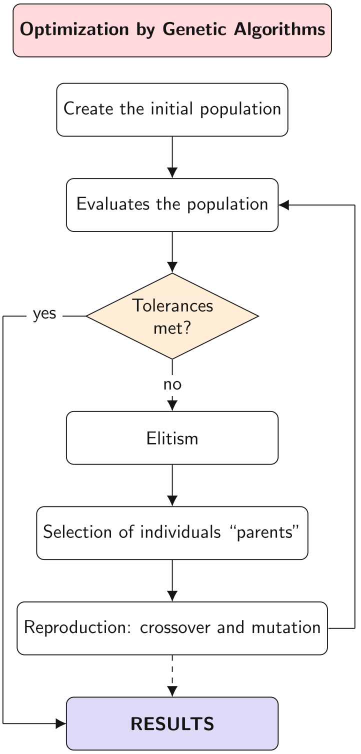

Optimization by Genetic Algorithms
Create the initial population
Evaluates the population
Tolerances met?
yes
no
Elitism
Selection of individuals "parents"
Reproduction: crossover and mutation
RESULTS

Fig. 3. Genetic algorithms flowchart.

Obtaining β involves a mapping transformation of random variables [18,20]. For the Nataf's transformation case, this operation starts from an original space $\mathbf{X}$ to a normal space $\mathbf{Z}$, and then from the space $\mathbf{Z}$ to a standard normal uncorrelated space $\mathbf{Y}$. Equations 9 and 10 show the chain rule for the Jacobian matrices, used in the process:

$$\mathbf{J}_{yx} = \left[\frac{\partial y_i}{\partial x_k}\right] = \left[\frac{\partial y_i}{\partial z_j}\frac{\partial z_j}{\partial x_k}\right] = \mathbf{L}^{-1}(\mathbf{D}^{neq})^{-1} = \mathbf{J}_{yz}\mathbf{J}_{zx} \tag{9}$$

$$\mathbf{J}_{xy} = \left[\frac{\partial x_i}{\partial y_k}\right] = \left[\frac{\partial x_i}{\partial z_j}\frac{\partial z_j}{\partial y_k}\right] = \mathbf{D}^{neq}\mathbf{L} = \mathbf{J}_{xz}\mathbf{J}_{zy} \tag{10}$$

where $\mathbf{L}$ is the lower triangular matrix obtained from the Cholesky decomposition and $\mathbf{D^{neq}}$ is the diagonal matrix of standard deviations of equivalent normal variables. Thus, $\mathbf{x}$ and $\mathbf{y}$ variables can be obtained by Eqs. 11 and 12:

$$\mathbf{y} = \mathbf{J}_{yx}\left\{\mathbf{x} - \mu^{neq}\right\} \tag{11}$$

$$\mathbf{x} = \mathbf{J}_{xy}\left\{\mathbf{y} + \mu^{neq}\right\} \tag{12}$$

where μ^{neq} is the normal equivalent mean.

FORM calculates the reliability index β by means of the following steps (based on [7]):

1. Calculation of non-normal distributions parameters;
2. Determination of equivalent correlation coefficients and the Cholesky decomposition matrix $\mathbf{L}$;
3. Determination of Jacobian matrices $\mathbf{J}_{yz}$ and $\mathbf{J}_{zy}$;

$$\mathbf{J}_{yz} = \mathbf{L}^{-1} \tag{13}$$

$$\mathbf{J}_{zy} = \mathbf{L} \tag{14}$$

4. Choice of the starting point x_k, for $k = 0$ (beginning of the iterative process);
   Start of the iterative process
5. Calculation of equivalent normal distributions parameters;
6. Updating the Jacobian matrices $\mathbf{J}_{yx}$ and $\mathbf{J}_{xy}$;
7. Transformation of point x_k from $\mathbf{X}$ to $\mathbf{Y}$;
8. Limit state function $g(x_k)$ assessment;
9. Calculation of gradients:
   a. Calculation of the partial derivatives of $g(\mathbf{X})$ in the design space $\mathbf{X}$;
   b. Gradient transformation to $\mathbf{Y}$;
   c. Calculation of linearized sensitivity factors $\alpha(y_k)$;
10. Calculation of the new point y_{k+1};
11. Transformation of y_{k+1} to space $\mathbf{X}$;
12. Convergence check. If convergence criteria are met, the algorithm is interrupted. Otherwise, the iteration number is increased and it returns to step 5. Convergence criteria:

$$1 - \frac{|\nabla g(\mathbf{y}_{k+1})^t \mathbf{y}_{k+1}|}{||\nabla g(\mathbf{y}_{k+1})||\ ||\mathbf{y}_{k+1}||} < \varepsilon \tag{15}$$

$$|g(\mathbf{y}_{k+1})| < \delta \tag{16}$$

13. Evaluation of the reliability index at the design point: $\beta = ||y^*||$.

3.2 Reliability-Based Design Optimization - RBDO

In RBDO methodology, the uncertainties related to each variable of a problem are directly taken into account in the optimization process [7]. Failure probabilities or targets reliability indices are defined as optimization constraints. Hilton and Feigen [14] first proposed the method in their work: *Minimum weight analysis based on structural reliability*, in 1960.

In this way, we must add the reliability constraint to the optimization problem presented in Subsect. 2.1:

$$P[g_i(\mathbf{X})] \leq P_f, \quad i = 1, 2, ..., n \tag{17}$$

or:

$$\beta_i(\mathbf{X}) \geq \beta_T, \quad i = 1, 2, ..., n \tag{18}$$

where $P[g_i(\mathbf{X})]$ is the failure probability of a structure for a given limit state function $g_i(\mathbf{X})$; P_f is the failure probability calculated by Eq. 19 below; $\beta_i(\mathbf{X})$ is the reliability index of a structure; β_T is the target reliability index.

$$P_f = P[\mathbf{X} \in \Omega_f] = \int_{\Omega_f} f_X(\mathbf{X})\, d\mathbf{X} \tag{19}$$

Ω_f is the failure domain; $f_X(\mathbf{X})$ is the probability density function for the random variable $\mathbf{X}$.

Table 1 shows a sequence of steps for the RBDO analysis, using a double-loop approach, where optimization is the outer loop and reliability assessment is the inner one.

4 Numerical Example: RBDO Analysis of a Single Floor Steel Frame

4.1 General Information

A RBDO analysis is made for the single floor steel frame shown in Fig. 4. The problem has 8 random variables, whose statistical characteristics are in Table 2, including the applied loads D, L and W; the section properties: area A, inertia I_x and plastic section modulus Z_x; material properties: Young's modulus E and yield strength F_y.

Table 3 shows the W-shapes characteristics, from AISC database (2017), in which the optimizer searches for the best configuration to satisfy the constraint (a target reliability index) and the objective function, which is minimize the total mass. This frame has been studied by several authors [5,12,19], but originally as a reliability problem only.

4.2 Limit State Function

For the reliability analysis carried out by FORM, one ultimate limit state is verified, which is flexure and axial force acting on column element 4, node 4. This interaction of efforts shall be limited by Eqs. 20 and 21 [2]:

(i) If $\frac{P_r}{P_c} \geq 0.2$

$$\frac{P_r}{P_c} + \frac{8}{9}\left(\frac{M_{rx}}{M_{cx}} + \frac{M_{ry}}{M_{cy}}\right) \leq 1.0 \tag{20}$$

Table 1. General sequence of steps for the RBDO analysis.

RBDO double-loop approach (MATLAB®)

```
Define $n$ (number of repetitions);

For $i$ from 1 to $n$, do:

    Structural analysis: Calls the FEM Program CS-ASA;

    opts = optimoptions(@ga, ...); (set genetic algorithms options)

    $A$ = [...]; $b$ = [...]; $A_{eq}$ = [...]; $b_{eq}$ = [...]; $lb$ = [...]; $ub$ = [...]; $intcon$ = [...];
    (set behavior and lateral constraints)

    [x,fval] = ga(@Fobj,nvars,A,b,Aeq,beq,lb,ub,@nonlcon,intcon, opts);
    (calls GA, performs the optimization and returns the variables X
    optimized and also the value of the objective function fval)

        ↪ Inner loop - Reliability constraints evaluation: using FORM,
        the solver checks whether the found reliability index is above
        the target;

    WriteNewFile(X); (calls the function that will rewrite the file to
    CS-ASA, for a new structural analysis with the updated variables X);

End-For
```

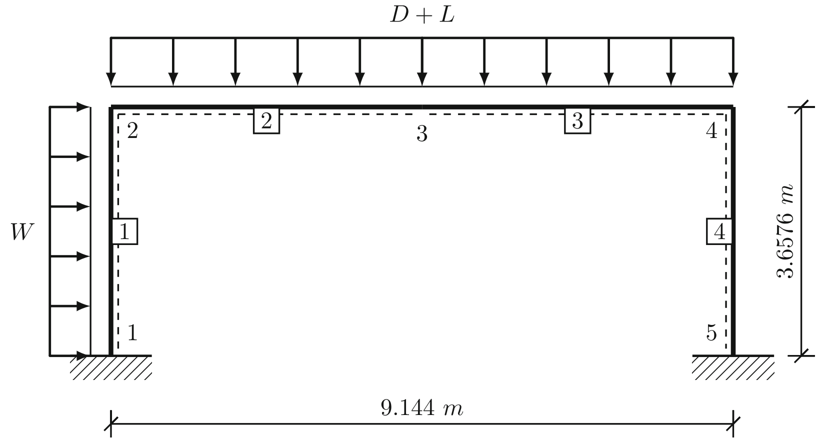


Fig. 4. Single Floor Steel Frame.

(ii) If $\frac{P_r}{P_c} < 0.2$

$$\frac{P_r}{2P_c} + \left(\frac{M_{rx}}{M_{cx}} + \frac{M_{ry}}{M_{cy}}\right) \leq 1.0 \tag{21}$$

Table 2. Statistical properties of random variables [12].

Variable	Unit	Mean	Coefficient of variation	Distribution function
D	kN/m	6.42	0.10	Normal
L	kN/m	0.73	0.25	Ext. Value - Type 1 (largest)
W	kN/m	5.98	0.37	Ext. Value - Type 1 (largest)
A	cm^2	Table 3	0.05	Normal
I	cm^4	Table 3	0.05	Normal
Z_x	cm^3	Table 3	0.05	Normal
E	MPa	199947.96	0.06	Normal
F_y	MPa	273.03	0.11	Normal

where P_r: required axial strength; P_c: available axial strength (Eqs. 22 and 23, if tension or compression); M_r: required flexural strength; M_c: available flexural strength (Eq. 24); x: major axis bending; y: minor axis bending.

$$P_{c,ten} = AF_y \tag{22}$$

$$P_{c,com} = AF_{cr} \tag{23}$$

$$M_c = Z_x F_y \tag{24}$$

F_{cr} is the critical stress given by Eq. 25 or Eq. 26:

(i) If $\lambda_c \leq 1.5$

$$F_{cr} = \left(0.658^{\lambda_c^2}\right) F_y \tag{25}$$

(ii) If $\lambda_c > 1.5$

$$F_{cr} = \left(\frac{0.877}{\lambda_c^2}\right) F_y \tag{26}$$

λ_c is the reduced slenderness ratio [1], calculated by Eq. 27:

$$\lambda_c = \frac{KL}{\pi}\sqrt{\frac{AF_y}{EI_x}} \tag{27}$$

where K is the effective length factor and L the laterally unbraced length of the member.

4.3 Design Variables

The variables are W-shapes, taken as discretes by GA optimizer, varying from 1 to 18 and then mapped to Table 3, whose characteristics like linear mass, area (A), inertia (I_x) and plastic section modulus (Z_x) are used in the process.

For first and second order analyses, 3 possibilities were studied: 1- considering all elements with the same W-shape *(1 optimization variable)*; 2- considering beam elements and columns elements with different W-shapes *(2 optimization variables)*; 3- considering beam elements with the same W-shape, but allowing columns with different W-shapes *(3 optimization variables)*.

Table 3. W-shapes properties of the design space.

n°	Label	Mass $[kg/m]$	A $[10^{-3}m^2]$	I_x $[10^{-5}m^4]$	Z_x $[10^{-4}m^3]$
1	W250x17.9	17.9	2.28	2.24	2.06
2	W200x19.3	19.3	2.48	1.65	1.87
3	W310x21.0	21.0	2.68	3.69	2.85
4	W250x22.3	22.3	2.85	2.87	2.62
5	W200x22.5	22.5	2.86	2.00	2.23
6	W150x22.5	22.5	2.86	1.21	1.77
7	W310x23.8	23.8	3.04	4.29	3.29
8	W150x24.0	24.0	3.06	1.34	1.92
9	W250x25.3	25.3	3.22	3.41	3.06
10	W200x26.6	26.6	3.39	2.58	2.79
11	W130x28.1	28.1	3.59	1.09	1.09
12	W310x28.3	28.3	3.59	5.41	4.05
13	W250x28.4	28.4	3.63	4.01	3.54
14	W150x29.8	29.8	3.79	1.72	2.46
15	W200x31.3	31.3	3.97	3.13	3.34
16	W250x32.7	32.7	4.19	4.91	4.26
17	W200x35.9	35.9	4.57	3.44	3.79
18	W150x37.1	37.1	4.74	2.22	3.10

4.4 Design Constraints

Besides the lateral constraints of the previous item, there is also the reliability constraint, given by a target value, as Eq. 28. Three scenarios were proposed: $\beta_{T,1} = 2.0$, $\beta_{T,2} = 2.5$ and $\beta_{T,3} = 3.0$.

$$c = \frac{\beta_{T,i}}{\beta_i} - 1 \leq 0 \tag{28}$$

4.5 Objective Function

The objective function $M(\mathbf{X})$, which represents the minimum mass of the structure, is given by Eq. (29):

$$M(\mathbf{X}) = \sum_{i=1}^{n} m_i l_i, \tag{29}$$

where n is the number of variables; m_i is the linear mass for a given W-shape (Table 3); l_i is the length of the bar.

4.6 Optimization Algorithm Setting

Genetic Algorithms were used to optimize the structure, with the following specific settings:

- Population size ('PopulationSize'): 10 - 12 individuals;
- Creation function ('CreationFcn'): 'gacreationuniform' (default);
- *Crossover* function ('CrossoverFcn'): 'crossoverscattered' (default);
- Mutation function ('MutationFcn'): 'mutationgaussian' (default);
- Elite individuals ('EliteCount'): 5% of population size (default);
- Maximum number of generations ('MaxGenerations'): 200;
- Algorithm for handling nonlinear constraints ('NonlinearConstraintAlgorithm'): 'penalty';
- Tolerance for objective function ('FunctionTolerance'): 10^{-6};
- Tolerance for constraints ('ConstraintTolerance'): 10^{-3};

4.7 Results

Case A: One Design Variable. Table 4 shows the results obtained for case where all the elements have same section. It can be seen that for $\beta_{T,3} = 3.0$, the second order analysis reaches a economy of material of 11.8%, when compared to the first order case. Furthermore, second order case presents smaller constraint violations and, consequently, higher calculated reliability indices, when both approaches use the same structural configuration.

Table 4. Case A: one design variable.

β_T	Analysis	Mass (kg)	W-shape	Const. c	β_i
2.0	1° Order	345.6	W310 × 21.0	−0.318	2.933
	2° Order	345.6	W310 × 21.0	−0.391	3.284
2.5	1° Order	345.6	W310 × 21.0	−0.147	2.933
	2° Order	345.6	W310 × 21.0	−0.239	3.284
3.0	1° Order	391.7	W310 × 23.8	−0.192	3.713
	2° Order	345.6	W310 × 21.0	−0.087	3.284

Case B: Two Design Variables. Table 5 shows the results obtained for case where beam elements and column elements have different sections. It can be seen that for $\beta_{T,3} = 3.0$, the second order analysis reaches a economy of material of 6.1%, when compared to the first order case. Furthermore, second order case presents smaller constraint violations and, consequently, higher calculated reliability indices, when both approaches use the same structural configuration.

Table 5. Case B: two design variables.

β_T	Analysis	Mass (kg)	W-Columns	W-Beam	Const. c	β_i
2.0	1° Order	317.3	W310 × 21.0	W250 × 17.9	−0.267	2.729
	2° Order	317.3	W310 × 21.0	W250 × 17.9	−0.339	3.025
2.5	1° Order	317.3	W310 × 21.0	W250 × 17.9	−0.084	2.729
	2° Order	317.3	W310 × 21.0	W250 × 17.9	−0.174	3.025
3.0	1° Order	337.8	W310 × 23.8	W250 × 17.9	−0.135	3.468
	2° Order	317.3	W310 × 21.0	W250 × 17.9	−0.008	3.025

Case C: Three Design Variables. Table 6 shows the results obtained for case where beam elements are composed of the same W-shape and columns elements may differ in their sections. It can be observed that in the 3 scenarios, both approaches had the same structural configuration. Furthermore, second order case presents smaller constraint violations and, consequently, higher calculated reliability indices.

Table 6. Case C: three design variables.

β_T	Analysis	Mass (kg)	W-Column 1	W-Beam	W-Column 4	Const. c	β_i
2.0	1° Order	306.0	W250 × 17.9	W250 × 17.9	W310 × 21.0	−0.238	2.626
	2° Order	306.0	W250 × 17.9	W250 × 17.9	W310 × 21.0	−0.315	2.918
2.5	1° Order	306.0	W250 × 17.9	W250 × 17.9	W310 × 21.0	−0.048	2.626
	2° Order	306.0	W250 × 17.9	W250 × 17.9	W310 × 21.0	−0.143	2.918
3.0	1° Order	316.2	W250 × 17.9	W250 × 17.9	W310 × 23.8	−0.104	3.348
	2° Order	316.2	W250 × 17.9	W250 × 17.9	W310 × 23.8	−0.180	3.657

5 Conclusions

This work presented theoretical topics and an example for the RBDO method, using FORM and Genetic Algorithms, through a double loop approach. Despite the simplicity of the numerical application, a considerable computational effort was used in the process, with the analysis time being quite dependent on the computer's processing power.

It was possible to notice that as far as the design variables were increased, closer was the mass found by the first or second order analyses. However, second order case presented smaller constraint violations and, consequently, higher calculated reliability indices, when both approaches used the same structural configuration.

Acknowledgments. The authors acknowledge the support given by UFOP (Federal University of Ouro Preto), FAPEMIG, FAPES, CNPq and CAPES (this study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001).

References

1. ABNT NBR 8800-2008: Projeto de estruturas de aço e de estruturas mistas de aço e concreto de edifícios. Associação Brasileira de Normas Técnicas, Rio de Janeiro (2008)
2. AISC: Manual of steel construction: Load and Resistance Factor Design. American Institute of Steel Construction, Chicago (2016)
3. Aoues, Y.: Optimisation fiabiliste de la conception et de la maintenance des stuctures. Ph.D. Thesis, University Blaise Pascal, Clermont-Ferrand (2008)
4. Azevedo, I.S., Silva, A.R.D., Silveira, R.A.M.: Influence of inverted-v-braced system on the stability and strength of multi-story steel frames. REM - Int. Eng. J. **76**, 39–46 (2023)
5. Baingo, D.: A framework for stochastic finite element analysis of reinforced concrete beams affected by reinforcement corrosion. Ph.D. Thesis, University of Ottawa, Ottawa (2012)
6. Batelo, E.A.P.: Análise Dinâmica Avançada de Estruturas de Aço com Ligações Semirrígidas e Interação com o Solo. Programa de Pós-Graduação em Engenharia Civil, Ouro Preto (2018)
7. Beck, A.T.: Confiabilidade e segurança das estruturas. Elsevier Brasil (2019)
8. Chen, Z., Qiu, H., Gao, L., Su, L., Li, P.: An adaptive decoupling approach for reliability-based design optimization. Comput. Struct. **117**, 58–66 (2013)
9. Cheng, G., Xu, L., Jiang, L.: A sequential approximate programming strategy for reliability-based structural optimization. Comput. Struct. **84**, 1353–1367 (2006)
10. Du, X., Chen, W.: Sequential optimization and reliability assessment method for efficient probabilistic design. J. Mech. Des. **126**, 225–233 (2004)
11. Gomes, W.J.S., Beck, A.T.: Global structural optimization considering expected consequences of failure and using ANN surrogates. Comput. Struct. **126**, 56–68 (2013)
12. Haldar, A., Mahadevan S.: Reliability Assessment Using Stochastic Finite Element Analysis. John Wiley & Sons, Hoboken (2000)
13. Hasofer, A.M., Lind, N.C.: An exact and invariant first-order reliability format. J. Eng. Mech. **100**, 111–121 (1974)
14. Hilton, H.H., Feigen, M.: Minimum weight analysis based on structural reliability. J. Aerosp. Sci. **27**, 641–652 (1960)
15. Holland, J.H.: Adaptation in natural and artificial systems. University of Michigan Press, Ann Arbor (1975)

16. Lemes, I.J.M.: Estudo numérico avançado de estruturas de aço, concreto e mistas. Programa de Pós-Graduação em Engenharia Civil, Ouro Preto (2018)
17. Liang, J., Mourelatos, Z.P., Tu, J.: A single-loop method for reliability-based design optimization. In: International Design Engineering Technical Conferences and Computers and Information in Engineering Conference (2004)
18. Madsen, H.O., Krenk, S., Lind, N.C.: Methods of Structural Safety. Prentice Hall, Englewood Cliffs (1986)
19. Mapa, D.L.S.: Confiabilidade estrutural de pórticos metálicos planos. Programa de Pós-Graduação em Engenharia Civil, Ouro Preto (2016)
20. Melchers, R.E.: Structural Reliability Analysis and Prediction, 2nd edn. John Wiley and Sons, New York (1999)
21. Messac, A.: Optimization in practice with MATLAB®: for engineering students and professionals. Cambridge University Press, New York (2015)
22. Rao, S.S.: Engineering Optimization: Theory and Practice. John Wiley & Sons, Florida (2019)
23. Shayanfar, M., Abbasnia, R., Khodam, A.: Development of a GAbased method for reliability-based optimization of structures with discrete and continuous design variables using OpenSees and Tcl. Finite Elem. Anal. Des. **90**, 61–73 (2014)
24. Silva, A.R.D.: Sistema computacional para análise avançada estática e dinâmica de estruturas metálicas. Programa de Pós-Graduação em Engenharia Civil, Ouro Preto (2009)
25. Truong, V.H., Kim, S.-E.: An efficient method for reliability-based design optimization of nonlinear inelastic steel space frames. Struct. Multidiscip. Optim. **56**, 331–351 (2017)
26. Valdebenito, M.A., Schuëller, G.I.: Reliability-based optimization considering design variables of discrete size. Eng. Struct. **32**, 2919–2930 (2010)
27. Yang, Y.B., Kuo, S.B.: Theory and Analysis of Nonlinear Framed Structures. Prentice Hall, Hoboken (1994)

A Calibration Method for Random Models with Dependent Random Parameters: The Applied Case of Tumor Growth

Carlos Andreu-Vilarroig(✉), Juan-Carlos Cortés, Cristina-Luisovna Pérez, and Rafael-Jacinto Villanueva

Instituto de Matemática Multidisciplinar, Universitat Politècnica de València,
Camí de Vera, s/n, 46020 Valencia, Spain
caranvi1@upv.es

Abstract. In the real world, multiple dynamic biological phenomena present an intrinsic randomness due to their nature. One of the most common ways of modeling them is to use random differential or random difference equations, whose parameters are considered as random variables. However, since these are complex models, independence between these parameters is usually assumed just for simplicity, without even having tested this hypothesis in the phenomenon under study. On the other hand, the impossibility of solving the calibration of random models with classical deterministic optimization techniques has given rise to new stochastic calibration techniques, such as bio-inspired algorithms. In this paper, we present a calibration method based on the Multi-Objective Particle Swarm Optimization (MOPSO) algorithm of a random model with a set of random parameters without assuming independence between them. The calibration goal is to find the multivariate probability distribution of the random parameters vector that best captures the uncertainty of the data by minimizing two fitness functions. To show the value of the method, we will apply it to a simple first-order difference model for the evolution of the growth of breast cancer.

Keywords: random model · calibration · multivariate · Particle Swarm Optimization

1 Introduction

The use of differential equations has allowed us to better understand and predict physical phenomena. Nowadays, with the advances in computer science, even unknown solutions to complex equations can be estimated. However, it is important to note that when we build a model to describe real-life phenomena we are not able to capture this behavior without error. When building a model one often omits some details and makes assumptions about the model that could be wrong. Additionally, if we want our modeling to imitate real-life processes, the use of data is necessary. This data, which is often scarce, contains measurement

J. E. S. De Cursi (Ed.): UQSM 2023, LNME, pp. 112–121, 2024.
https://doi.org/10.1007/978-3-031-47036-3_10

errors caused both by instrumentation and by human errors. Hence, although traditional deterministic models have been well studied and documented, the introduction of uncertainty in models appears to be necessary to account for the errors and better understand the world around us. This uncertainty can be introduced by random models, which assume that the model parameters are random variables instead of deterministic values, and where the outcome is a stochastic process with a time-varying probability density function: the 1-PDF. With this function, the uncertainty of the random model is quantified, and it is possible to estimate statistics of interest such as the expected value, variance, confidence intervals, etc.

However, the introduction of uncertainty into models increases its difficulty. To simplify calculations the parameters are assumed to be independent in most models with uncertainty. Nevertheless, the testing of the independence assumption is often omitted although in many cases it can either not be proven. A good example of this is the Bertalanffy model which is used to model the weight of a fish at a time t. This model assumes that the change in weight depends on energy losses and energy acquisition [4]. If this model is considered as an initial valued problem, then another parameter is the initial weight of the fish. In this example it is natural to consider the parameters as a vector of random variables with an underlying correlation structure. This was done and calibrated using real data, giving an estimate of the covariance matrix which confirms that the parameters are indeed correlated [8].

If, more generally, independence between parameters is not assumed, then the random vector of model parameters follows a multivariate probability distribution, characterized by its joint probability density function (or joint PDF). However, this approach presents several challenges, such as the appropriate choice of the joint PDF or the calibration of its parameters from the data, i.e., to find those parameters of the joint PDF of the model parameter vector whose samples generate simulations well fitted to the real data. Faced with this fact, the classical deterministic optimization algorithms, unable to calibrate random models, have given way to a new generation of stochastic algorithms, such as bio-inspired algorithms [10]. However, despite the possibilities offered by this field, few researches are working without the assumption of independence in their models.

In this paper, we propose a calibration method of its parameters based on the bio-inspired Multi-Objective Particle Swarm Optimization (MOPSO) algorithm [3] oriented to random models with multivariate distributions, without assuming independence between their parameters. The calibration goal is to find the multivariate probability distribution of the random parameters vector that best captures the uncertainty of the data by minimizing several fitness functions. In Sect. 2, we formally expose the proposed a calibration method based on MOPSO algorithm for a general random model. In Sect. 3, we present a practical application case on a simple first-order difference equation which describes a tumor growth, and the obtained results. Finally, in Sect. 4, we present several conclusions.

2 Calibration Method

Suppose we have a random dynamic model $\mathcal{M}_\Theta(t),\ t > 0$ with a random vector of parameters $\Theta = (\theta_1, \ldots, \theta_k) \sim \mathcal{D}(p_1, \ldots, p_m)$, where $\mathcal{D}(p_1, \ldots, p_m)$ is a multivariate distribution with m parameters. If the explicit expression of D is known, we only need to estimate the parameters $(p_1, \ldots, p_m)$ to fully describe the behavior of Θ and, consequently, the model $\mathcal{M}_\Theta(t)$.

As this is a computational calibration method, the model evaluation is carried out for a discrete and finite set of time instants $T = \{t_0, t_1, \ldots, t_e\}$. Thus, let us denote the data sequence for the model calibration as $\mathbf{X^{data}} = \{X_t^{\text{data}} : t \in T\}$, and a set of N simulations of the model $\mathcal{M}_\Theta(t)$ as $\mathbf{X} = \{X_t^{(n)} :\ t \in T;\ n = 1, \ldots, N\}$. We define the goal of the calibration method as finding the parameters $(p_1, \ldots, p_m)$ for a distribution $\mathcal{D}(p_1, \ldots, p_m)$ that generate a set of model simulations $\mathbf{X}$ best approaching $\mathbf{X^{data}}$.

To quantify the goodness of fit between data and model simulations, we define two objective functions:

Standard deviation function:

$$F_\sigma(\mathbf{X}) = \sum_{t \in T} \sqrt{\frac{1}{N} \sum_{n=1}^{N} \left(X_t^{(n)} - \bar{X}_t \right)^2}, \tag{1}$$

Inside-outside function:

$$F_{io}\left(\mathbf{X}, \mathbf{X^{data}}\right) = \sum_{t \in T} d\left(X_t^{\text{data}}, \left[Q_{\alpha/2}(\mathbf{X}_t), Q_{1-\alpha/2}(\mathbf{X}_t)\right]\right), \tag{2}$$

where

$$\bar{X}_t = \frac{1}{N} \sum_{n=1}^{N} X_t^{(n)}, \tag{3}$$

$$\mathbf{X}_t = \left\{ X_t^{(n)} :\ n = 1, \ldots, N \right\}, \tag{4}$$

$$d(x, [q, Q]) = \begin{cases} 0 & \textbf{if } x \in [q, Q], \\ \min\{|x - q|, |x - Q|\} & \textbf{if } x \notin [q, Q], \end{cases} \tag{5}$$

and where $Q_\alpha(\mathbf{X}_t)$ is the α-order quantile function for a vector $\mathbf{X}_t$.

If we only calibrate with the *inside-outside function*, the algorithm will tend to generate distributions $\mathcal{D}(p_1, \ldots, p_m)$ with a high variance in order to widen its confidence interval (i.e. generate a larger space between the quantiles) and thus capture all the points of the $\mathbf{X^{data}}$ series within the confidence interval. Therefore we need to introduce a second objective function, the *standard deviation function*, in order to penalize distributions with high variances.

The behavior of both functions is antagonistic (orthogonal), since when we try to minimize the inside-outside function, we will necessarily widen the variability between simulations to obtain a wider confidence interval and, consequently, the standard deviation function will be larger. On the contrary, if we try to minimize the standard deviation function, i.e. the variability between simulations, the confidence interval between simulations will be narrower, thus leaving more points of the data series outside the interval and increasing the inside-outside function [11].

Therefore, we are faced with a multi-objective problem, where both objective functions, F_σ and F_{io}, are minimized simultaneously, and where the result of the optimization problem is not a single solution, but a set $P^* = \{(p_1, \ldots, p_m)_s : s = 1, \ldots, S\}$ of S Pareto-optimal solutions, whose solutions are not dominated by each other [1].

The parameter search by the calibration algorithm will be carried out in the domain

$$I = I_{p_1} \times \cdots \times I_{p_m},$$

where $I_{p_i} = [a_{p_i}, b_{p_i}]$ is the interval where each parameter p_i is defined. The domain can also be restricted to certain regions to speed up the search for good solutions.

To calibrate the model $\mathcal{M}_\Theta$ with two objective functions we use the Multi-Objective Particle Swarm Optimization (MOPSO) algorithm. In this algorithm, a solution or "particle" represents a vector $(p_1, \ldots, p_m)$. The search process goes as follows:

1. L particles each containing $P_l = (p_1, \ldots, p_n)$ for $l = 1, \ldots, L$ are randomly sampled from our parameter space I^m.
2. Given $\mathbf{X}^{\mathbf{data}}$ for each particle P_l:
   (a) N vectors Θ are sampled from $\mathcal{D}(P_l)$ to simulate $\mathbf{X}$ set.
   (b) $F_{io}(\mathbf{X}, \mathbf{X}^{\mathbf{data}})$ and $F_\sigma(\mathbf{X})$ are computed. P_l is then classified as local best, global best, or other. If $l = 0$ then it is considered a global best. Otherwise, P_l is considered a local best if it dominates in the Pareto sense P_{l-1} and a global best if it dominates $P_0, \ldots, P_{l-1}$.
   (c) If P_l is a local or global best, P_{l+1} is generated in one of two ways:
      i. Sample from I^m (10% probability).
      ii. Otherwise (with 90% probability), update P_l by adding a velocity term and then performing a mutation operation (with 10% probability) as done in [7]
3. Steps (a)-(c) are reiterated up to a maximum number of iterations M_{max} large enough to achieve convergence. The global Pareto-optimal solutions are returned.

3 Application: The Tumor Growth Case

3.1 Data

Data is extremely important to argue that a model we build replicates real physical phenomena. The data we study in this work is a time series of the size of tumours in mice. Although to better understand the dynamics of cancer in humans, data of human cancers will be ideal, it is not something readily available because many tumors are removed as soon as detected and of the ethical concerns of doing such a study. This data is, nevertheless, is an estimation of the overall behaviour of tumors through time. Experiments in mice have been of great impact for the understanding of illnesses and medication development in areas such as immunology and oncology [6,9].

In the data used, the rumors of breast cancer in mice was measured in mm^3 using a technique named xenograft, where human tissue was injected into mice. This technique is described in [5]. Our data is of the following form:

$$\mathbf{X^{data}} = \{X_t^{\text{data}}, t \in T\}$$

where $T = \{t_0, t_1, \dots, t_m\}$ is a series of organized time points and each X_t represents the tumor size at time t this data is shown in Table 1 and Fig. 1.

Table 1. Data series table [5].

Time t	Tumor size X_t in (mm^3)
0	45.74
16	194.257
30	675.14
33	936.53
43	1941.7
48	2558.6

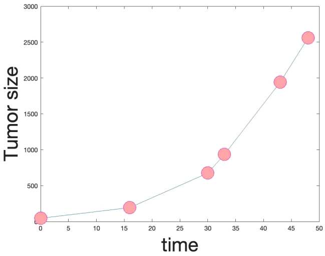


Fig. 1. Data series graph.

3.2 Model

Among the many dynamic models to study the tumour growth [2], the simplest ones are given by exponential and logistic equations. For this problem, according to the data, we choose to build the simplest model that describes the early stages of tumor growth, i.e., a first order linear model of the following form:

$$X_{t+1} = X_t(1+K), \quad X_0, K \in \mathbb{R}_+, \ t \in \mathbb{N}, \tag{6}$$

where X_0 is the size of the tumor at time $t=0$ and K is the tumour growth rate. Given this form of the model, the random sequence $\{X_1, X_2, \dots\}$ is built recursively. In this problem, we also assume that $\Theta = (K, X_0) \in \mathbb{R}_+^2$ is a random vector, so that $X_t, \forall t$ will be a random variable.

As the random vector is restricted to $\mathbb{R}_+^2$ domain, let $\Theta = (K, X_0) \sim$ Log-normal$(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where

$$\boldsymbol{\mu} = (\mu_K, \mu_{X_0})$$

is the mean vector, and $\boldsymbol{\Sigma}$ is the covariance matrix, with the form

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_K^2 & \rho\sigma_K\sigma_{X_0} \\ \rho\sigma_K\sigma_{X_0} & \sigma_{X_0}^2 \end{pmatrix}.$$

A log-normal distribution has been chosen because it is a way to restrict a normal distribution to $\mathbb{R}_+$ wihout adding more parameters to estimate. The joint probability density function (joint PDF) of the bivariate log-normal distribution is given by

$$f_\Theta(\theta) = f_{K,X_0}(k,x_0) = \frac{1}{2\pi k x_0 \sqrt{|\mathbf{\Sigma}|}} e^{-\frac{1}{2}\left(\log(k,x_0)-\boldsymbol{\mu}\right)^T \mathbf{\Sigma}\left(\log(k,x_0)-\boldsymbol{\mu}\right)} \tag{7}$$

Hence to fully describe our model $\mathcal{M}_\Theta$ defined in Equation (6), we need to estimate $(\mu_K, \mu_{X_0}, \sigma_K, \sigma_{X_0}, \rho)$ parameters of the log-normal distribution. To do so we apply to the problem the calibration scheme described in Sect. 2.

3.3 Calibration and Results

In this problem, the objective of the calibration is to find a set of Pareto-optimal solutions $P^* = \{P_s = (\mu_K, \mu_{X_0}, \sigma_K, \sigma_{X_0}, \rho)_s : \ s = 1, \ldots, S\}$. To bound the search domain of the PDF parameters, a nonlinear deterministic fit of the model to the data has been carried out, obtaining that

$$\theta^* = (k, x_0)^* = (0.0739, 85.568)$$

is the model parameters vector that minimizes the Root Mean Square Error (RMSE) of the difference between the model output and the data. We have also checked, by evaluating different values of k and x_0, that for $k \in [0.0682, 0.0785]$ and for $x_0 \in [66.446, 104.690]$, the RMSE deviates no more than 5% with respect to the minimum (see [11] for a more complete explanation). With this analysis, it also follows that the parameter k is more sensitive than x_0. Knowing that each bivariate log-normal distribution component satisfies that

$$\mathbb{E}[\Theta_i] = e^{\mu_i + \frac{\sigma_i^2}{2}}; \quad \mu_i = \log \mathbb{E}[\Theta_i] - \frac{\sigma_i^2}{2},$$

we establish that

- $\sigma_K \in I_{\sigma_K} = \left[10^{-5}, 10^{-2}\right]$ and $\sigma_{X_0} \in I_{\sigma_{X_0}} = \left[10^{-5}, 0.5\right]$ (a sufficiently wide interval for both standard deviations).
- $\mu_K \in I_{\mu_K} = \left[\log 0.0682 - \frac{\left(10^{-2}\right)^2}{2}, \log 0.0785 - \frac{\left(10^{-5}\right)^2}{2}\right] = [-2.686, -2.544]$, and $\mu_{X_0} \in I_{\mu_{X_0}} = \left[\log 66.446 - \frac{0.5^2}{2}, \log 104.690 - \frac{\left(10^{-5}\right)^2}{2}\right] = [4.071, 4.651]$. These intervals have been constructed by taking the k and x_0 values from the deterministic analysis of the model, and combining them with the deviation values that generate the wider search intervals.
- $\rho \in [-1, 1]$, as we know that correlation coefficient is always between $[-1, 1]$.

Thus, the search space is defined as

$$\begin{aligned} I &= I_{\mu_K} \times I_{\mu_{X_0}} \times I_{\sigma_K} \times I_{\sigma_{X_0}} \times I_\rho \\ &= [-2.686, -2.544] \times [4.071, 4.651] \times \left[10^{-5}, 10^{-2}\right] \times \left[10^{-5}, 0.5\right] \times [-1, 1]. \end{aligned} \tag{8}$$

For the execution of the calibration method, a total of $L = 60$ particles and $M_{max} = 10000$ maximum iterations were used, and the number of simulations of the model for each particle was $N = 1000$. Additionally, to ensure the replicability of the results, the random numbers generation (for simulations) has been carried out with a different seed for each MOPSO particle. The programming of the algorithm has been carried out in Python v.3.11 [11,12].

The Pareto front obtained in the calibration is shown in Table 2 and Fig. 2a. As can be seen, $S = 13$ Pareto-optimal solutions (global bests) have been found. It should be noted that the optimal values of means μ of the Pareto front are very similar (due to their restricted search interval), and that the optimal order of the deviations is mainly around 10^{-3} (for σ_K) and 10^{-2} (for σ_{X_0}), meanwhile the correlation coefficients are highly variable between -1 and 1.

Table 2. Pareto front solutions obtained by the MOPSO algorithm.

Solution	Particle $(\mu_K, \mu_{X_0}, \sigma_K, \sigma_{X_0}, \rho)$	Objective functions (F_{io}, F_σ)
1	$(-2.647, 4.581, 0.00757, 0.0255, -0.519)$	$(186.994, 153.679)$
2	$(-2.574, 4.307, 0.00575, 0.0592, 0.444)$	$(39.899, 422.790)$
3	$(-2.592, 4.352, 0.00303, 0.00525, -0.582)$	$(390.670, 42.442)$
4	$(-2.603, 4.454, 0.00821, 0.0181, -0.912)$	$(247.070, 68.855)$
5	$(-2.577, 4.152, 0.00679, 0.163, -0.812)$	$(0.000, 761.355)$
6	$(-2.600, 4.226, 0.00300, 0.00646, -0.897)$	$(1167.891, 21.987)$
7	$(-2.582, 4.316, 0.00185, 0.0530, 0.920)$	$(42.913, 368.820)$
8	$(-2.591, 4.275, 0.00671, 0.0991, 0.0131)$	$(13.378, 567.747)$
9	$(-2.580, 4.277, 0.000488, 0.0921, 0.890)$	$(15.034, 540.836)$
10	$(\mathbf{-2.558, 4.286, 0.00117, 0.0734, -0.290})$	$(\mathbf{28.276, 460.505})$
11	$(-2.607, 4.372, 0.00424, 0.0754, -0.559)$	$(42.691, 406.310)$
12	$(-2.590, 4.412, 0.00453, 0.0268, -0.333)$	$(96.463, 167.512)$
13	$(-2.569, 4.224, 0.00625, 0.102, -0.0636)$	$(9.356, 592.523)$

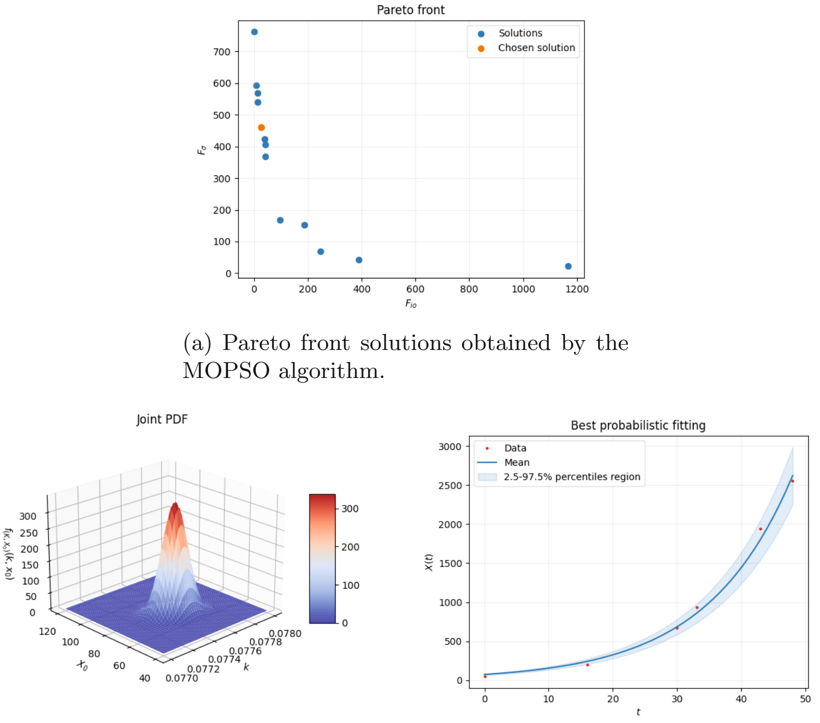
Often the solution that gives an equilibrium between both objective functions is chosen but in this work solution number 10 has been chosen, prioritizing slightly the inside-outside function over the variance one. With the chosen solution, the log-normal joint PDF of the model parameters $\Theta = (K, X_0)$ has been defined and represented in Fig. 2b, and N pairs of parameters have been sampled from it to simulate a set $\mathbf{X}$ of N simulations. From these simulations, it is possible to estimate the expected value as

$$\bar{X} = \left\{\bar{X}_t : t \in T\right\},$$

where $\bar{X}_t$ is the average of the simulations values at time t defined in Equation (3), and the 95% confidence interval as

$$\mathrm{CI}_{95\%} = \{[Q_{0.025}(\mathbf{X}_t), Q_{0.975}(\mathbf{X}_t)] : \; t \in T\},$$

where $\mathbf{X}_t$ is the simulations values vector at time t defined in Equation (4) and $Q_\alpha(.)$ is the α-order quantile function. The results of the probabilistic model fitting, shown in Fig. 2c, show that the random model, without assuming independence between its parameters, correctly captures the uncertainty of the data within its 95% confidence interval.



(a) Pareto front solutions obtained by the MOPSO algorithm.

(b) Best joint PDF.

(c) Best probabilistic model fitting

Fig. 2. .

4 Conclusions

In this work we have proposed and successfully applied a method of calibration of a random model - in our case, a tumor growth first-order difference equation - without assuming the hypothesis of independence between its random vector of parameters $\Theta = (K, X_0) \sim \text{Log-normal}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Since from the deterministic calibration of the model we have obtained information on the parameters expected values $\mathbb{E}[K]$ and $\mathbb{E}[X_0]$, related to the vector of means $\boldsymbol{\mu} = (\mu_K, \mu_{X_0})$, the challenge of the calibration has been to determine the width (in both dimensions) of the joint PDF, determined by the deviations σ_K and σ_{X_0}, and the correlation coefficient ρ, which determine the $\boldsymbol{\Sigma}$ matrix.

Given a wide search margin for deviations, the results have shown that the Pareto-optimal deviation values should be on the order of 10^{-3} for K and 10^{-2} for X_0. Such low deviation values suggest that the model is particularly sensitive to variations in its parameters, as the model is of the exponential type. However, the correlation coefficient, which is highly variable in the Pareto front solutions (given deviations within the same order), suggests that there is no preferred direction of variances in the distribution that generates better simulations than others. In other words, simulations with correlation coefficients very high or very close to 0 yield good solutions (for the same order in the values of the deviations, i.e., for similar distribution widths), so that the correlation coefficient would not be a determining parameter in the shape of the distribution. In our specific problem, independence between K and X_0 parameters is a hypothesis that can be assumed or not without consequences in the results: in both cases, we reach good solutions.

The main limitation of the method is that the objective functions are oriented to the data and simulations of the random model, but not to the shape of the joint PDF. It could be the case that, in other random models, the optimization algorithm generates Pareto-optimal solutions (distributions) where its density is concentrated around parameters' regions with low error, but omitting other good parameters' regions. In view of this, a possible improvement would be the introduction of a criterion of preference for wider distributions, either as a selection criterion for Pareto front solutions or as a new objective function for calibration. Another limitation is the lack of more time series that would allow a more robust statistical analysis, although since this is a tumor growth model for a specific patient, only a single record is available.

With all this, we believe that our proposal takes a step forward in the technique of stochastic fitting of random models without losing its generality by making strong assumptions such as the parameters independence.

References

1. Censor, Y.: Pareto optimality in multiobjective problems. Appl. Math. Optim. **4**(1), 41–59 (1977)
2. Adam, J.A., Bellomo, N.: A Survey of Models for Tumor-immune System Dynamics. Springer Science & Business Media, Boston (1997). https://doi.org/10.1007/978-0-8176-8119-7
3. Coello, C.A., Lechuga, M.S.: MOPSO: a proposal for multiple objective particle swarm optimization. In: Proceedings of the 2002 Congress on Evolutionary Computation. CEC 2002 (Cat. No.02TH8600). https://doi.org/10.1109/cec.2002.1004388
4. Lester, N.P., Shuter, B.J., Abrams, P.A.: Interpreting the von Bertalanffy model of somatic growth in fishes: the cost of reproduction. Proc. R. Soc. London. Ser. B Biol. Sci. **271**(1548):1625–1631 (2004). https://doi.org/10.1098/rspb.2004.2778, https://royalsocietypublishing.org/doi/abs/10.1098/rspb.2004.2778
5. Worschech, A., et al.: Systemic treatment of xenografts with vaccinia virus GLV-1h68 reveals the immunologic facet of oncolytic therapy. BMC Genom. **10**(1), 1–22 (2009)

6. De Groot, J.F., et al.: Tumor invasion after treatment of glioblastoma with bevacizumab: radiographic and pathologic correlation in humans and mice. Neuro-oncology **12**(3), 233–242 (2010)
7. Marini, F., Walczak, B.: Particle swarm optimization (PSO). A tutorial. Chemometr. Intell. Lab. Syst. **149**, 153–165 (2015). ISSN 0169-7439. https://doi.org/10.1016/j.chemolab.2015.08.020, https://www.sciencedirect.com/science/article/pii/S0169743915002117
8. Casabán, M.-C., Cortés, J.-C., Navarro-Quiles, A., Romero, J.-V., Roselló, M.-D., Villanueva, R.-J.: Computing probabilistic solutions of the Bernoulli random differential equation. J. Comput. Appl. Math. **309**, 396–407 (2017). ISSN 0377-0427.https://doi.org/10.1016/j.cam.2016.02.034, https://www.sciencedirect.com/science/article/pii/S0377042716300814
9. Masopust, D., Sivula, C.P., Jameson, S.C.: Of mice, dirty mice, and men: using mice to understand human immunology. J. Immunol. **199**(2), 383–388 (2017)
10. Molina, D., LaTorre, A., Herrera, F.: An insight into bio-inspired and evolutionary algorithms for global optimization: review, analysis, and lessons learnt over a decade of competitions. Cogn. Comput. **10**(4), 517–544 (2018)
11. Andreu-Vilarroig, C., et al.: Evolutionary approach to model calibration with uncertainty: an application to breast cancer growth model. In: Proceedings of the Genetic and Evolutionary Computation Conference Companion, pp. 1895–1901 (2022)
12. URL: https://www.python.org/

Mechanical Property Characterization of a 3D Printing Manufacturing System

Luiz H. M. S. Ribeiro[1](✉), Claus Claeys[2,3], Adriano T. Fabro[4],
Dimitrios Chronopoulos[2], and José R. F. Arruda[1]

[1] University of Campinas, Campinas, SP 13083-860, Brazil
luiz.marra@outlook.com
[2] KU Leuven, 3001 Leuven, Belgium
[3] Flanders Make@KU Leuven, Leuven, Belgium
[4] University of Brasilia, Brasilia 70910-900, Brazil

Abstract. Additive manufacturing is making it possible to increase the complexity of designed mechanical structures. However, the variability inherent to this manufacturing process can influence significantly the performance of structural elements, specially in phononic crystals and metamaterials since their working principles relies on the repetition of identical cells with a dedicated designed geometry. In this work, first, a design of experiments approach is applied to a determine a sampling strategy in order to characterize an additive manufacturing machine. Then, mechanical properties of the samples are inferred using material properties measured with an ultrasound transducer. The material density was measured using the weight of the samples, both dry and immersed in water, using the buoyancy force expression. It is known that the elastic modulus measured via ultrasound is biased. Therefore, the distributions inferred using ultrasound measurements were updated using experimental forced responses of sample rods and dynamic models via the Spectral Element Model. Updated values are used in statistical regression modeling to infer the stochastic field over print are of the 3D printer. The presented work is a first step in the longer term research goal: to show how to model the overall variability of a given additive manufacturing process, which is usually obtained in the statistical process control, and explain how to use it in the design of robust phononic crystal and metamaterial designs. The printing direction presented a statistically significant relationship with the elastic modulus and with the mass density, while only the printing direction presented a statistically significant relationship for the shear modulus.

Keywords: Uncertainty quantification · Statistical regression · Statistical inference · Kernel smoother

1 Introduction

Geometrically complex designs, which include metamaterials and phononic crystals, can be printed using additive manufacturing [14]. However, the variability of

J. E. S. De Cursi (Ed.): UQSM 2023, LNME, pp. 122–129, 2024.
https://doi.org/10.1007/978-3-031-47036-3_11

such manufacturing process can influence the printed structures substantively, specially the mechanical properties [16] are more impacted than what occurs typically in other manufacturing processes [15,17] and, thus, statistical process control can be a tool to assure that the manufacturing process will coincide with the design [18]. The inferred variability can be propagated through a deterministic model to obtain the stochastic result, and it can be used in a robust optimization approach as showed in [2].

The objective of this research is to show how statistical modeling can be applied to the data obtained from statistical process control to estimate stochastic fields that represent the variability of the mechanical properties of 3D printed parts. This estimation can be, for instance, combined with a robust optimization for designing phononic crystals and metamaterials that are robust against these types of variability.

2 Design of Experiments

In the current research, we assumed variability in the mass density (ρ), Young's modulus (E), and shear modulus (G). These variables were assumed as dependent variables, which could be tested if they are statistically related to the independent variables: thickness (T), printing position ($\boldsymbol{P} = \{P_x, P_y, P_z\}$), which is the position the parts are printed inside the 3D printer, printing direction ($\boldsymbol{D} = \{D_x, D_y, D_z\}$). The measurements were taken from samples defined as parallelepipeds. In addition, samples were taking over the weeks to see if the machine settings would influence the mechanical properties. For this paper a 3D Printer of the type "Prusa MK3S" with a print volume of approximately $11\,\mathrm{dm}^3$ and making use of the fused deposition modeling technique is studied. In addition, by taking samples over time, the assumption of the variability being the same over the time due to substantive changes in the used material and the setting were checked through statistical test.

Before the measurements, we assumed each observation for E follows a normal distribution with a mean of 2.1 and standard deviation of 0.5, as found by [3] in a similar manufacturing process, and consequently, the mean of samples follows a t-distribution. Assuming the α value was defined as 5%, and the type II error for 11 statistical degrees-of-freedom for a distribution X_A with mean $\bar{X}_A$ 1.5 distant from $\bar{X}$ is lower than 0.001%, and assuming that no more than four variables are going to be tested at the same time, the sample size of 15 is defined.

Each week, for 3 weeks, 15 samples were printed in one print, whose positions were defined by dividing the batch into 216 (6 in x direction, 6 in y direction, and 6 in z direction) smaller boxes of $30 \times 30 \times 30$ mm. Then, the printing position of each sample was defined by taking samples without replacement from a discrete uniform distribution $U(1, 6)$ for x, y, and z directions, defining, thus, the vector $\boldsymbol{P}$. Hence, rectangular parallelepipeds with dimensions of $1.5d_1$x1.5xd_1xd_1 mm, were defined, where the term $1.5d_1$ was sampled from the continuous uniform distribution $U(5, 25)$. Thus, the thickness can be defined as $T = d_1$. The variable time was also included in the analysis as week number (0, 1, 2).

For each sample, the values of E and G were observed using a Olympus 38DL Plus acquisition system with the M106 and V152 longitudinal and shear wave contact ultrasound transducer of 2.25 MHz. The mass density was estimated using the highly precise Acculab Atilon scale, where the samples' weights were measured dry and submerged in water. Once the estimations of E, G, and ρ are done, their statistical relationships with the independent variables can be made using statistical regression modeling.

3 Statistical Regression Models

For each dependent random variable, the vector of values $\boldsymbol{y}$ can be defined and modeled using the matrix of observations of independent variables weighted by the parameter vector $\boldsymbol{\beta}$ as presented in Eq. (1) [10].

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \tag{1}$$

Assuming that each element on vector $\boldsymbol{\varepsilon}$ follows a normal distribution with zero mean and variance σ^2, and that all these distributions are independent of each other, i.e., $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 I_{N_{iv}})$, where $I_{N_{iv}}$ is the identity matrix. Then, the maximum likelihood estimator to estimate $\boldsymbol{\beta}$ and σ^2 are respectively given by [9]

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}, \tag{2a}$$

$$\hat{\sigma}^2 = \frac{(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})}{N}. \tag{2b}$$

After that, a statistical test can be applied to test the hypothesis of β_j being statistically significant, i.e., $\beta_j \neq 0$ using a significance level of 5% [9].

The relationships between the dependent variables E, G, and ρ, with the independent variables T, $\boldsymbol{P}$, and $\boldsymbol{D}$ were hence verified. Variables $\boldsymbol{D}$ were included as dummy variables ($\delta(D_c)$ with $c = \{x, y, z\}$) [11], where 0 indicates the absence and 1 presence of these variables. All the combinations of the independent variables at the power of one, two, and three were used in the analyses. The Akaike information criterion [7,8] was in the selection of the models that presented statistically significant parameters and presented valid assumptions.

The Breusch-Pagan test [4] was used to check if the variance is the same for all the terms in the vector $\boldsymbol{\varepsilon}$, the Shapiro-Wilk test [5] was used to verify the normality, and the Durbin-Watson [6] test was used to verify the assumption of independency on the residuals.

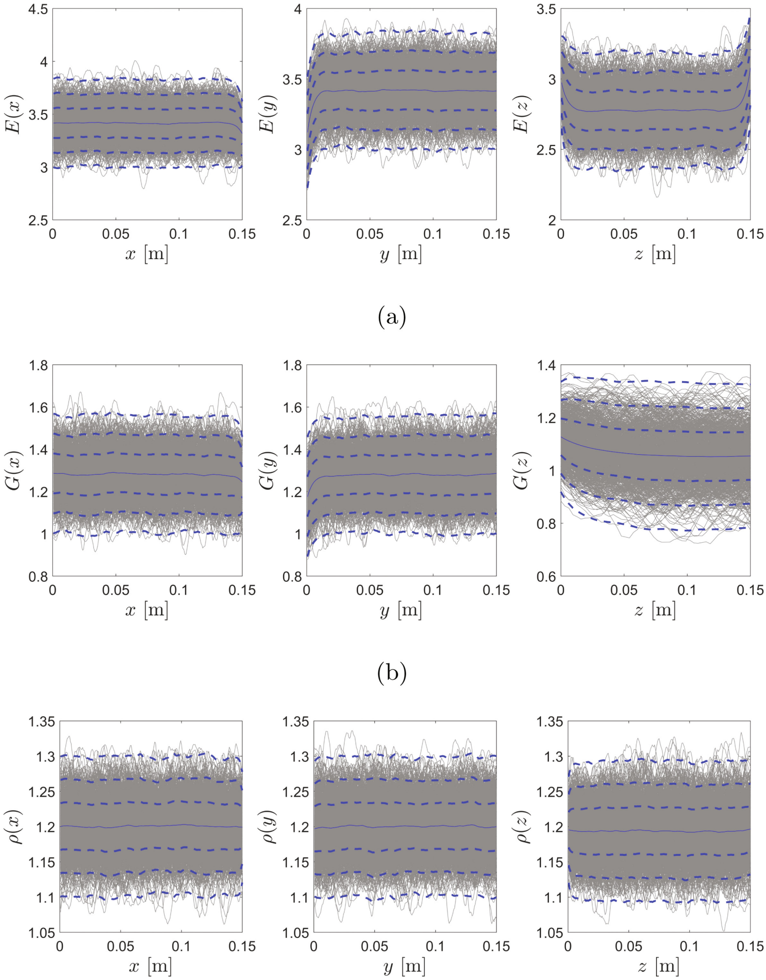
Using the proposed methodology, the stochastic regression models in Eq. (3) were obtained.

$$E_m \sim \mathrm{N}(2.9740 + 0.34445 \times (\delta(P_x) + \delta(P_y)), 0.4035), \tag{3a}$$

$$G_m \sim \mathrm{N}(1.1247 + 0.1200 \times (\delta(P_x) + \delta(P_y)), 0.2718), \tag{3b}$$

$$\rho_m \sim \mathrm{N}(1.1956 + 0.0036 \times (\delta(P_x) + \delta(P_y)), 0.0974). \tag{3c}$$

As can be observed by the estimated equations, the parts printed in the z direction have significantly lower Young's modulus, shear modulus, and mass density than the parts printed in x and y directions.



(a)

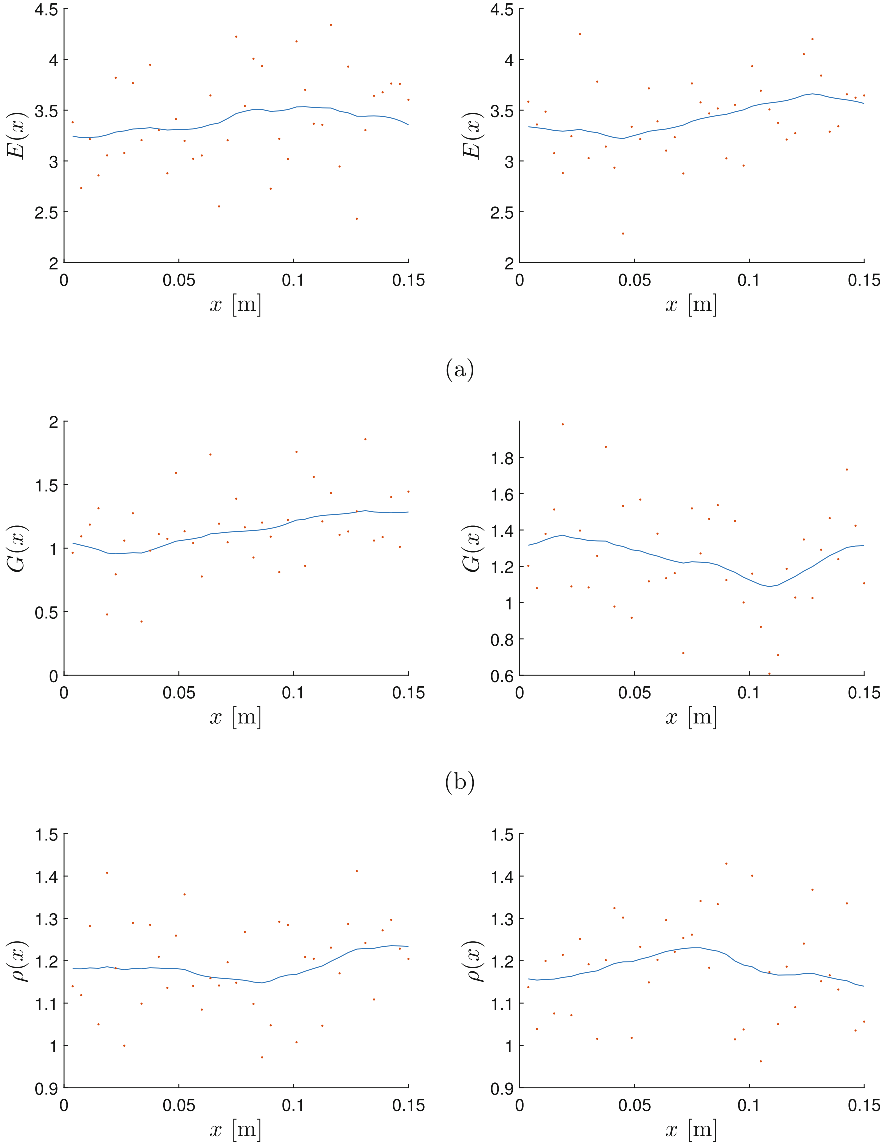
(b)

(c)

Fig. 1. Illustration of two simulated field of $\boldsymbol{E}_{3d}$ (a), $\boldsymbol{G}_{3d}$ (b), and $\boldsymbol{\rho}_{3d}$ (c) with 1,000 samples each. The straight line is the field mean, the dashed lines are the intervals that contains one, two, and three standard deviations.

3.1 Kernel Smoother

Using the models in Eq. (3), discrete samples of E, G, and ρ in the three-dimensional space can be defined as the vectors of length n $\boldsymbol{E}_{3d}$, $\boldsymbol{G}_{3d}$, and $\boldsymbol{\rho}_{3d}$. When premultiplying one of those vectors by the Kernel matrix $\boldsymbol{K}$, the sampled vector can be made smoother, whose r-th row and s-th column element is given by [1]



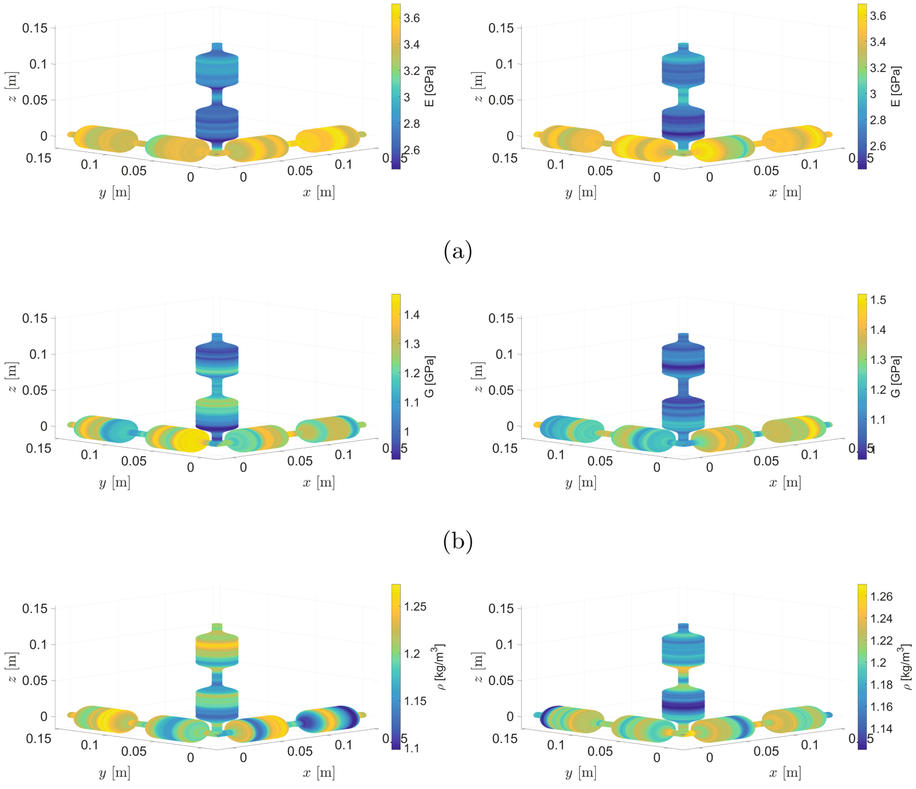
(a)

(b)

(c)

Fig. 2. Two raw (red dots) then smoothed (blue lines) samples of the stochastic fields of Young's modulus (a), shear modulus (b), and mass density (c). (Color figure online)

$$K_{rs} = \frac{f_K\left(\frac{d_{3d,r} - d_{3d,s}}{\zeta}\right)}{\sum_{s=1}^{n} f_K\left(\frac{d_{3d,r} - d_{3d,s}}{\zeta}\right)} \tag{4}$$



(a)

(b)

(c)

Fig. 3. Illustration of two smoothed samples of $\boldsymbol{E}_{3d}$ (a), $\boldsymbol{G}_{3d}$ (b), and $\boldsymbol{\rho}_{3d}$ (c) with $n = 100$.

where the function $f_K(d)$ can be an exponential function of the spatial distance (d) between $d_{3d,r}$ and $d_{3d,s}$ with correlation length ζ [12,13].

The 1,000 samples of the spatial E, G, and ρ sampled from Eq. (3) and smoothed using the Kernel smoother, whose elements are given by Eq. (4), are illustrated in Fig. 1.

Two samples of the smoothed $\boldsymbol{E}_{3d}$, $\boldsymbol{G}_{3d}$, and $\boldsymbol{\rho}_{3d}$ are illustrated in Fig. 2. For large enough samples, after convergence of the Monte Carlo method, the three-dimensional spatial field of E, G, and ρ can be simulated. Figure 3 illustrates two samples of the smoothed field of the mechanical properties on a specific frame structure.

4 Final Remarks

In the current research, we have used some data, simulating the data from a statistical process control from a manufacturing process, and we have shown how to infer and estimate the variability via stochastic field. The estimated stochastic field can be used to check if the manufacturing process is under control, and it can be used in robust optimization via combining engineering design and statistical process control. In the current research, we found a significant lower Young's modulus, shear modulus, and mass density for the parts printed in z direction than the ones printed in x and y directions.

Acknowledgments. The authors gratefully acknowledge the financial support of the São Paulo Research Foundation (FAPESP) through processes number 2020/15328-5, 2019/00315-8, 2018/18774- 6 and 2018/15894-0, and the Brazilian National Council of Research CNPq (Grant Agreement ID: 305293/2021-4).

References

1. Ribeiro, L.H.M.S., Dal Poggetto, V.F., Arruda, J.R.F.: Robust optimization of attenuation bands of three-dimensional periodic frame structures. Acta Mech. **223**(2), 445–475 (2022). https://doi.org/10.1007/s00707-021-03118-x
2. Cantero-Chinchilla, S., Fabro, A.T., Meng, H., Yan, W., Papadimitriou, C., Chronopoulos, D.: Robust optimised design of 3D printed elastic metastructures: a trade-off between complexity and vibration attenuation. J. Sound Vib. **529**, e116896 (2022). https://doi.org/10.1016/j.jsv.2022.116896
3. Sousa, A.M., Pinho, A.C., Piedade, A.P.: Mechanical properties of 3d printed MouthGuards: influence of layer height and device thickness. Mater. Des. **203**, e109624 (2021). https://doi.org/10.1016/j.matdes.2021.109624
4. Breusch, T.S., Pagan, A.R.: A simple test for heteroscedasticity and random coefficient variation. Econom. J. Econom. Soc. 1287–1294 (1979). https://doi.org/10.2307/1911963
5. Royston, J.P.: An extension of Shapiro and Wilk's W test for normality to large samples. J. Roy. Stat. Soc.: Ser. C (Appl. Stat.) **31**(2), 115–124 (1982). https://doi.org/10.2307/2347973
6. Durbin, J., Watson, G.S.: Testing for serial correlation in least squares regression. Biometrika **37**(3), 409–428 (1950). https://doi.org/10.2307/2332391
7. Sakamoto, Y., Ishiguro, M., and Kitagawa, G.: Akaike information criterion statistics. Dordrecht, The Netherlands: D. Reidel. **81**(10), 26853 (1986). https://doi.org/10.1080/01621459.1988.10478680
8. Akaike, H.: Information theory and an extension of the maximum likelihood principle. In: Selected Papers of Hirotugu Akaike, pp. 199–213 (1998). https://doi.org/10.1007/978-1-4612-1694-0_15
9. Montgomery, D.C., Peck, E.A., Vining, G.G.: Introduction to Linear Regression Analysis. John Wiley & Sons, Hoboken (2012). https://doi.org/10.1111/insr.12020_10
10. Ribeiro, L.H.M.S., Beijo, L.A., Salgado, E.G., Nogueira, D.A.: Bayesian modelling of number of ISO 9001 issued in Brazilian territory: a regional and state level analysis. Total Qual. Manage. Bus. Excell. **33**(9), 1183–1212 (2022). https://doi.org/10.1080/14783363.2021.1944083
11. Draper, N.R., Smith, H.: Applied Regression Analysis. John Wiley & Sons, Hoboken (1998). https://doi.org/10.1002/9781118625590

12. Aydin, D.: A comparison of the nonparametric regression models using smoothing spline and kernel regression. World Acad. Sci. Eng. Technol. **36**, 253–257 (2007). https://doi.org/10.5281/zenodo.1332448
13. Ribeiro, L.H.M.S., Dal Poggetto, V.F., Beli, D., Fabro, A.T., Arruda, J.R.F.: Quantifying spatial uncertainty and inferring the stochastic wave attenuation. In: Proceedings of COBEM 2021 (2021)
14. Gattin, M., Bochud, N., Rosi, G., Grossman, Q., Ruffoni, D., Naili, S.: Ultrasonic bandgaps in viscoelastic 1D-periodic media: mechanical modeling and experimental validation. Ultrasonics. e106951 (2023). https://doi.org/10.1016/j.ultras.2023.106951
15. Santoro, R., Mazzeo, M., Failla, G.: A computational framework for uncertain locally resonant metamaterial structures. Mech. Syst. Signal Process. **190**, e110094 (2023). https://doi.org/10.1016/j.ymssp.2023.110094
16. Santoro, R., Mazzeo, M., Failla, G.: Wave attenuation and trapping in 3D printed cantilever-in-mass metamaterials with spatially correlated variability. Sci. Rep. **9**(1), e5617 (2019). https://doi.org/10.1038/s41598-019-41999-0
17. Tsung, F., Zhang, K., Cheng, L., Song, Z.: Statistical transfer learning: A review and some extensions to statistical process control. Qual. Eng. **30**(1), 115–128 (2018). https://doi.org/10.1080/08982112.2017.1373810
18. He, K., Qian, Z., Yili, H.: Profile monitoring based quality control method for fused deposition modeling process. J. Intell. Manuf. **30**, 947–958 (2019). https://doi.org/10.1007/s10845-018-1424-9

Stochastic Analysis Involving the Computational Cost of a Monte-Carlo Simulation

Héctor E. Goicoechea(✉), Roberta Lima, and Rubens Sampaio

Departamento de Engenharia Mecânica, Pontifícia Universidade Católica do Rio de Janeiro, Rua Marquês de São Vicente 255, Gávea, Rio de Janeiro, RJ 22451-900, Brazil
h.e.goicoechea@gmail.com

Abstract. To present ideas, a model problem consisting of a moving mass-belt system with random friction showing the stick-slip phenomenon is treated. The dynamics is simulated. The objective of this work is to assess the behaviour of the computation cost in terms of the run-time, which is random, and its relationship with some of the output variables that define the dynamical behaviour of the mechanical system, such as the duration of the phases present in the simulation, sticks and slips, and the number of phases that occur in each realisation. All this is analysed from a stochastic perspective. However, the probabilistic model to analyse the distribution of a three-dimensional random vector, formed by the run-time, duration and number, belongs to R^4, thus it is difficult to characterise and visualise. Hence, in this study, the use of random variable transformations to produce new independent variables is explored as an attempt to reduce the number of dimensions that need to be considered. Also, the change of variables is used to assess the link between the behaviour of the results and the chosen integration method. It is shown that the predictions obtained with the Monte Carlo method combined with a Multiple Scales analytical approximation are influenced by the number of transition phases rather than their durations.

Keywords: computational costs · stochastic run-time · multiple Scales method

1 Introduction

The Monte Carlo (MC) method is an important tool to deal with stochastic problems, as it can be used to construct statistical models for random object transformations [1,2]. The method deals with the stochastic problem by partitioning it into many deterministic ones where, in each, a realisation of the random input is used. First, a random sample of the input is generated. After collecting a sufficiently large number of realisations, the next step involves transforming each realisation of the sample according to some mathematical transformation. The method is based on the law of large numbers, hence a large amount of realisations is required [3] to assure convergence of the results. For the current application,

J. E. S. De Cursi (Ed.): UQSM 2023, LNME, pp. 130–140, 2024.
https://doi.org/10.1007/978-3-031-47036-3_12

the transformation is given by the equations of motion of the oscillator described in [4]. A sample of the output is obtained by collecting the results obtained in each of the realisations. Thus, these output parameters are also random variables that need to be stored and that, in the end, will be used to construct the statistical model.

Even though each realisation is deterministic in nature, neither the inputs, the outputs, nor the computational costs are. The uncertainty associated with the inputs is propagated to the output parameters, as shown in [5,6]. Moreover, the elevated number of calculations required to assure an accurate statistical model makes the MC method a big data problem, especially when the transformation is given by a differential equation that is solved by numerical integration, as shown in [7]. Taking into account that the computational resources are limited, the computational costs and, particularly, the total run-time can be of the utmost importance.

In this paper, the run-time associated with the simulation of a random oscillator is studied from a stochastic perspective. This is achieved by comparing the results of the MC method combined with a Runge-Kutta numerical integration scheme, and the MC combined with an analytical approximation based on the multiple scales method. Up to the authors' knowledge, this study is a novelty given that most papers concerning stochastic simulations ignore the fact that the run-times are also of stochastic nature, [8,9]. Eventually, the run-times play a role in determining whether a stochastic analysis is feasible or not, and their behaviour could have an impact on the efficient assignment of the available resources.

In this paper, the influence of the integration method in the run-times associated with the Monte Carlo simulations is explored. For this task, the results considering three different integration strategies are compared: an analytical approach based on the Multiple Scales method, a Runge-Kutta numerical scheme with variable time-step, and a Runge-Kutta approach with fixed time-step. The idea of comparing the different methods is also assessing how the features of each integration technique affect the stochastic behaviour of the run-times.

2 The Mass-Belt Model

As it was stated, the system that is used for the analysis is described in details [4]. A brief description of the system, without providing all the details of the formulation involved, is given herein to grasp the basics properties of the system that will be used.

The dynamical problem is that of a mass-spring-damper that moves over a belt is analysed. A sketch of the system is given in Fig. 1, along with a sketch depicting the general characteristics of the friction model that is use. For the present application, the belt in contact with the mass-spring-damper system moves at a constant speed. The equation that described the dynamics of the system in terms of y, the position of the mass, is given by

$$m\,\ddot{y}(t) + \gamma\,\dot{y}(t) + k\,y(t) = f_{at}\,, \tag{1}$$

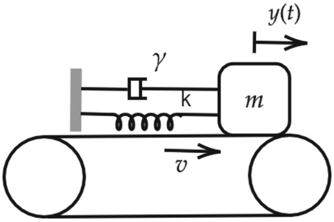


Fig. 1. Mass-spring-damper system with dry friction.

where m, k and γ are the mass, the spring constant, and the damping coefficient. The friction force is indicated as f_{at} in that equation and its behaviour is depicted in Fig. 2.

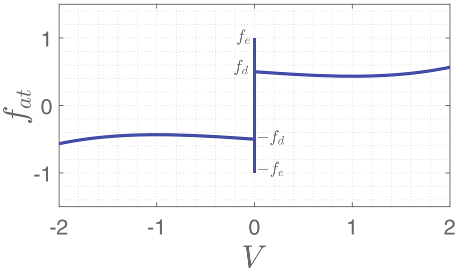


Fig. 2. Friction force model with $a = 0.1$ and $f_d = 0.5$.

In [4] a description of the general behaviour of the system is also presented. It was shown that the dynamics present stick-slip oscillations. Thus, the behaviour can be mathematically expressed in terms of a piece-wise function where stick-slip phases alternate. The characteristics of a stick-phase, which in this problem imply that the mass moves at a constant speed over some period, assure the existence of an exact solution for those periods. However, during the slip phases, due to the non-linear behaviour of the system, an exact solution can no longer be obtained. In that context, an analytical approximation was developed using the multiple scales method. In the end, [4] provides an approximation to the solution of the problem given by a piece-wise function that combines an exact solution for the slip phases with the multiple-scales analytical approximation for the slip ones.

Another alternative to find an approximation to the solution of this problem is to use a numerical integration scheme, such as the Runge-Kutta method. In this type of approximation the solution is advanced in time by taking small time-steps. However, this characteristic usually makes the method more demanding than the analytical approach from [4].

The idea of this paper is to quantify the computational cost of each approach when a stochastic problem is considered. For that purpose, the Monte Carlo method is used (see [2] for a detailed description of the method). In this paper, the dynamic friction force is modelled as a random variable given by a uniform distribution with support $[0.8, \quad 1.8]$. The variability in the friction influences the output variables of interest, such as the transition instants, the phase duration, the number of sticks, the position of the system, and the computational cost.

3 Results and Discussions

The stochastic problem is tackled by combining the Monte Carlo method with three different approximation strategies to integrate the equation of motion. These strategies are:

- AN: Monte Carlo combined with a Multiple Scale method analytical approximation;
- NV: Monte Carlo combined with a numerical integration scheme based on the Runge-Kutta method of 4^{th} and 5^{th} order. This case uses an automatic variable time-stepping scheme;
- NF: Monte Carlo combined with a numerical integration scheme based on the Runge-Kutta method of 4^{th} order with fixed time step.

The numerical and analytical approximations were simulated with the parameters shown in Tab. 1. The simulated time is $[0\,, 2000\,]$s. A total number of 10^5 realisations were used for the approaches AN and NV, and due to time constraints, $4 \cdot 10^4$ realisations were used for the NF approach.

Table 1. Parameters used in the simulations.

Parameter	Value	Unit	Parameter	Value	Unit
m	1	Kg	k	0.1	N/m
v	-2	m/s	$y(0)$	1	m
γ	1	(N s)/m	$\dot{y}(0)$	4	m/s
a	0.1	(Kg s)/(m^2)	f_e	2	N
ϵ	0.0001	-	g	9.81	m/s^2

The normalised histograms associated with the run-times of each approach are depicted in Fig. 3 for a) AN; b) NV and c) NF. By direct inspection of the results, it is observed that the support is orders of magnitude different. This is supported by the statistics associated with these histograms, which are presented in Table 2. Differences in the shape and the statistics of these distributions are found. NV took, in mean, $\approx$ 10 times longer to simulate than AN, and NF took $\approx$ 300 times longer than AN. These differences are exclusively due to characteristics intrinsic to each of the approximation strategies, given that the same input sample for the friction coefficient was used for the realisations in the three cases.

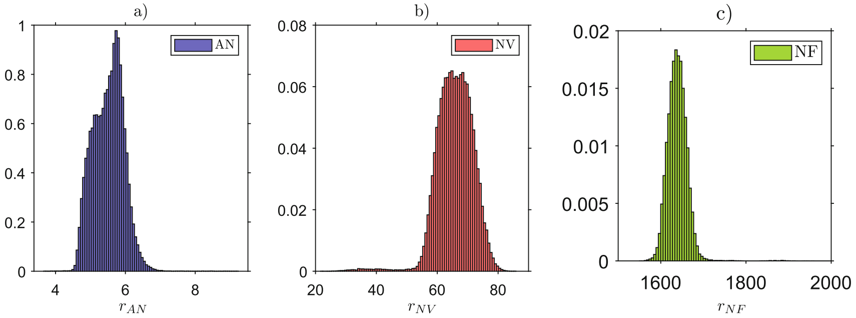


Fig. 3. Normalised histograms for the run-times obtained with MC combined with: a) AN, b) NV e c) NF.

Table 2. Some statistics for the variables R_{AN}, R_{NV} e R_{NF}.

Approach	Mean value	Standard deviation	3^{rd} order moment	4^{th} order moment
AN	5.51 s	0.43 s	0.16	3.62
NV	65.80 s	6.02 s	-1.08	7.69
NF	$1,64 \cdot 10^3$ s	$54,21$ s	$23,01$	$718,53$

To understand why those differences occurred, in what follows the run-times and their relations with other characteristic variables in a stick-slip problem are further analysed. To do this, the run-times are discriminated into two groups: the ones associated with the stick phases and the ones linked with the slip phases.

3.1 Further Understanding the Behaviour of the Run-Times with An

The run-times associated with the AN approach are depicted in Fig. 4, where a) shows the time taken to complete the integration of the stick phases; b) shows the time taken to complete the integration of the slip phases; and c) the total run-time which is the sum of the two previous.

When analysing the support, it is observed that the run-time associated with the stick phases is almost negligible when compared to those of the stick phases. In fact, the stick phases are calculated ≈ 1000 times faster than the slip ones. The reason for this difference lies in the integration strategy itself. For instance, the stick phases are governed by a simple equation (see [4] for details), and they correspond to a motion with constant speed. An exact solution for this phase is easy to obtain, as well as the time instant where the phase ends, which can be calculated in a straightforward manner from the same equation by isolating t. In contrast, with the slip phases the equation of motion is approximated using a Multiple Scales expansion where the initial parameters of the piecewise approximation need to be calculated. This process now requires finding the roots of a system of non-linear system of equations, as indicated in [10], at the beginning of each phase. In addition, finding the transition instant where a phase ends also requires solving the roots of a non-linear equation. Given that these manipulations are more demanding in terms of computation costs, it is expected that they have an impact on the differences in the run-time between the stick and the slip phases.

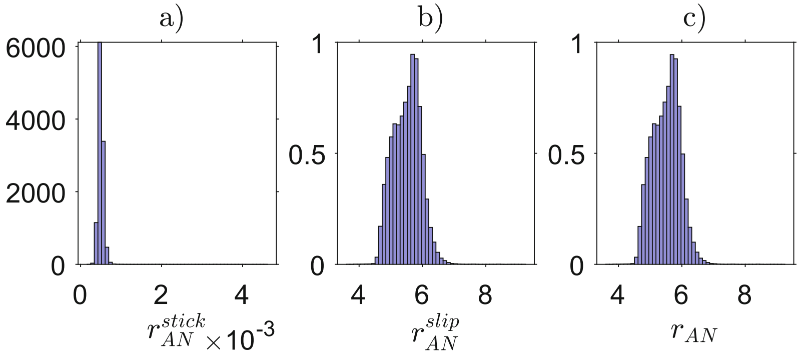


Fig. 4. Normalised histograms of the run-times for a) the stick phases; b) the slip phases; and c) the total run-time. All correspond to the simulation case AN.

To further understand the behaviour of these run-times, in what follows the random vectors $[R_{AN}^{stick}, D^{stick}]$, $[R_{AN}^{stick}, N^{stick}]$ and $[R_{AN}^{slip}, D^{slip}]$, $[R_{AN}^{slip}, N^{slip}]$ are analysed. For this an analysis, a transformation of random variables will be used. The justification for this transformation comes from observing the behaviours depicted in Fig. 5. In the figure, a superior view of the joint histogram of some of these random vectors, and the height of the histogram is represented with a colourmap. On top of these graphs, linear regressions represented in solid black lines were overlayed.

When observing the behaviour of the data of these histograms with respect to the regression line, some tendencies are observed. For instance, the histograms associated with $[R_{AN}^{stick}, N^{stick}]$ and $[R_{AN}^{slip}, N^{slip}]$ tend to follow the direction of the regression line, a behaviour which is not observed with either $[R_{AN}^{stick}, D^{stick}]$ nor $[R_{AN}^{slip}, D^{slip}]$. This tendency leads to formulating the hypothesis that, by considering a transformation of variables, that is, considering a different frame of reference to evaluate the responses, more information about the dependency among these random vectors can be stated. In fact, it is sought to evaluate if, for a given linear transformation, the new variables behave as independent stochastic variables.

To assess the previous hypothesis, the following general transformation rule given by π_1 is considered.

$$\pi_1 : \begin{cases} r_2 = r - \tilde{r} = r - \alpha_2 n - \beta_2 \\ n_2 = n \\ d_2 = d \end{cases} \tag{2}$$

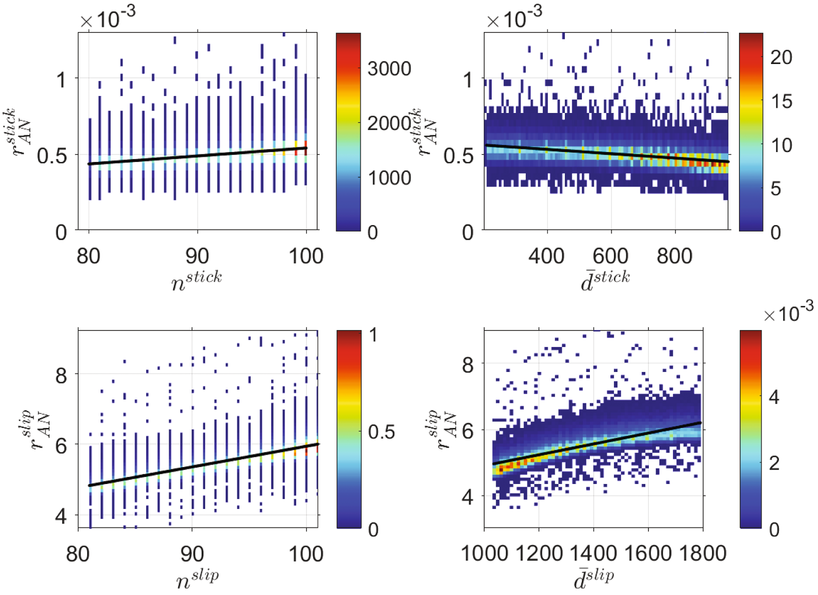


Fig. 5. Superior view of the joint histograms of the random vectors associated with the An approach: a) r_{AN}^{stick} e n^{stick}; b) r_{AN}^{stick} e d^{stick}; c) r_{AN}^{slip} e n^{slip}; e d) r_{AN}^{slip} e n^{slip}.

where α_2 and β_2 are the parameters associated to the regression, r_2 is a new transformed variable, r is the run-time (either of the stick or the slip phase), n is the number of phases and d their duration.

In the present study, the following two expressions for the regressions in the form of $\tilde{r} = \alpha_2 n + \beta$ are used

$$\tilde{r}_{AN}^{stick} = 5,2 \cdot 10^{-6}\; n^{stick}\; + 0.21 \cdot 10^{-4} \tag{3}$$

and

$$\tilde{r}_{AN}^{slip}\; = \; 5,9 \cdot 10^{-2}\, n^{slip}\; + 0.071. \tag{4}$$

The following results are obtained in the transformed space. Figure 7 depicts the results for the transformed vectors $[R2_{AN}^{stick}, D_2{}^{stick}]$, $[R2_{AN}^{stick}, N_2{}^{stick}]$ and $[R2_{AN}^{slip}, D_2{}^{slip}]$, $[R2_{AN}^{slip}, N_2{}^{slip}]$, obtained with π_1 and (3).

The graph shows very interesting results. To the left, in a) and c), the joint normalised histograms obtained with the transformed variables in the new coordinate system are depicted. To the right, in b) and d), the products of the marginals are shown. It is observed that in the new variable space, the resulting vector $[R2_{AN}^{stick}, N_2{}^{stick}]$ is independent in the stochastic sense. But a new and interesting result was found: with this transformation also $[R2_{AN}^{stick}, D_2{}^{stick}]$ is independent.

An analogue observation can be found when analysing $[R2_{AN}^{slip}, {N_2}^{slip}]$ and $[R2_{AN}^{slip}, {D_2}^{slip}]$, whose joint histograms are depicted in Fig. 6 a) and c), and the histogram associated with the product of the marginals in b) and d). The results in this graph are obtained using π_1 and (4).

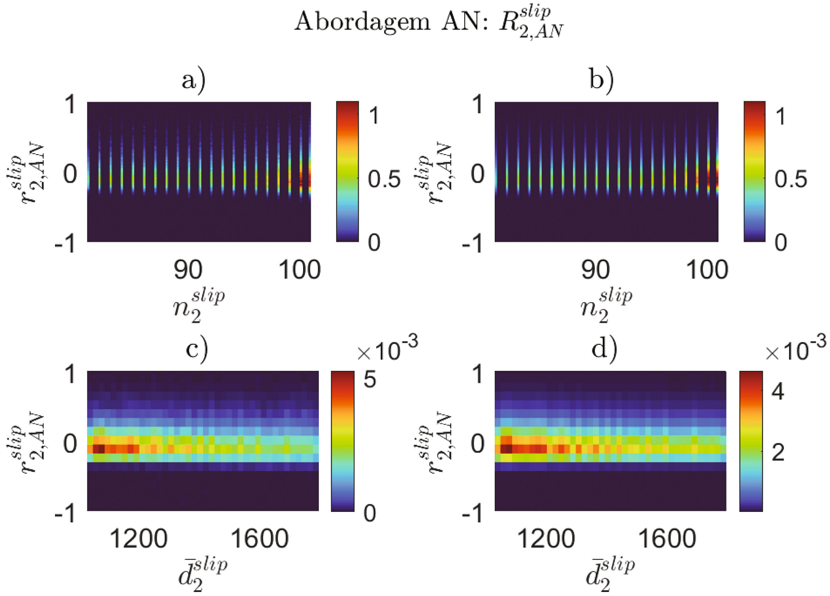


Fig. 6. A superior view of the histograms of the random vectors associated with slip phases in the AN approach is depicted for figures a) and c), while a representation constructed using the product of the marginals is shown in b) and d).

An interesting and unexpected result is found with the behaviours of $[R2_{AN}^{stick}, {D_2}^{stick}]$ and $[R2_{AN}^{slip}, {D_2}^{slip}]$, which now are independent from a stochastic perspective. Although it is not explored herein, the reason for this behaviour could lie in the fact that the durations and the number of phases are deterministically related. Therefore, eliminating the dependency in the stochastic sense with one of the two, in this case, the number of phases, may also eliminate the dependency with the duration. But this observation is left as a hypothesis to be explored in further studies.

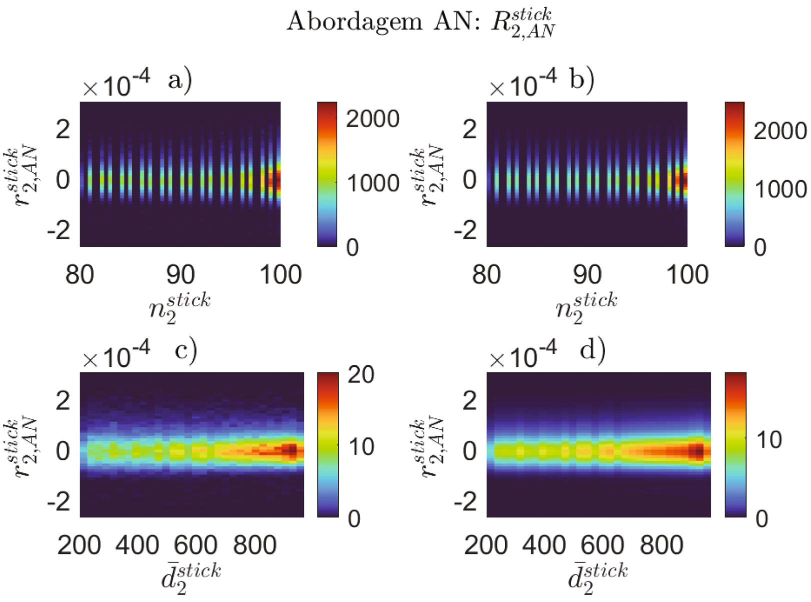


Fig. 7. A superior view of the histograms of the random vectors associated with stick phases in the AN approach is depicted for figures a) and c), while a representation constructed using the product of the marginals is shown in b) and d).

4 Conclusions

In this study, the problem of an oscillator that exhibits stick-slip was addressed. Stochastic dry friction was considered, with the dynamic friction modelled as a uniform probability distribution. The focus of this study was set on the computational aspects of the stochastic simulations.

Considering the deterministic problem, which is non-linear due to the dry friction, an exact closed-form solution does not exist, and approximation techniques must be used. As a consequence, the stochastic problem was solved by means of three strategies: a combination of the Monte Carlo method with numerical integration with variable time-step, Monte Carlo with numerical integration with fixed time-step, and the combination of the Monte Carlo method with an analytical integration, where an approximation is obtained by using the Multiple Scale method.

With the numerical integration, an approximation was obtained by discretising the equation in the time domain following the Runge-Kutta integration scheme. The method relies on taking small time-steps to guarantee the accuracy of the solution. Whereas, with the analytical integration, an approximation given by a closed-form expression can be obtained for the stick and the slip phases.

The analytical approximation is constructed as a piecewise function, due to the sudden change in the behaviour between stick and slip phases. But once the general form of the closed-form expression is known, only the information at the transition instants and the initial conditions at the beginning of each phase is required to completely define the behaviour of the system at any time. This has a direct impact on the computational costs involved with the analytical approximation, and it is advantageous if compared to the numerical one where many intermediate time steps in between the transition instants are required to assure accuracy in the approximation. Moreover, it can be a decisive factor in defining the feasibility of a stochastic study, given the large number of realisations that need to be performed, and the run-time involved in such calculations.

All these previous aspects are reflected in the results obtained for the stochastic oscillator problem. To understand the link between the run-times and the integration strategies, joint histograms considering the run-time and some characteristic variables for the problem, in this case, the duration and the number of phases were analysed. After that, a change of variables was proposed, and it was shown that the new variables are independent, from a stochastic perspective. This allowed us to further understand the relationship between the different behaviours in the run-time and the adopted integration strategy. It was observed that the results obtained with the analytical approximations were influenced by the number of phases rather than by their durations. This is associated with how an analytical approximation is constructed, and the way each of the phases is solved without using intermediate time-steps. It is left for future work to produce a similar analysis for the other approaches with the numerical integration schemes.

Acknowledgments. The authors acknowledge the support given by FAPERJ, CNPq and CAPES.

References

1. de Cursi, E.S., Sampaio, R.: Uncertainty Quantification and Stochastic Modeling with MATLAB, vol. 67, 1st edn. ISTE Press Ltd., Elsevier Ltd., London, UK (2015)
2. Sampaio, R., Lima, R.: Modelagem Estocástica e Geração de Amostras de variáveis e Vetores Aleatórios, vol. 70. SBMAC - Notas em Matemática Aplicada, São Carlos - SP (2012)
3. Sobol, I.M.: A Primer for the Monte Carlo Method, 1st edn. CRC Press, Boca Raton (1994)
4. Gomes, M., Goicoechea, H.E., Lima, R., Sampaio, R.: Stochastic evaluation of the run-time of a stick-slip oscillator problem. In *Mecánica Computacional*, vol. 39, Bahía Blanca, 2022. Asociacion Argentina de Mecánica Computacional (2022)
5. de Cursi, E.S., Sampaio, R.: Modelagem Estocástica e Quantificação de Incertezas, vol. 66. SBMAC - Notas em Matemática Aplicada, São Carlos - SP (2012)
6. Lima, R., Sampaio, R.: What is uncertainty quantification? J. Braz. Soc. Mech. Sci. Eng. **40**, 155 (2018)
7. Lima, R., Sampaio, R.: Random stick-slip oscillations in a multiphysics system. Eur. Phys. J. Plus **136**(8), 879 (2021)

8. Wilhelm, R., et al.: The worst-case execution-time problem-overview of methods and survey of tools. ACM Trans. Embed. Comput. Syst. **7**(3), 1–58 (2008)
9. Lee, M., Jeon, J., Bae, J., Jang, H.-S.: Parallel implementation of a financial application on a GPU. In: Proceedings of the 2nd International Conference on Interaction Sciences: Information Technology, Culture and Human, pp. 1136-1141, New York, NY, USA, 2009. Association for Computing Machinery
10. Gomes, M., Lima, R., Sampaio, R.: Multiscale method: a powerful tool to reduce the computational cost of big data problems involving stick-slip oscillations. In: De Cursi, J.E.S. (ed.) Uncertainties 2020. LNME, pp. 69–79. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-53669-5_5

A Comparison of Different Approaches to Find the Probability Distribution of Further Generations in a Branching Process

João Pedro Freitas(✉), Roberta Lima, and Rubens Sampaio

Laboratório de Vibrações, Departamento de Engenharia Mecânica,
Pontifícia Universidade Católica do Rio de Janeiro, Rua Marquês de São Vicente,
255, Gávea, 22451-900 Rio de Janeiro, RJ, Brazil
joaopxf@aluno.puc-rio.br, {robertalima,rsampaio}@puc-rio.br

Abstract. In this paper, the spread of a general epidemic over time is modeled as a branching process. It is a stochastic process sorted as an individual-based model, which records population growth over generations with uncertainties to its size. The source of randomness is inherently related to the individual behavior of each member in a population. In this context, the transmissibility of the disease, i.e., the contagion from an infected person to susceptible ones is the root. Therefore, a discrete random variable models the number of infections per infector and rules the branching process. Given the probabilistic model of the contagion, the objective of the paper is to compare three methodologies to evaluate the mass functions of further generations of the branching process: probability generating functions (pgf), Markov chains (MC) and Monte Carlo simulations (MCS). The former gives analytical expressions, that can be symbolic computed, to evaluate the probability of an arbitrary number of infected members for a desired generation, whereas MC is a semi-numerical methodology and the latter is indeed a numerical one. The comparison between all of them relies on computational cost (runtime and storage) and limitation of applicability in relation to the mass function of the contagion. One of the characteristics of interest in the analysis is the determination of which methodologies allow the calculation of the mass function of a further generation without computing the mass functions of previous ones. This feature is referred in here as not time-dependent. Another characteristic of interest is the determination of which methodologies allow the computation of just some values of the mass function of a generation, i.e., probabilities related to the same generation can be achieved independently from the others. This is so-called a local property.

Keywords: epidemic · branching process · probability generating functions · Markov chains · Monte Carlo simulations

1 Introduction

The transmissibility of an epidemic is related to how easily a disease can spread from the contagion of an infected person (infector) to susceptible ones

J. E. S. De Cursi (Ed.): UQSM 2023, LNME, pp. 141–155, 2024.
https://doi.org/10.1007/978-3-031-47036-3_13

(infectees) [6]. It relies on hardly traceable features, such as the individual behavior and the pathogen's infectivity itself. Therefore, it is more suitable to model the transmissibility, including its inherent uncertainty.

The usual disease transmission models [11] are classified into three major groups: population-level, metapopulation and individual-level. The former splits the population according to the health state of its members. These subgroups are named compartments. The dynamic of the transitions between them are based on some averages. Metapopulation models slice the population into two groups, each one with specific parameters according to the disease and general characteristics. The latter is composed by networks and individual-based models (IBM) and it is getting more attention because it deals with stochastic outbreaks.

Such an IBM model is the branching process. It deals with the so-called demographic stochasticity [5], i.e., the population randomness over time is a consequence of the individual's uncertainty. This way, it entails the stochastically features of transmissiblity. The Bienaymé-Galton-Watson (BGW) is the classical version. It is a discrete state process in discrete time [2,3]. There are also the continuous ones in time [1,8] and the ones that the state space is continuous (CSBP), such as the Feller type.

In here, probabilistic descriptions of the size of subsequent generations of infected people are dealt. Therefore, it is meaningful to quantify the uncertainty over the process. Since some popular set of statistics, for example mean and standard deviation, mean and coefficient of variation or Shannon entropy, are not suitable as a proper measure for this task [7], the cumulative distribution function (CDF) still remains the best option. For this reason, the aim of this paper is to compare approaches to find the values of the mass function of further generations in a discrete state branching process that models an epidemic's spread.

This paper is organized as follows. Section 2 introduces the mathematical formulation of the branching process and the context of the epidemic's propagation over time to this model. Section 3 presents three different methodologies to find the values of the mass function to further generations in order to quantify the uncertainty: probability generating function, Markov chain and Monte Carlo simulation. In Sect. 4, there is a comparison among the approaches taking into account computational cost, applicability of random variables and other features later on discussed.

2 Epidemics Spreading Stochastically over Time

In a certain population, an epidemic takes off with a single infector, who is individual number 1 from the ramification tree in Fig. 1. This infected member belongs to the so-called 0th generation. The number of new members infected per infector is ruled by a discrete random variable in here named contagion C. As a consequence of the contact with individual number 1, only another person was infected (individual number 2). Therefore, the size of the 1st generation is unitary. Then, another realization related this time to individual number 2 of C is done. The outcome is that this infector spread the disease to individuals number 3 and 4, and hence the 2nd generation's size is two. For each one of these two infected people, a new realization of C is done. The evolution over time follows the same dynamic and in Fig. 1 is depicted this realization of the branching process up to the 5th generation.

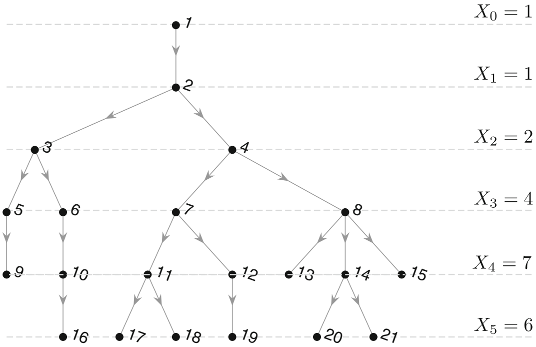


Fig. 1. Realization of the branching process up to the 5th generation.

The family of random variables of the branching process $\mathcal{X} = \{X_t\}, t \in \mathbb{N}^0$ is in this context the size of each generation of members infected. For definition, it starts with the deterministic statement $X_0 = 1$, which means a single initial infector. In this model, the contagion random variable for each member infected is independent and identically distributed (i.i.d.), which means that the transmissibility does not rely on the population size and the evolution over time. The size of a subsequent generation X_{t+1} is determined by the amount of infectors of the most previous one X_t. For each infector member of the latter, a realization of contagion is done and the total sum is the size. Generally speaking, we have according to Schinazi [10] that

$$X_{t+1} = \sum_{k=1}^{X_t} C, \, t \in \mathbb{N}^0. \tag{1}$$

The operation in Eq. (1) is a sum of a random quantity of random variables and in order to find the values of the mass functions of any generation beyond the first (further generations), three methodologies are presented next.

3 Methodologies to Find the Values of Mass Functions of Further Generations

3.1 Probability Generating Functions

The values of a mass function $\mathbb{P}(X = x)$ of a non-negative integers-valued random variable X can be rewritten as a sequence of probabilities that respects the normalization condition. One way to get this sequence is through its probability generating function (pgf), which is unique for each discrete random variable. The pgf of X is the function $G_X(s)$ defined by

$$G_X(s) := \mathbb{E}\left(s^X\right) = \sum_{x=0}^{\infty} s^x \, \mathbb{P}(X = x) = p_0 + p_1 s + p_2 s^2 + \dots \,. \tag{2}$$

A classical example is the pgf from a Binomial distribution. This probability distribution models in this context a society with strictly social distancing rules, in which an infector contacts with m infectees at most and the probability of infection is p for each of them. Its pgf is given as

$$G_X(s) = \sum_{x=0}^{\infty} \binom{m}{x} p^x (1-p)^{m-x} s^x = \left[(1-p) + p\, s\right]^m . \tag{3}$$

The pgf from any further generation X_{t+1} can be related to the pgf from the contagion random variable according to Grimmett and Welsh [4]

$$\begin{aligned} G_{X_{t+1}}(s) &= \sum_{x=0}^{\infty} s^x \, \mathbb{P}(X_{t+1} = x) = \mathbb{E}\left(s^{X_{t+1}}\right) \\ &= \sum_{i=0}^{\infty} \mathbb{E}\left(s^{X_{t+1}} \mid X_t = i\right) \mathbb{P}(X_t = i) \\ &= \sum_{i=0}^{\infty} \mathbb{E}\left(s^{\underbrace{C + C + \ldots + C}_{X_t \text{ times}}}\right) \mathbb{P}(X_t = i) \\ &= \sum_{i=0}^{\infty} \underbrace{G_C(s)}_{\text{argument}}{}^{i} \, \mathbb{P}(X_t = i), \text{ since i.i.d.} \\ &= G_{X_t}(G_C(s)), \text{ as result of the definition in Eq. (2).} \\ &= G_C(G_C(\ldots(G_C(s)))), \text{ recurrence happens } t \text{ times.} \end{aligned} \tag{4}$$

In order to find a specific probability $\mathbb{P}(X_{t+1} = k)$ up from the pgf $G_{X_{t+1}}(s)$, it must be k times differentiated, divided for the factorial of k and then evaluated in $s = 0$,

$$\mathbb{P}(X_{t+1} = k) = \frac{1}{k!} \left. \frac{d^{(k)}\left[G_C(G_C(\ldots(G_C(s))))\right]}{ds^{(k)}} \right|_{s=0} . \tag{5}$$

This analytical approach allows us to calculate the values of the mass function of any further distribution without knowing the previous ones. This feature is referred in here as not time-dependent. The probabilities related to the same generation are also achieved individually from the others and it is so-called a local methodology. For instance, we are looking to find the values of the mass function of the second generation when the contagion random variable is modeled as $C \sim \mathit{Binomial}(2, 0.7)$. The pgf from it is $\left[0.7\,(0.7\, s + 0.3)^2 + 0.3\right]^2$ according to Eqs. (3) and (4). In order to find the whole values of the support of the mass function from X_2, the operation in Eq. (5) must be done individually as many times as the possible outcomes of the realization of this random variable, which in this case is five:

$$\mathbb{P}\left(X_2 = 0\right) = \frac{1}{0!}\left.\left[0.7\left(0.7\,s + 0.3\right)^2 + 0.3\right]^2\right|_{s=0} = 0.132.$$

$$\mathbb{P}\left(X_2 = 1\right) = \frac{1}{1!}\left.\frac{d^{(1)}\left[0.7\left(0.7\,s + 0.3\right)^2 + 0.3\right]^2}{ds^{(1)}}\right|_{s=0} = 0.213.$$

$$\mathbb{P}\left(X_2 = 2\right) = \frac{1}{2!}\left.\frac{d^{(2)}\left[0.7\left(0.7\,s + 0.3\right)^2 + 0.3\right]^2}{ds^{(2)}}\right|_{s=0} = 0.336.$$

$$\mathbb{P}\left(X_2 = 3\right) = \frac{1}{3!}\left.\frac{d^{(3)}\left[0.7\left(0.7\,s + 0.3\right)^2 + 0.3\right]^2}{ds^{(3)}}\right|_{s=0} = 0.202.$$

$$\mathbb{P}\left(X_2 = 4\right) = \frac{1}{4!}\left.\frac{d^{(4)}\left[0.7\left(0.7\,s + 0.3\right)^2 + 0.3\right]^2}{ds^{(4)}}\right|_{s=0} = 0.118.$$

The mass function of X_2 is displayed next at Fig. 2.

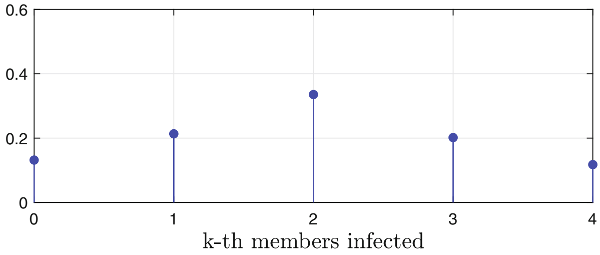


Fig. 2. Mass function of X_2 for $C \sim \mathit{Binomial}\,(2, 0.7)$.

3.2 Markov Chain Property

The branching process is Markovian, i.e., the sequence of random variables $\mathcal{X} = \{X_t\}_{t \in \mathbb{N}}$ defined in the discrete and finite state space $\mathbb{S}$ follows the rule: if we are looking to reach any state $i_{t+1} \in \mathbb{S}$ of some random variable X_{t+1}, it is only necessary to know the conditional probability of it based on the current random variable X_t reaching the state $i_t \in \mathbb{S}$, regardless of its past [4,9],

$$\begin{gathered}\mathbb{P}\left(X_{t+1} = i_{t+1} \mid X_0 = i_0, X_1 = i_1, \ldots, X_t = i_t\right) = \\ \mathbb{P}\left(X_{t+1} = i_{t+1} \mid X_t = i_t\right).\end{gathered} \tag{6}$$

It is fundamental in here that the state space $\mathbb{S}$ is finite. For this reason, the contagion random variable C of the branching process must also have a finite support. The conditional probabilities $p_{i,j}\left(t\right) = \mathbb{P}\left(X_{t+1} = j \mid X_t = i\right), t > 0$, as in Eq. (6) are named the t-th one-step tranisition probabilities. They are elements of the stochastic matrix known as the t-th one-step transition matrix $\mathbf{T}\left(t\right)$. The last follows the normalization condition for each row: $\sum_{j\in\mathbb{S}} p_{i,j} = 1$. It is responsible to relate the mass function of the distribution of X_{t+1} with the X_t one. It also enables to link the mass function of X_{t+1} with the X_1 one. In order to do this, the values of the initial distribution, i.e., X_1, along the whole state space $\mathbb{S}$ must be organized in a row vector λ_{X_1} (which is the same of the contagion, λ_C) and all the one-step transition matrices $\mathbf{T}\left(1\right), \mathbf{T}\left(2\right), \ldots, \mathbf{T}\left(t\right)$ must be known. The procedure is done in Eq. (7).

$$\lambda_{X_{t+1}} = \lambda_C \, \mathbf{T}\left(1\right) \mathbf{T}\left(2\right) \ldots \mathbf{T}\left(t\right). \tag{7}$$

The state space $\mathbb{S}$ of the branching process is defined according to the furthest generation of which the values of the mass function is desired to know. The dimension of all one-step transition matrices is then $|\mathbb{S}| \times |\mathbb{S}|$. But this is not a homogeneous Markov chain, which means the one-step transition matrices are not the same over the time. The reason of it is that each generation has a different upper limit of infectees, despite the fact that the whole branching process is defined on $\mathbb{S}$. This limit comes from the stochastic network structure of the branching process. The t-th generation for instance have ${q_C}^t$ infectees at most, in which q_C is the upper possible value from a realization of C. The t-th one-step transition probabilities $p_{i,j}\left(t\right)$ are then obtained according to

$$p_{i,j}\left(t\right) = \begin{cases} 1, \text{ if } \begin{cases} i = j = 0 \\ i > {q_C}^t, \, j = 0 \end{cases} \\ 0, \text{ if } \begin{cases} i = 0, \, j \neq 0 \\ i > {q_C}^t, \, j \neq 0 \\ 0 < i \leq {q_C}^t, \, j > i \times q_C \end{cases} \\ \mathbb{P}\left(\displaystyle\sum_{k=1}^{X_t = i} C = j\right), \text{ otherwise }. \end{cases} \tag{8}$$

The last statement above in Eq. (8) is calculated with the help of the pgfs. Unlike the Eq. (1), the size of the most previous generation, $X_t = i$, is already known. The pgf $G_{X_{t+1}}$ is related to G_C as in Grimmett and Welsh [4]

$$\begin{aligned} G_{X_{t+1}}(s) &= \sum_{x=0}^{\infty} s^x \, \mathbb{P}(X_{t+1}=x) = \mathbb{E}\left(s^{X_{t+1}}\right) \\ &= \mathbb{E}\left(s^{\overbrace{C+C+\ldots+C}^{}_{i \text{ times}}}\right) = \mathbb{E}\left(s^C s^C \ldots s^C\right) \\ &= \mathbb{E}\left(s^C\right)\mathbb{E}\left(s^C\right)\ldots\mathbb{E}\left(s^C\right), \text{ from independence} \\ &= G_C(s)\, G_C(s) \ldots G_C(s), \text{ by Eq. (2)} \\ &= {G_C}^i(s). \end{aligned} \tag{9}$$

Finally, the probability of the sum of a given number of random variables is

$$\mathbb{P}\left(\sum_{k=1}^{X_t=i} C = j\right) = \frac{1}{j!} \left. \frac{d^{(j)}\left[{G_C}^i(s)\right]}{ds^{(j)}} \right|_{s=0}. \tag{10}$$

This semi-numerical technique is a time-dependent and not a local one. In order to find the values of the mass function of X_{t+1}, the operation in Eq. (7) sequentially evaluates all the values of mass functions of the previous generation. The mass function is always found as a row vector of the whole state space $\mathbb{S}$ and is not possible to find the probabilities locally per state.

Taking the same example before. The state space of the Markov chain is $\mathbb{S} = \{0,1,2,3,4\}$. Therefore, the first one-step transition matrix has size 5×5 and the row vector of initial probability distribution has size 1×5. Notice that the actual support of the contagion random variable $C \sim \mathit{Binomial}\,(2, 0.7)$ is $[0,1,2]$. As a consequence, it is necessary to define probabilities for the remain states. Their probabilities $\mathbb{P}(C > 2)$ are zero. This way, the values of the mass function of X_2 are obtained from the relation in Eq. (7),

$$\lambda_{X_2} = \lambda_C \, \mathbf{T}(1),$$

where the row vector of the initial probability has size five

$$\lambda_C = \left[0.090\ 0.420\ 0.490\ 0\ 0\right]$$

and the first one-step transition matrix has its components calculated according to Eq. (8) and the last statement with the help of Eq. (10)

$$\mathbf{T}(1) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.090 & 0.420 & 0.490 & 0 & 0 \\ 0.008 & 0.076 & 0.265 & 0.412 & 0.240 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

The operation results in the whole row vector at once

$$\lambda_{X_2} = \left[0.132\ 0.213\ 0.336\ 0.202\ 0.118\right].$$

3.3 Monte Carlo Simulation

Some transformations of random objects are hard to make analytically. Instead of that, we make realizations of these objects, then the transformations are applied individually in each generated experiment. In this case, it means to make n_r realizations of the branching process. Each one of them is an experiment. They consist in one sample.

Sample statistics can be taken from the n_r realizations of the branching process. They are actually random variables since they depend intrinsically on the random numbers generator. In order to deal with this uncertainty, a convergence analysis must be made. Thus, an error ξ_1 must be fixed. For a certain quantity of experiments generated, if the sample statistics do not satisfy this tolerance, another sample with a greater number than n_r realizations must be generated and sample statistics reevaluated. For this case, the difference between the sample mean $\hat{\mu}_{X_{t+1}}$ and the expectation of the generation's random variable X_{t+1} is compared with the tolerance

$$\hat{\xi}_{t+1} = \left|\mathbb{E}\left(X_{t+1}\right) - \hat{\mu}_{X_{t+1}}\right| < \xi_1. \tag{11}$$

Despite the fact that we don't know previously the probability distribution of any further generation, its expectation can be found from Grimmett and Welsh [4] with the help of the pgfs and Abel's lemma

$$\begin{aligned} \frac{d^{(1)}G_{X_{t+1}}\left(s\right)}{ds^{(1)}} &= \frac{d^{(1)}}{ds^{(1)}} \sum_{x=0}^{\infty} s^x \, \mathbb{P}\left(X_{t+1} = x\right) \\ &= \sum_{x=0}^{\infty} \frac{d^{(1)}}{ds^{(1)}} s^x \, \mathbb{P}\left(X_{t+1} = x\right) \\ &= \sum_{x=0}^{\infty} x s^{\,x-1} \, \mathbb{P}\left(X_{t+1} = x\right). \end{aligned} \tag{12}$$

Taking $s = 1$ in Eq. (12), the expectation is found

$$\begin{aligned} \left.\frac{d^{(1)}G_{X_{t+1}}\left(s\right)}{ds^{(1)}}\right|_{s=1} &= \sum_{x=0}^{\infty} x \, \mathbb{P}\left(X_{t+1} = x\right) \\ &= \mathbb{E}\left(X_{t+1}\right). \end{aligned} \tag{13}$$

From Eq. (4), the pgf from the generation X_{t+1} is a multicomposition function of t recurrences of the pgf from the contagious random variable, so

$$\mathbb{E}\left(X_{t+1}\right) = \left.\frac{d^{(1)}\left[G_C\left(G_C\left(\ldots\left(G_C\left(s\right)\right)\right)\right)\right]}{ds^{(1)}}\right|_{s=1}. \tag{14}$$

This numerical approach is a time-dependent and not a local methodology. Once the number of experiments n_r is determined, each realization of the branching process gives values of infected members per generation. After that, a normalized histogram of each desired generation is done and the approximated mass function along the support is visualized.

4 Comparison Among the Methodologies

The comparison among the methodologies showed above shall not be done based on a single perspective. There are many variables presented in this task, e.g., computational cost (runtime and storage) and the law of the mass function chosen for the contagion random. Some other points are relevant, such as time-dependency and local properties. The machine that runs the MATLAB codes is a MacBook Air M2, 16 GB of RAM and 512 GB of storage.

In this work, the contagion random variable's law is $C \sim \mathit{Binomial}\,(m, p)$. The former parameter changes in the range $m = [1, 2, \ldots, 5]$, while the latter one in $p = [0.1, 0.3, \ldots, 0.9]$. This distribution is also chosen because the Markov's chain technique requires a discrete random variable with finite support. The furthest generation observed in here is X_6 due its already high values of runtime (CPU time) for some methodologies.

Figure 3 shows the runtime spent of the pgf approach to find the whole values of the mass function per generation. This methodology is at first hand not dependent on the parameter p. The main reason of it is the fact that the binomial pgf has p as a scalar and as a constant in its analytical expression. The only exception is the case of $p = 0.5$, in which the runtimes are faster. The explanation is that in this situation the expression in Eq. (3) does a factorization that decreases the runtime. If an expansion of the analytical function of the pgf is previously done, the runtime remains close to the other cases.

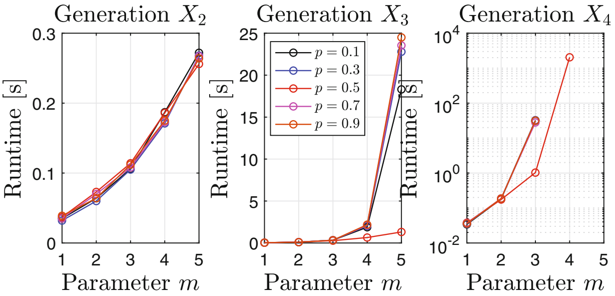


Fig. 3. Runtime per generation for the pgf approach.

On the other hand, the parameter m occurs as an exponent. Figure 3 shows indeed that the runtime has a significantly dependency on the value of m in every generation. As further as the generation is, the runtime of this approach increases the most. It is a consequence of the complexity of the multicomposition function from Eq. (5). The runtime spent to find the values of the mass function for the whole support up the fourth generation is not feasible anymore, except for $p = 0.5$. An interesting aspect of the runtime in this methodology is that it is a deterministic property despite of computational noises. Furthermore, this is the only local and time-independent technique. If the aim is to find specific values of the mass function of some generation, this might be an interesting approach.

The Markov chain technique has a similar effect of the parameter p as the pgf approach. In this case, there is not a remarkable influence coming from $p = 0.5$ as in the previous methodology. Figure 4 shows that the meaningful dependency relies on the parameter m, once the Eq. (10) for the t-th one step transition probabilities uses also the binomial pgf.

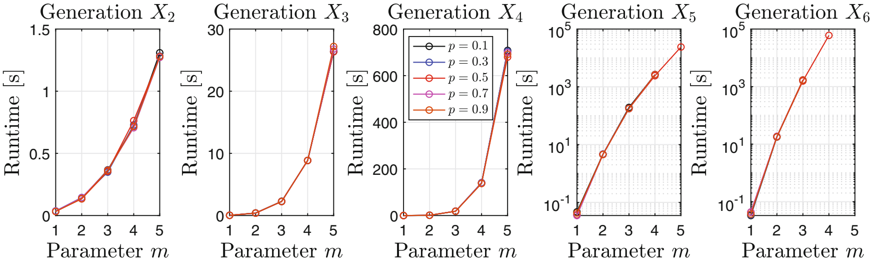


Fig. 4. Runtime per generation for the MC approach.

The main difference is that, this time, there are not multicompositions functions, which makes smaller runtimes than the pgf approach. For $m = 1$, there is indeed a homogeneous Markov chain and the runtime remains almost the same regardless the generation. The reason is that the state space in this situation is $\mathbb{S} = \{0, 1\}$ and the only possible time-evolution to this branching process is that each generation has a single infector. As a consequence, the one-step transition matrix does not change over time. It also has a significant small dimension of 2×2 and the operation in Eq. (7) is way more feasible. The remain cases show an amount growth of the runtime per generation. It has also a deterministic runtime.

For the Monte Carlo simulation, the convergence study was realized with the mass function of $C \sim \mathit{Binomial}\,(5, 0.9)$ and the third generation was the one to perform Eq. (11), because the number of experiments required to further ones increases a lot. The tolerances was $\xi_1 = 10^{-3}$. The required number of experiments were 28738. For this reason, all the Monte Carlo simulations next were done with $n_r = 30000$. Figures 5 and 6 show respectively the convergence study and the evolution of histograms of the mass function of X_3 per different values of n_r.

Different from the other techniques, this is a purely numerical one and does not rely on any pgf. Since it is dependent on the realizations of the branching process, and, as a consequence, on a series of realizations of the contagion random variable, greater values of p and m result in longer runtimes. Despite of the approximate values from the mass functions, this technique shows the greatest benefit in runtime for further generations, as shown in Fig. 7. On the other hand,

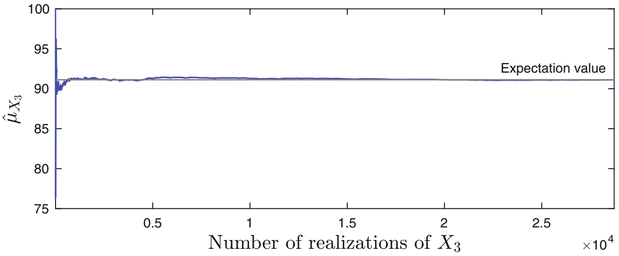


Fig. 5. Convergence study of the third generation for $C \sim \mathit{Binomial}\,(5, 0.9)$.

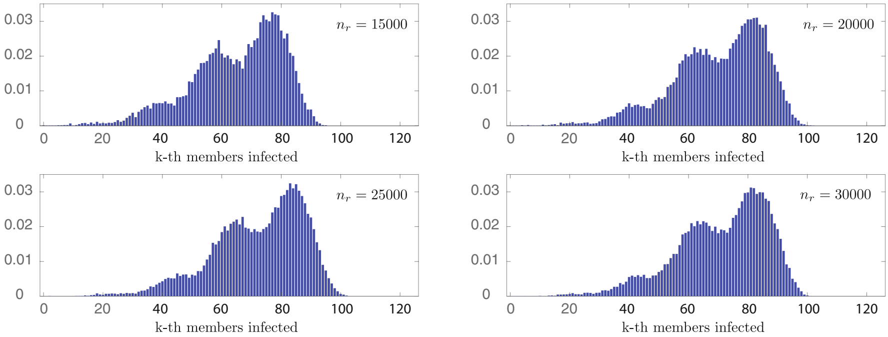


Fig. 6. Normalized histogram evolution of X_3 for $C \sim \mathit{Binomial}\,(5, 0.9)$ per n_r.

the convergence for further generations than the third one for the same criteria of tolerances requires a greater number of n_r than the one used. Furthermore, features of the probabilistic model, e.g., the mathematical relation between the contagion and any generation's size random variable is lost in this methodology.

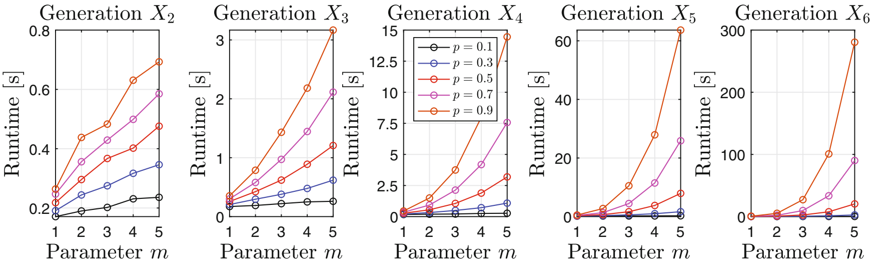


Fig. 7. Runtime per generation for the MCS approach.

The runtime in here is a stochastic object. It is dependent on the sample generator. Figure 8 shows the runtime spent for each experiment also from the third generation for $C \sim Binomial\,(5, 0.9)$. The total value of runtime of the simulation is therefore a sum of random experiments' runtime.

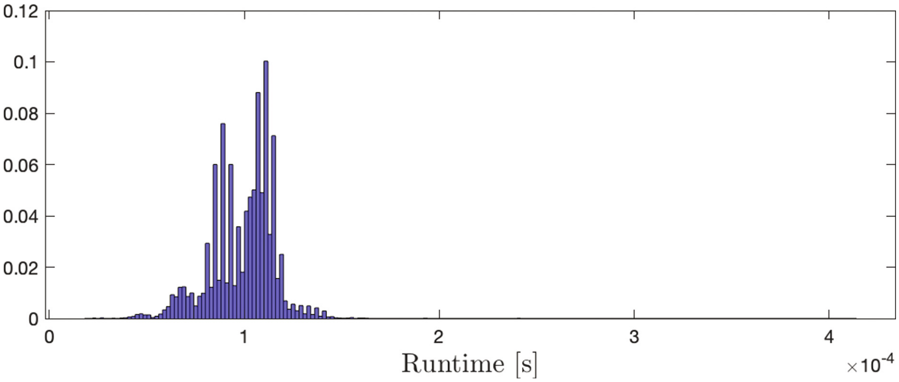


Fig. 8. Normalized histogram of runtime per experiment.

Next, it is presented in Figs. 9, 10 and 11 a storage analysis of the methodologies. The saved files from the pgf approach contain the symbolic expressions of the Eq. (5) for each quantity of infectors, the runtime spent to operate it and the values of the mass function. The ones from the MC technique carry the vector of the initial distribution, the required one-step transition matrices and the probabilities related to the support of the desired generation and the previous ones. Finally, for the MCS methodology, the files have the number of infected members, the runtime spent per generation of each experiment and the probabilities related to all infected members of the generations up to the desired one.

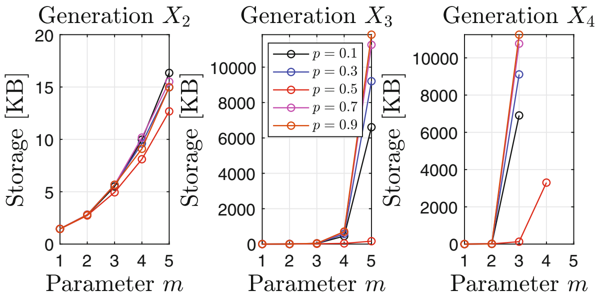


Fig. 9. Storage per generation for the pgf approach.

The greatest increase of storage size per generation happens to the pgf methodology. This is related to the data that includes an extensive use of symbolic computation. Since it is a local methodology, a symbolic expression is required for each value of the mass function. This time the parameter p shows a stronger influence than in the case of the runtime analysis. For further generations, the storage size increases as the parameter p turns greater, except for $p = 0.5$. In this case, there is still a major benefit. Another similar fact is that the storage increases exponentially according to the value of the parameter m, reflecting the complexity of the expression in Eq. 5 according to this value. This is also a deterministic feature.

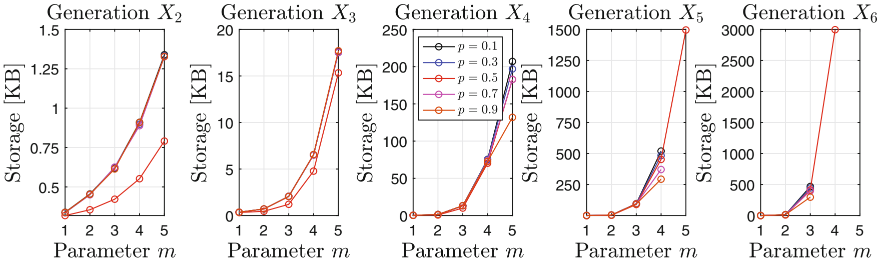


Fig. 10. Storage per generation for the MC approach.

For the MC technique, the similarity of the influence of the parameter m remains. On the other hand, the parameter p shows some stronger influence on the storage size, specially when $p = 0.5$. Up to third generation, this approach has the lowest values of storage per generation. Another relevant aspect depicted in Fig. 10 is that for the 5th and 6th generations, the higher values of p seem to decrease the storage size.

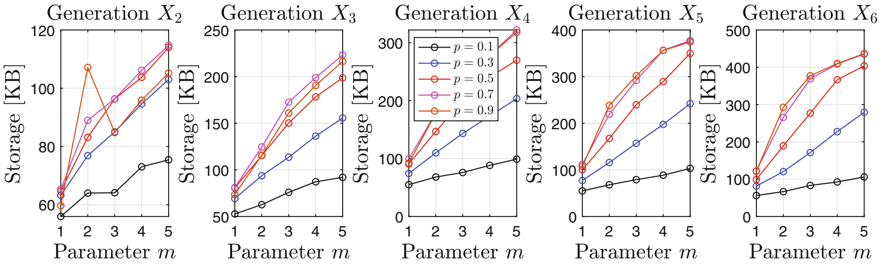


Fig. 11. Storage per generation for the MCS approach.

The storage size for the MCS approach is a stochastic feature. The data files collect generation per generation the results of a binomial random generator for several times. This is reflected specially in the second generation in Fig. 11 when any recognizable pattern is not found based on what it is usually expected for the combination of greater values of m and p. As generations goes by, for

$p = 0.1$ and $p = 0.3$, the storage size seems to increase in a linear way on m. On the other hand, for the values of $p = 0.5$, $p = 0.7$ and $p = 0.9$, it seems to increase asymptotically to a limit storage value. For the furthest generations, this methodology has the lowest values of storage size.

5 Conclusions

In this work, three different methodologies to obtain the values of the mass functions over generations are used and compared. The pgf methodology presents the advantage of having an analytical expression that links each value of the mass function of some generation to the random variable chosen to model the contagion. Despite this beneficial property, when the aim is to find the whole probabilities of the number of members infected for any generation, it was found that the methodology struggles in terms of computational costs. In the MC technique, there is a numerical connection between the values of the mass functions and the values of the contagion one. From the perspective of the computational costs, this connection can be used to find the probabilities related to the whole support of some generation in a more feasible way. Nevertheless, the application of the this methodology is limited, as it can only be applied to distributions where finite and discrete supports are considered. Finally, the MCS approach shows the best performance in terms of computational costs. However, the convergence of further generations reflects a high number of experiments, and a big data problem is a reality.

Acknowledgments. The authors acknowledge the support given by FAPERJ, CNPq and CAPES.

References

1. Bansaye, V., Méléard, S.: Stochastic Models for Structured Populations: Scaling Limits and Long Time Behavior, 1st edn. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-21711-6
2. Borges, B., Lima, R., Sampaio, R.: Análise estocástica de propagação de doenças epidemiológicas. Revista Mundi Engenharia, Tecnologia e Gestão **6**(3), 352–01, 352–11 (2021). 10.21575/25254782rmetg2021vol6n31636
3. Borges, B., Lima, R., Sampaio, R.: How the spread of an infectious disease is affected by the contagion's probabilistic model. In: XIV Encontro Acadêmico de Modelagem Computacional, pp. 1–10 (2021)
4. Grimmett, G., Welsh, D.: Probability: An Introduction, 2nd edn. Oxford University Press, Oxford (2014)
5. Haccou, P., Jagers, P., Vatutin, V.A.: Branching Processes: Variation, Growth, and Extinction of Populations, 1st edn. Cambridge University Press, Cambridge (2005)
6. Leung, N.H.: L: Transmissibility and transmission of respiratory viruses. Nat. Rev. Microbiol. **19**, 528–545 (2021). https://doi.org/10.1038/s41579-021-00535-6
7. Lima, R., Sampaio, R.: What is uncertainty quantification? J. Brazil. Soc. Mech. Sci. Engi. **40**(155) (2018). https://doi.org/10.1007/s40430-018-1079-7

8. Pardoux, E.: Probabilistic Models of Population Evolution, 1st edn. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-30328-4
9. Privault, N.: Understanding Markov Chains, 1st edn. Springer, Cham (2013). https://doi.org/10.1007/978-981-13-0659-4, www.springer.com/series/3423
10. Schinazi, R.B.: Classical and Spatial Stochastic Processes, 2nd edn. Springer, Cham (2014). https://doi.org/10.1007/978-1-4939-1869-0
11. Verelst, F., Willem, L., Beutels, P.: Behavioural change models for infectious disease transmission: a systematic review (2010-2015). J. R. Soc. Interface **13** (2016). https://doi.org/10.1098/rsif.2016.0820

Data Assimilation Using Co-processors for Ocean Circulation

Marcelo Paiva[1], Sabrina B. M. Sambatti[2], Luiz A. Vieira Dias[3], and Haroldo F. de Campos Velho[1](✉)

[1] National Institute for Space Research (INPE), São José dos Campos, SP, Brazil
haroldo.camposvelho@inpe.br

[2] Independent researcher, São José dos Campos, SP, Brazil

[3] Aeronautics Institute of Technology (ITA), São José dos Campos, SP, Brazil

Abstract. Data Assimilation is a procedure for fusion from the observational system and previous forecasting to calculate the initial condition – also called *analysis* – for the next prediction cycle. Several methods have been developed and applied for data assimilation (DA). We can cite the Kalman filter, particle filter, and variational approach as methods employed for DA. However, the mentioned methods are computer intensive. One alternative to reduce the computational effort for DA is to apply a neural network (NN) to emulate a computationally expensive technique. The NN approach has been also applied for ensemble prediction to address uncertainty quantification for each forecasting cycle. A self-configuring framework was applied to design the best NN architecture to emulate the Kalman filter for DA, using the metaheuristic: Multiple Particles Collision Algorithm (MPCA). The optimal artificial neural network is implemented on two types of co-processors: FPGA (Field-Programmable Gate Array), and TPU (tensor processing unit). Shallow water 2D system is designed to simulate ocean circulation dynamics, where the finite difference scheme is used for numerical integration of the model. The artificial neural network was effective, with reduction of processing time. The use of FPGA or TPU as co-processors for data assimilation have similar precision in comparison with analysis calculated by software. The better processing time performance among multi-core CPU, FPGA, and TPU was obtained by the TPU when the number of grid points ($N \times N$) is greater than 150. For $N \leq 150$, the CPU presented a smaller execution time.

Keywords: Data assimilation · ocean circulation · artificial neural network · co-processors: FPGA and TPU

1 Introduction

Modern forecasting systems are based on time integration of differential equations by numerical methods [10,14]. For weather and climate operational prediction centers, supercomputers with high processing power are used for this task. This is a research field under permanent development [4].

J. E. S. De Cursi (Ed.): UQSM 2023, LNME, pp. 156–164, 2024.
https://doi.org/10.1007/978-3-031-47036-3_14

In order to produce a better prediction as possible, one important issue is to compute the initial condition (IC). The calculation of the IC is called *data assimilation* (DA). DA is a procedure for combining data from an observation system and data from the mathematical model [11]. The DA procedure is named *analysis*. The computation of the analysis demands a huge computational effort, due to the dimension of the observational data and the number of grid points in computer models of the atmosphere and/or ocean dynamics. Therefore, there are many investigations for new methods for DA, and the innovative hardware architectures.

To save processing time, new algorithms based on artificial neural networks [7] have been proposed. The goal of the present paper is to report results using co-processors for CPU multi-core: FPGA (field programmable gate array), and TPU (tensor processing unit).

Cintra and co-authors applied neural networks for data assimilation for a 3D atmospheric global circulation model [8]. The neural network was designed to have a similar analysis as that obtained by the local ensemble transform Kalman filter. The uncertainty evaluation is computed by using an *ensemble prediction* approach [11]. Neural networks were trained using the ensemble average only. The trained neural network was applied for each ensemble member producing an ensemble prediction. Fifteen ensemble members were employed in the numerical experiment, showing a practical way to use neural network as a data assimilation with ensemble framework. From this consideration, the forecasting uncertainty quantification can be calculated by using the same strategy as Cintra et al. [8].

A shallow water 2D (SW-2D) system is used as a simplified model to simulate ocean circulation [5]. The finite difference method is employed to discretize the SW-2D system, and the time integration is carried out by a forward-backward scheme [13].

Here, a neural network is the method for data assimilation. The optimal multi-layer perceptron neural network (MLP-NN) is configured to emulate the Kalman filter [6]. The self-configuring MLP-NN is performed by solving an optimization problem with the meta-heuristic multi-particle collision algorithm (MPCA) [12]. The MLP-NN for the data assimilation process was codified to the FPGA and TPU. The DA processing itself was faster with the FPGA, but considering the time for data transfer between the CPU and the co-processor plus the DA processing the TPU was more effective. Indeed, the TPU processing is superior to the CPU when the number of grid space points ($N \times N$) is greater than 150 ($N > 150$).

2 Shallow Water Model for Ocean Dynamics

The shallow-water system is used to represent a 2D ocean circulation. The ocean depth is rigid surface defined by: $z = h(x, y)$, and the ocean interface with the atmosphere is a free surface defined as: $z \equiv q(x, y, t) + h(x, y)$. Two dimension equations are obtained by vertical integration of the Euler equation on vertical [1]. The average depth denoted by H is coupled to the velocity vector, with components (u, v). A limited area is the domain for ocean circulation [5], where the differential equations for time $t > 0$ over a 2D region $D(x, y)$ are given by:

$$\frac{\partial u}{\partial t} - fv + g\frac{\partial q}{\partial x} + r_u u = F_u \tag{1}$$

$$\frac{\partial v}{\partial t} + fu + g\frac{\partial q}{\partial y} + r_v v = F_v \tag{2}$$

$$\frac{\partial q}{\partial t} + H\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) + r_q q = 0 \tag{3}$$

where $D(x, y) = (0, L_x) \times (0, L_y)$, g is the gravity acceleration, f is the Coriolis parameter, parameters $(r_u,\ r_v,\ r_q)$ are the damping coefficients [5], and (u, v) is the ocean velocity components. The heterogeneous terms F_u and F_v are written as:

$$F_u = -C_d\, \rho_a\, u_a^2/(H\, \rho_w)\ , \tag{4}$$

$$F_v = 0\ . \tag{5}$$

being C_d the drag coefficient is the parameter C_d, ρ_a and ρ_w, are the air and ocean water densities, respectively, and zonal wind u_a has a constant value. Boundary conditions $\{\partial D(x, y)\}$ are the same used by Bennet [5] – see also Campos Velho et al. [6].

Table 1 shows the numerical values for numerical discretizations, and the parameters assumed to represent the ocean dynamics – see Bennett [5].

Table 1. Parameters used in the integration for the SW-model.

Parameter	Value	Parameter	Value
Δt (s)	180	r_u (s^{-1})	1.8×10^4
N_t	200	r_v (s^{-1})	1.8×10^4
t_{max} (h)	3.6×10^4	r_q (s^{-1})	1.8×10^4
Δx (m)	10^5	ρ_q (kg m^{-3})	1.275
Δy (m)	10^5	ρ_w (kg m^{-3})	1.0×10^3
N_x	40	g (m s^{-2})	9.806
N_y	40	f (s^{-1})	1.0×10^{-4}
H (m)	5000	C_d	1.6×10^{-3}
		u_a (m s^{-1})	5

3 Data Assimilation

The best combination between a previous prediction (*background*) and from data from observation system is called data assimilation (DA), producing the *analysis*. The analysis is the initial condition for next prediction cycle.

The configuration for supervised multilayer perceptron neural network (MLP-NN) [9] to emulate the Kalman filter is found by MPCA meta-heuristic. The MLP-NN is applied over each grid point, where the inputs are the observation and the background, and the output is the analysis for that grid point.

The cost function to be minimized by the MPCA is given by:

$$J(X) = penalty \times \left[\frac{\rho_1\, E_{\mathrm{Learn}} + \rho_2\, E_{\mathrm{Gen}}}{\rho_1 + \rho_2} \right] \; ; \tag{6}$$

$$penalty = c_1\, e^{\{\#\mathrm{neurons}\}^2} + c_2\, \{\#\mathrm{epochs}\} + 1 \; . \tag{7}$$

where the square difference between the analysis obtained from the Kalman filter and the MLP-NN output is the *training* (E_{Learn}) and *generalization* (E_{Gen}) errors, respectively. The cost function also looks for the simplest NN, with a balance between errors with the smallest number of neurons and faster convergence for the training (learning) phase – see Eq. (7). Other parameters in Eq. (6) are free ones. For the present problem, it is assumed the values $(\rho_1,\, \rho_2,\, c_1,\, c_2) = (0.5,\, 0.5,\, 5 \times 10^8,\, 5 \times 10^5)$ [2].

As a first experiment, the data assimilation is executed at each 10 time-steps, with the grid points expressed in Table 1 ($N_x \times N_y = 40 \times 40$). The result for the experiment with MLP-NN emulating a Kalman filter for data assimilation is shown in Fig. 1. The experimental values are collected at 25 points for the ocean

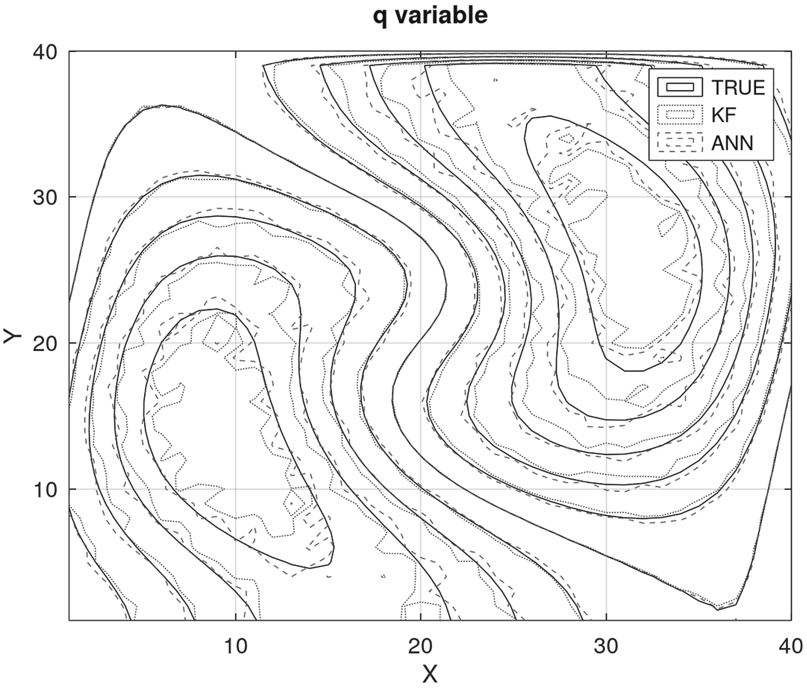


Fig. 1. Data assimilation for ocean circulation at time-step $t = 30$ by using Kalman filter and MLP neural network.

level $q(x, y)$. After 3 cycles of DA (time-step $t = 30$), isocurves for the q-variable are displayed in Fig. 1: true dynamics ("TRUE"), Kalman filter ("KF"), and neural network ("ANN"). The difference between the analysis and the reference values (TRUE lines) is smaller when the data assimilation is executed with the neural network than Kalman filter.

4 Co-processors: FPGA and TPU

4.1 FPGA

A field-programmable gate array (FPGA) is an integrated circuit containing an array of programmable logic blocks, with reconfigurable interconnects. Logic blocks can be combined in different ways to calculate complex functions. Generally, FPGA can be configured by using a hardware description language (HDL), in a similar way as in the application-specific integrated circuit (ASIC).

Cray XD1 is a heterogeneous computer system integrating CPU multi-core and reconfigurable computing technologies. Cray XD1 is based on Direct Connected Processor (DCP) architecture. The machine has six interconnected nodes (blades). Each blade contains two CPUs (2.4 GHz AMD Opteron) and one FPGA (Xilinx Virtex II Pro). Figure 2 shows the architecture of a Cray XD1 node (blade).

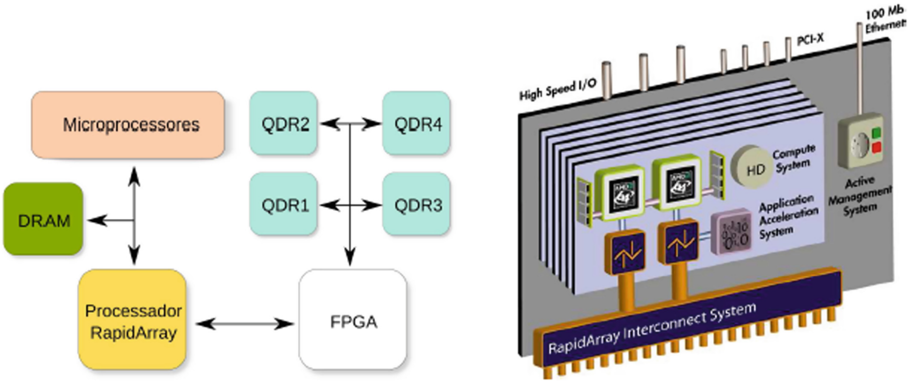


Fig. 2. Sketch for the Cray XD1 blade.

4.2 TPU

Different from scalar processors, operating on single data only, a central processing unit (CPU) designed for a type of parallel computation with instructions to operate on one-dimensional arrays of data, this type of parallel CPU is called vector processor.

The tensor processing unit (TPU) is also an Application-Specific Integrated Circuit (ASIC). It is designed to operate two-dimensional arrays optimizing matrix multiplication and accumulation operations. Google company developed a specific hardware for the Google's Tensorflow project – deep learning AI framework. The first TPU version (TPUv1) appeared in 2016. The second TPU version (TPUv2) was announced in 2017. The TPUv2 card is also compatible with computer available in the market. Years 2018 and 2021, Google presented the third and fourth versions (TPUv3 and TPUv4, respectively) of the TPU chips. The hardware of the respective TPU versions are illustrated in Fig. 3.

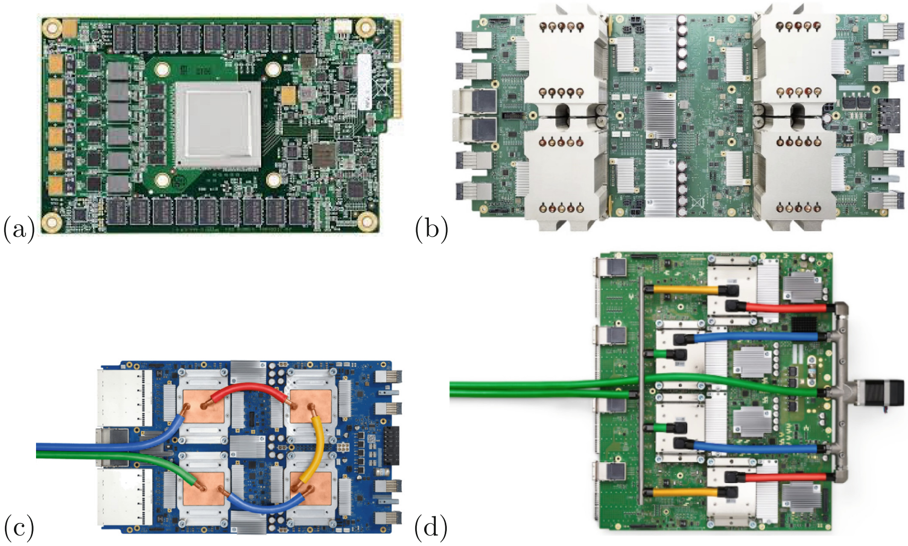


Fig. 3. TPU versions: (a) TPUv1, (b) TPUv2, (c) TPUv33 (d), TPUv4.

5 Results

5.1 FPGA Results

The numerical experiment was carried out with space domain discretization ($N_x = N_y = N$) as shown in Table 1 ($N = 40$), with DA cycle activated at each 10 time-steps. Software processing (CPU) was executed with 121709 µs, while the hardware execution spent 209187 µs.

The CPU execution was much faster than FPGA. Actually, there are more advanced FPGA than that implemented in the Cray XD1 blade nowadays. Looking at closer the experiment execution time, there are two types of time to be considered: data transfer, and effective processing time. Table 2 shows the time for the transfer of data between CPU-FPGA and FPGA-CPU, and the effective processing by the FPGA. Table 2 shows the time for transfer data between CPU-FPGA and FPGA-CPU, and the effective processing by the FPGA. There is a difference of time for transferring data from CPU-FPGA (greater) than FPGA-CPU (lower). This difference is because for FPGA-CPU transfer the data are storaged on the DRAM – see Fig. 2 –, a memory accessed by CPU and FPGA. However, the FPGA time processing for effective execution is very small (2μs).

Table 2. FPGA: time for effection processing × time for data transfer.

Process	time
CPU-FPGA transfer	181635 μs
FPGA-CPU transfer	9455 μs
FPGA processing	2 μs

5.2 TPU Results

The numerical experiment with TPU was carried out on the Google Colaboratory platform, also called as "Colab". Colab is a cloud service, hosted by the Google company. The main language for the Colab is Python, but the platform also works with other computer languages: Fortran, R, Julia, Swift, among other. The Google Colab has three options for processing: CPU, GPU, and TPU. The experiments used an CPU Intel® Xeon® 32-bit 2.20 GHz, with only one TensorCore type TPUv2, containing eight TensorCores.

The TPUv2 was executed with several number of grid points. Depending on the number of grid points, the CPU presented a better performance than TPU. Therefore, other executions were necessary to evaluate the performance with enhancing the number of grid points. Table 3 shows execution times (seconds) for CPU and TPU for several number of grid points. For this application with the described hardware, if the number of space discretization points ($N \times N$) is lower than 150 the CPU has a better performance than TPUv2.

Table 3. CPU vs TPUv2.

Matrix: $N \times N$	CPU time	TPU time
100	0,2110564709	0,3778991699
150	0,3215475082	0,3667955399
200	0,4424006939	0,3672273159
250	0,6224746704	0,3842322826

6 Final Remarks

Data assimilation for a simplified ocean circulation model was carried out by self-configuring multi-layer perceptron neural network. The optimal MLP-NN was designed to emulated the Kalman filter (KF). Campos Velho et al. [6] did an algorithm complexity analysis showing that NN has a lower complexity than KF, with computer experiments showing a significant reduction of the processing time.

In a different fashion, the focus of the present paper is to apply co-processors for speeding up the DA processing by the MLP-NN. Results with two co-processors were presented: FPGA and TPU.

The MLP-NN was implemented on FPGA (cray XD1 heterogeneous computer) and TPUv2 (CoLab Google cloud). The results with FPGA had a slower performance than CPU. From Table 2, it is clear that the time for transfer data between CPU and co-processor was the main factor for the difference of the processing time. There is new hardware embracing CPU and FPGA in the same chip. One example is the Zynq-7000 family, integrating an ARM Cortex-A9 CPU with FPGA on SoS (system-on-chip). But, this hardware is not standard for HPC servers. The TPUv2 performance was superior than the CPU with number of grid points greater than 150. However, the results with TPU are still preliminaries.

Ensemble prediction [11] is the standard procedure to address the predictability, the forecasting uncertainty quantification, by the operational weather and ocean circulation prediction centers. The predictability can be computed as the strategy developed by Cintra et al. [8] for ensemble prediction.

Acknowledgments. Authors wish to thank the National Institute for Space Research (INPE), the Aeronautics Institute of Technology (ITA): ITA Computer Engineering Undergraduate Program, ITA Electronic and Computer Engineering Graduate Program (ITA-PG/EEC-I), the Google company: for making available, free of charge, the Google Colaboratory development platform (Colab), Google hardware Tensor Processing Unit (TPU) version-2; Dr. Roberto Pinto Souto (researcher from the LNCC: National Laboratory for Scientific Computing, Brazil) for providing the MPCA source code, as well as the SW2D model. Authors also thanks to the National Council for Research and Development (CNPq) – Brazilian agency for research support – for the financial project support (Grant: 422614/2021-1, and Grant: 314660/2020-8 (author: HFCV)).

References

1. Altaie, H.O.: New techniques of derivations for shallow water equations. Int. J. Adv. Sci. Tech. Res. **3**(6), 131–151 (2016)
2. Anochi, J.A., Campos Velho, H.F., Hernandez Torres, R.: Two geoscience applications by optimal neural network architecture. Pure Appl. Geophys. **176**(1), 1–21 (2019)
3. Anochi, J.A., Hernández Torres, R., Campos Velho, H.F.: Climate precipitation prediction with uncertainty quantification by self-configuring neural network. In: De Cursi, J.E.S. (ed.) Uncertainties 2020. LNME, pp. 242–253. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-53669-5_18
4. Bauer, P., Thorpe, A., Brunet, G.: The quiet revolution of numerical weather prediction. Nature **4**(525), 47–55 (2015)
5. Bennett, A.F.: Inverse Modeling of the Ocean and Atmosphere. Cambridge University Press, Cambridge (2002)
6. Campos Velho, H.F., et al.: Data assimilation by neural network for ocean circulation: parallel implementation. Supercomput. Front. Innov. **9**(1), 131–151 (2022)

7. Cintra, R.S.C., Campos Velho, H.F.: Data Assimilation by Artificial Neural Networks for an Atmospheric General Circulation Model. Intech – Chapter 14, pp. 265–285 (Ed. by, El-Shahat). Advanced Applications for Artificial Neural Network (2018)
8. Cintra, R.S.C., Cocke, S., Campos Velho, H.F.: Data assimilation by neural networks with ensemble prediction. In: International Symposium on Uncertainty Modeling and Analysis (joint ICVRAM ISUMA), Florianópolis (SC), Brazil, pp. 35–43 (2018)
9. Haykin, S.: Neural Networks: A Comprehensive Foundation, 2nd edn. Prentice Hall, Upper Saddle River (1994)
10. Haltiner, G.J., Williams, R.T.: Numerical Prediction and Dynamic Meteorology. Wiley, New York (1980)
11. Kalnay, E.: Atmospheric Modeling, Data Assimilation and Predictability. Cambridge University Press, Cambridge (2002)
12. Luz, E.F.P., Becceneri, J.C., Campos Velho, H.F.: A new multi-particle collision algorithm for optimization in a high performance environment. J. Comput. Interdiscip. Sci. **1**(1), 3–10 (2008)
13. Mesinger, M., Arakawa, A.: Numerical Methods Used in Atmospheric Models, World Meteorological Organization (WMO), Global Atmospheric Research Program (1976)
14. Washington, W.M., Parkinson, C.L.: An Introduction to Three-Dimensional Climate Modeling. University Science Books, Melville (1986)

Uncertainty Analysis of a Composite Plate Using Anti-optimization and PCE

Everton Grotti(✉), José G. P. Filho, Pedro B. Santana, and Herbert M. Gomes

Mechanical Engineering Department, Federal University of Rio Grande do Sul, UFRGS, Avenue Sarmento Leite, 425, sala 202, 2o. Andar., RS 90050-170 Porto Alegre, Brazil
{ewerton.grotti,jose.picoral,pedro.santana}@ufrgs.br, herbert@mecanica.ufrgs.br

Abstract. Uncertainty propagation has gained increasing attention in the research community in recent years. A better understanding of the uncertainty translates into a more efficient final product. Composite materials are susceptible to the aforementioned uncertainties, for instance by means of variations in material properties, loadings and manufacturing process. In this study, a composite plate uncertainty propagation problem is addressed with three techniques: Anti-optimization Interval Analysis, Polynomial Chaos Expansion (PCE), and the traditional Monte Carlo method. The dynamic mechanical response of the composite plate is analysed in the time domain. The anti-optimization interval analysis approach resulted in wider envelopes in the time histories (lower and upper bounds) when compared to PCE and Monte Carlo, especially in the last and more challenging example. Despite being unable to generate envelopes as broad as the other two approaches, PCE showed to be very attractive due to the small number of function evaluations used, especially in simpler problems. The adopted PCE algorithm is based in a non-intrusive approach: The Multivariate Collocation Method.

Keywords: Composite plate · PCE · anti-optimization

1 Introduction

Composite materials are widely employed in engineering applications due to their attractive mechanical qualities, such as high strength-to-weight ratio, stiffness, and fatigue resistance. However, uncertainties in load, fiber angle, and material properties may heavily impact the structural behavior. Therefore, the uncertainty propagation analysis is essential to predict the structural response and design a safe and efficient final product.

According to [1], Monte Carlo in association with finite element method is frequently used for uncertainty analysis in composite laminated structures. This approach is performed in [2], for example. Despite being the classical approach to uncertainty problems, Monte Carlo has a slow convergence rate and high computational cost.

J. E. S. De Cursi (Ed.): UQSM 2023, LNME, pp. 165–177, 2024.
https://doi.org/10.1007/978-3-031-47036-3_15

Recently, [3] used the anti-optimization and Monte Carlo approaches with convex hull to evaluate the uncertainty propagation in composite laminates. The comparison showed that anti-optimization approach is very attractive in the composite uncertainty propagation problem. Consequently, many authors have been using this approach such as [4–7].

The Polynomial Chaos Expansion (PCE) is another interesting technique, often used to perform recent uncertainty analysis. The great attraction regarding PCE is definitely the low number of numerical evaluations needed to solve problems (often dozens of function calls, compared to tens of thousands usually required by other approaches). The number of evaluations depends on the number of uncertain variables and the complexity of the problem, which may require polynomial with higher levels of expansion. Recent examples of PCE in composite materials are [8–10]. Extensive literature can be found in the subject, such as [11–15].

This study uses Polynomial Chaos Expansion (PCE), Monte Carlo, and Anti-Optimization techniques to access the uncertainty propagation in a composite plate. The structure is simulated in the time domain using the Newmark algorithm and the finite element method. Envelopes are generated by each approach and compared with each other in different problems: uncertainty in fiber orientation angle, excitation load, damping ratio, and finally, an example regarding a higher number of composite layers.

2 Methods

The main methods and assumptions used in this study will be briefly presented in this section, such as the finite element simulation of a rectangular plate in time domain, and the uncertainty propagation techniques: Monte Carlo sampling; Anti-optimization; and generalized polynomial chaos.

2.1 Finite Element Simulation of a Plate

In order to simulate the composite plate, a finite element analysis algorithm based on Mindlin plate theory is used. According to [16] the displacement field in this case is the same as in first order deformation theory (FSDT):

$$u(x, y, z) = u_0(x, y) + z\theta_x(x, y), \tag{1}$$

$$v(x, y, z) = v_0(x, y) + z\theta_y(x, y), \tag{2}$$

$$w(x, y, z) = w_0(x, y), \tag{3}$$

where u, v, and w are the displacements, θ_x is the rotation on the x axis, and θ_y is the rotation on the y axis.

The strain tensor can be found by deriving the displacements u, v, and w as follows

$$\begin{Bmatrix} \varepsilon_{xx} \\ \varepsilon_{yy} \\ \gamma_{xy} \\ \gamma_{xz} \\ \gamma_{yz} \end{Bmatrix} = \left\{ \frac{\partial u}{\partial x} \frac{\partial v}{\partial y} \left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} \right) \left(\frac{\partial u}{\partial z} + \frac{\partial w}{\partial x} \right) \left(\frac{\partial v}{\partial z} + \frac{\partial w}{\partial y} \right) \right\}^T \tag{4}$$

To define the stiffness matrix, one should start from the strain energy and integrate throughout each layer over the thickness direction. The membrane part of the stiffness matrix $\boldsymbol{K}_{mb}^{(e)}$, the membrane-bending coupling components $\boldsymbol{K}_{mb}^{(e)}$ and $\boldsymbol{K}_{bm}^{(e)}$, the bending $\boldsymbol{K}_{bb}^{(e)}$ and the shear component $\boldsymbol{K}_{ss}^{(e)}$ are obtained as follows

$$\boldsymbol{K}_{mm}^{(e)} = \sum_{k=1}^{nc} \int_A \boldsymbol{B}_m^T \boldsymbol{D_k} \boldsymbol{B}_m (Z_{k+1} - Z_k) dA \tag{5}$$

$$\boldsymbol{K}_{mb}^{(e)} = \sum_{k=1}^{nc} \int_A \boldsymbol{B}_m^T \boldsymbol{D_k} \boldsymbol{B}_b \frac{1}{2} (Z_{k+1}^2 - Z_k^2) dA \tag{6}$$

$$\boldsymbol{K}_{bm}^{(e)} = \sum_{k=1}^{nc} \int_A \boldsymbol{B_b}^T \boldsymbol{D_k} \boldsymbol{B}_m \frac{1}{2} (Z_{k+1}^2 - Z_k^2) dA \tag{7}$$

$$\boldsymbol{K}_{bb}^{(e)} = \sum_{k=1}^{nc} \int_A \boldsymbol{B_b}^T \boldsymbol{D_k} \boldsymbol{B_b} \frac{1}{3} (Z_{k+1}^3 - Z_k^3) dA \tag{8}$$

$$\boldsymbol{K}_{ss}^{(e)} = \sum_{k=1}^{nc} \int_A \boldsymbol{B_s}^T \boldsymbol{D_k} \boldsymbol{B_s} (Z_{k+1} - Z_k) dA \tag{9}$$

where nc is the number of layers, $\boldsymbol{D}$ is the rotated constitutive matrix, $\boldsymbol{B_m}$ and $\boldsymbol{B_b}$ are the shape function matrix for membrane and bending, respectively, and z_k is the coordinate for the k^{th} layer in the thickness direction with the reference being the middle plane of the composite plate (starting at $k = 1$ in the negative coordinate).

Thus, the element matrix results in

$$\boldsymbol{K}^{(e)} = \boldsymbol{K}_{mm}^{(e)} + \boldsymbol{K}_{mb}^{(e)} + \boldsymbol{K}_{bm}^{(e)} + \boldsymbol{K}_{bb}^{(e)} + \boldsymbol{K}_{ss}^{(e)}. \tag{10}$$

The global stiffness matrix $\boldsymbol{K}$ is then assembled by superposition. The corresponding global mass matrix $\boldsymbol{M}$ is obtained as usual in a similar way.

For the free vibration, [16] states that the equation of motion in Mindlin plates, apart from the damping, can be expressed as

$$\boldsymbol{M\ddot{u}} + \boldsymbol{Ku} = \boldsymbol{f}, \tag{11}$$

where $\boldsymbol{f}$ is the force vector, and $\boldsymbol{\ddot{u}}$ and $\boldsymbol{u}$ are the accelerations and displacements vectors. The natural frequencies and vibration mode shapes can be found by solving an eigenproblem, described as

$$(\boldsymbol{K} - \Omega^2 \boldsymbol{M})\Phi = \boldsymbol{0}, \tag{12}$$

with $\boldsymbol{\Omega}$ being a diagonal matrix containing the square of circular frequencies (ω_i^2), and $\boldsymbol{\Phi}$ is the eigen matrix with the columns being the corresponding vibration modal shape vectors ($\boldsymbol{\phi}_i$).

Table 1. Typical mechanical properties of epoxy/carbon fiber unidirectional lamina. [17].

Property	Symbol	Unit	Graphite/epoxy
Longitudinal elastic modulus	E_1	GPa	181
Transversal elastic modulus	E_2	GPa	10.30
Major Poisson's ratio	ν_{12}	-	0.28
Shear modulus	G_{12}	GPa	7.71

A cantilever composite plate was simulated in this study, presenting 0.15 m x 0.05 m, with 144 elements (mesh can be checked at Fig. 2). It was assumed epoxy/carbon composite which mechanical properties are listed in Table 1 [17], and the simulation was solved in time domain using Newmark scheme. All software was implemented in Matlab R2012a.

A shear correction factor $\kappa = 5/6$ is used, as shown in [16]. Other elastic properties follow the usual assumption for plane stress.

2.2 Monte Carlo Sampling Approach

Monte Carlo (MC) sampling is the classic approach for solving uncertainty propagation problems. Although being reliable, MC has a slow convergence rate, thus being unapplicable in models that require hours to days for a single evaluation [15]. Convergence rates can be increased by using Latin hypercube and quasi-Monte Carlo sampling, but the method remains exceedingly computationally expensive.

An advantage of MC sampling is that the number of uncertain input variables does not increase the convergence time. This is also true for Anti-optimization Interval Analysis approach, but, on the other hand, not true for Polynomial Chaos Expansion (PCE). Therefore, in problems with a high number of uncertain input variables the PCE method may become infeasible.

Although it is a robust method, usually taken as a reference for comparisons, works have shown [3] that for extreme system's response scenarios, its results may not be as accurate as Anti-Optimization methodology.

2.3 Anti-optimization Interval Analysis

Consider a system (Fig. 1) with uncertain input variables $\boldsymbol{Z} = (Z_1, Z_2, \ldots, Z_d)$, and parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_{\mathrm{i}})$, the result is an output vector $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n)$. There are two special combinations of input variables and parameters within that uncertainty limits ($\overline{\boldsymbol{Z}}$, $\underline{\boldsymbol{Z}}$, $\overline{\boldsymbol{\theta}}$, and $\underline{\boldsymbol{\theta}}$) that generates the maximum and minimum outputs ($\overline{\boldsymbol{Y}}$, $\underline{\boldsymbol{Y}}$). . The under and overscore symbol means minimum and maximum values. This is the basis for the approach to uncertainty propagation problem called interval analysis. It is important to highlight that this solution is not trivial, since the extreme output is not necessarily given by the combination of extreme input ($\overline{\boldsymbol{Y}} = f\left(\boldsymbol{Z}_u^*, \boldsymbol{\theta}_u^*\right)$ and $\underline{\boldsymbol{Y}} = f\left(\boldsymbol{Z}_l^*, \boldsymbol{\theta}_l^*\right)$, where $\boldsymbol{Z}^* = (Z_1^*, Z_2^*, \ldots, Z_d^*)$ and $\boldsymbol{\theta}^* = (Z_1^*, Z_2^*, \ldots, Z_d^*)$). A double optimization can be used in order to cope with this problem, which is usually referred to as anti-optimization interval analysis.

The anti-optimization analysis can be stated as:

$$\begin{aligned}&\mathit{Find}\ \ [\mathbf{Z}_l^*, \theta_l^*]\ \mathit{and}\ [\mathbf{Z}_u^*, \theta_u^*]\ \mathit{that\ minimizes\ and\ maxmimizes\ the\ output}\ [\overline{\boldsymbol{Y}}, \underline{\boldsymbol{Y}}]\\ &\mathit{Subject\ to}:\ \boldsymbol{Z}^* \in\ [\overline{\boldsymbol{Z}}, \underline{\boldsymbol{Z}}],\ \mathit{and}\ \boldsymbol{\theta}^* \in\ [\overline{\boldsymbol{\theta}}, \underline{\boldsymbol{\theta}}].\end{aligned} \tag{13}$$

In this anti-optimization scenario, the number of design variables is $d + i$, being d the number of uncertain input variables and i the number of parameters.

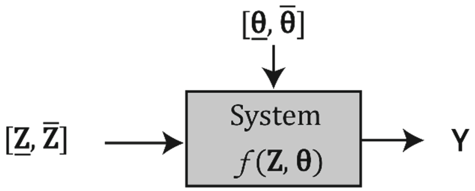

$[\underline{\boldsymbol{\theta}}, \overline{\boldsymbol{\theta}}]$
System
$f(\mathbf{Z}, \boldsymbol{\theta})$
$[\underline{\mathbf{Z}}, \overline{\mathbf{Z}}]$
Y

Fig. 1. Representation of a system with uncertain variables **Z**, parameters **θ** and output **Y**.

In this work the optimization procedure will be carried by a Spiral Optimization Algorithm (SOA) with the following optimization parameters: population $n = 20$, convergence rate $r = 0.99$, rotation angle $\beta = \pi/2$, randomization index $\alpha = 0.99$, contraction factor $\alpha_t = 0.99$, and convergence tolerance $tol = 0.01$. For more information on the anti-optimization interval analysis approach, please refer to [3].

2.4 Multivariate Generalized Polynomial Chaos Expansion (gPCE)

According to [15], the gPCE is a way of representing a stochastic process $u(z)$ parametrically. Consider a vector of uncertain variables $Z = (Z_1, Z_2, \ldots, Z_d)$ mutually independent, the gPCE expands the process using orthogonal polynomials Φ as follows

$$u^N(Z) = \sum_{|j| \leq \mathrm{N}} \hat{u}_j \Phi_j(Z), \tag{14}$$

where z represents the random variables, N is the degree of the expansion, and $\hat{u}$ is the N-th degree orthogonal gPCE projection (being $N = (n + p)!/(n!p!)$). There are two main approaches regarding the PCE: intrusive and non-intrusive. In the intrusive method, the PCE decomposition is introduced directly within the model. In the non-intrusive method, on the other hand, the model is assumed as a black box and sampled for integration at specific points. In this study only the non-intrusive method will be included since this allows easy implementation for any given system representation.

The original PCE was based on Hermite polynomials, and the application was effective only for Gaussian distributed parameters. The gPCE, however, uses orthogonal basis polynomials chosen according to the distribution of the stochastic parameters. Table 2 shows the correspondence between distributions and polynomial basis in gPCE.

Table 2. Distributions and polynomial basis. [13].

Distribution of input variables	gPCE basis polynomials	Support
Gaussian	Hermite	$(-\infty, \infty)$
Gamma	Laguerre	$[0, \infty)$
Beta	Jacobi	$[a, b]$
Uniform	Legendre	$[a, b]$

According to [13], one approach to non-intrusive method is called Stochastic Collocation Method (SCM), that relies on expansion through Lagrange interpolation polynomials.

$$\tilde{u}(Z) \equiv L(u) = \sum_{j=1}^{M} u(Z^j) h_j(Z), \tag{15}$$

where h_j are the Lagrange polynomials. It is important to note that $h_i\left(Z^j\right) = \delta_{ij}$ for $i, j \in [1, \ldots, M]$ (δ being Kronecker delta), so the interpolation $\tilde{u}(Z)$ is equal to the exact solution in each of the M collocation points. A set of collocation points are chosen $Z_M = \left\{Z^j\right\}_{j=1}^{M}$, then each of the nodes Z^j are solved for $j = 1, \ldots, M$ in a form of deterministic system. The points can be selected by Monte Carlo simulation, but usually a clever choice of collocation points is made (following the quadrature rule for example). This approach is called Multidimensional Colocation PCE.

The mean of the interpolation $\hat{u}$ can be computed as

$$E(\tilde{u}) = \sum_{j=1}^{M} u(Z^j) \int_T h_j(Z) \rho(z) dz, \tag{16}$$

where Γ is the random space in wich Z is defined and $\rho(z)$ is a distribution specific weight.

Using the quadrature rule to evaluate the integral, Eq. (16) gives

$$E(\tilde{u}) = \sum_{j=1}^{M} u(Z^j) \sum_{k=1}^{M} h_j(Z_k) \rho(z_k) wz, \tag{17}$$

where z_k and w_k are the quadrature points and weights respectively. Since the quadrature points are chosen as collocation points, previous formulation finally becomes Eq. (18)

$$E(\tilde{u}) = \sum_{j=1}^{M} u(Z^j) \rho(Z^j) w_j, \tag{18}$$

According to [14], the gPCE coefficients can be used to estimate the statistics of a random process directly as

$$\mu = E\left[\sum_{|j| \le N} \hat{u}_j \Phi_j(z)\right] = \hat{u}_0 \tag{19}$$

$$\sigma^2 = E\left[\left(\sum_{|j| \le N} \hat{u}_j \Phi_j(z)\right)^2\right] = \sum \hat{u}^2 \tag{20}$$

where μ denotes mean value and σ^2 the variance.

3 Simulation, Results and Discussion

For the following examples, a [±45] graphite/epoxy composite with a thickness of 0.25 mm is used. As mentioned, the dimension of the composite plate is 0.15 m x 0.05 m with a total of 144 finite elements, and the boundary condition is FFFC (one clamped edge and the others free, a cantilever plate). The composite is excited by a sustained 0.1 N force applied to the middle of the tip of the rectangular plate and maintained throughout the simulation, as shown in Fig. 2. The damping ratios for the first four vibration modes are taken as $\zeta_1 = 2 \times 10^{-2}$, $\zeta 2 = 3 \times 10^{-2}$, $\zeta 3 = 6 \times 10^{-2}$, $\zeta 4 = 1 \times 10^{-1}$.

Each simulation lasts 0.3 s and is solved in the time domain using Newmark method with $\Delta \mathrm{t} = 1 \times 10^{-3}$ s. The algorithm is implemented in MATLAB R2012a and run in an Intel® Core™ i5-9600KF 3.7 GHz CPU with 16 GB RAM in approximately 0.2 s. A modal analysis of the composite structure gives the vibration mode shapes presented in Fig. 3.

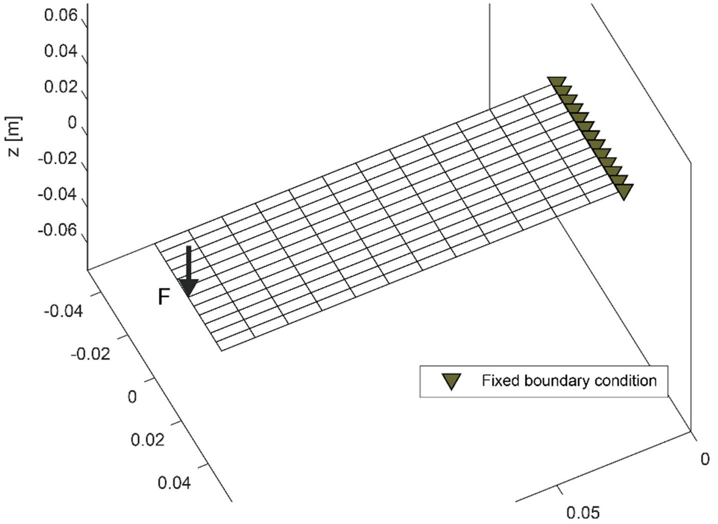


Fig. 2. Finite element mesh representation (clamped in one edge). F denotes the load.

For comparison purposes the mean and variance results obtained with gPCE will be used to find the envelopes assuming $\underline{Y} = (2\mu - \sqrt{12\sigma^2})/2$ and $\overline{Y} = (2\mu + \sqrt{12\sigma^2})/2$. Also, with the aim of a fair comparison, the number of Monte Carlo simulations will be limited according to the number of evaluations of the anti-optimization analysis.

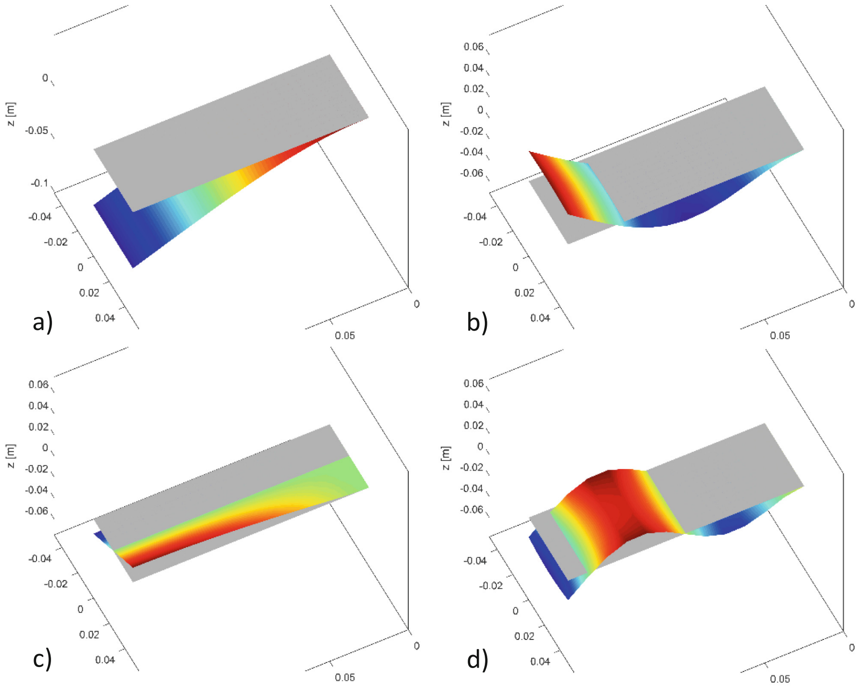


Fig. 3. a) first, b) second, c) third, and d) forth vibration modes.

The uncertainty analysis results will be presented in 4 subsections: uncertainty in fiber orientation angle, excitation load, damping ratio, and finally, an example regarding a higher number of composite layers.

3.1 Uncertainty in Fiber Orientation Angle

For this problem, the aforementioned configuration is subjected to a ±5° uncertainty in each lamina (uniform distribution). Figure 4 shows the uncertainty envelopes for MC, anti-optimization interval analysis and multivariate colocation gPCE. The vertical displacement is measured at the same point where the excitation force is applied (see Fig. 2).

From Fig. 4, it can be observed that both Monte Carlo and Anti-optimization approaches produced close results, although the anti-optimization envelop is broader for the same number of function calls (12600). The PCE method captured the general envelop behavior efficiently, given it consumed very little function calls (only 22 for a

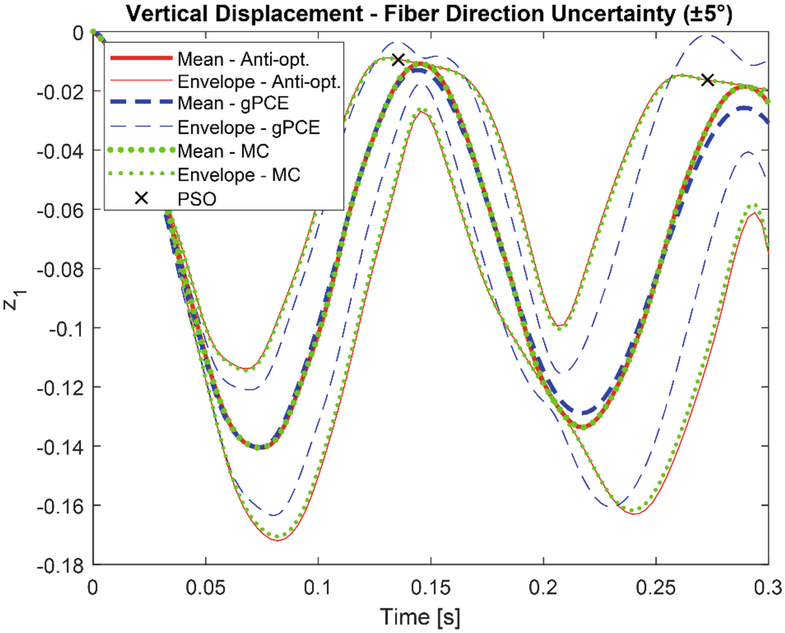


Fig. 4. Displacement envelope for ±5° uncertainty in fiber orientation.

level 2 polynomial expansion). However, its precision is comparatively lower than the other methods.

For the first valley ($t = 0.08$ s), the percentual increase in maximum vertical displacement due to uncertainties found by gPCE, Monte Carlo, and anti-optimization, was respectively 16.3%, 21.4% e $22.4\%([\mu - \text{envelopevalue}]/\mu)$.

To check main discrepancies in Fig. 4, it was selected 2 points marked with a cross. These points were exhaustively anti-optimized to find the maximum value for the displacement. The particle swarm algorithm (PSO) was used as an independent optimization engine. It was found that there was no combination of the input uncertain variables capable of further expanding the envelopes in these points, as indicated by the gPCE. This suggests that the overshoot found by gPCE is most likely due to an imprecision of the method. Moreover, increasing the level of polynomial expansion up to 6 did not result in a noticeable improvement in envelope accuracy. Furthermore, it is worth noting that regions where the anti-optimization obtained a broader envelope do not raise any concern regarding imprecisions. This is because the anti-optimization provides the exact input configuration that generated each solution, which is a significant advantage over the gPCE.

For replicability purposes, the PSO algorithm parameters used in the validation tests were: population $n = 100$, inertial moment $\omega = 0.9$, cognitive components 1 and 2 $c_1 = c_2 = 2.01$, convergence tolerance $tol = 10^{-6}$, initial mutation coefficient $\alpha = 0.9$ and mutation decay $\alpha_t = 0.01$.

3.2 Uncertainty in the Excitation Load

The composite plate is now subjected to a ±20% uncertainty in the excitation load (uniform distribution). The results can be checked in Fig. 5.

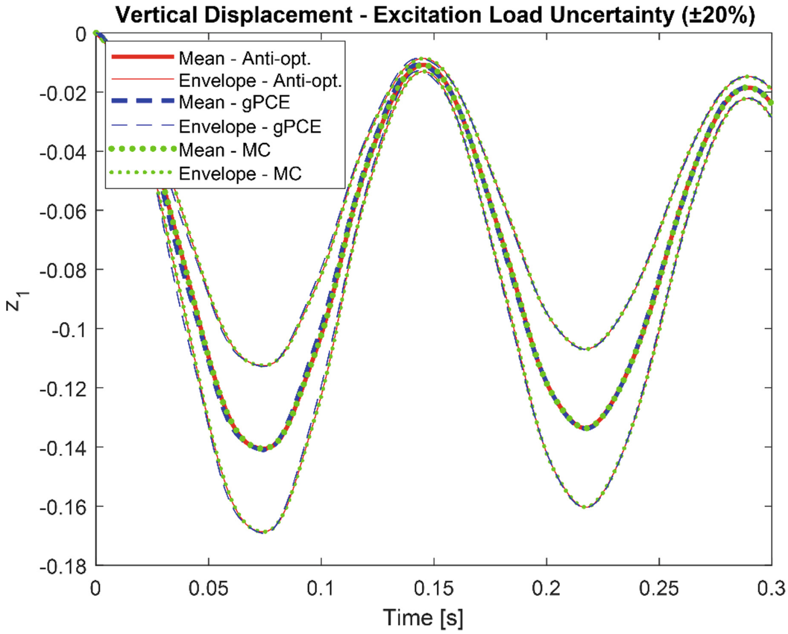


Fig. 5. Displacement envelope for ±20% uncertainty in load.

All uncertainty propagation approaches generated similar envelopes in this example, likely because of the low complexity level of the problem. For the first valley, the percentual increase in maximum vertical displacement due to uncertainties was 20% for all methods, as expected. The number of function evaluations for gPCE was 7, against 10600 for Monte Carlo and anti-optimization solutions.

3.3 Uncertainty in Damping Ratio

The aforementioned configuration is now subjected only to a ±40% uncertainty in the damping ratio (uniform distribution). Figure 6 shows the results for each approach.

From Fig. 6 one can note that the damping ratio uncertainty generated narrow envelopes when compared to load and fiber orientation examples. Thus, the structure sensitivity to damping is lower in comparison to the already analyzed uncertain variables.

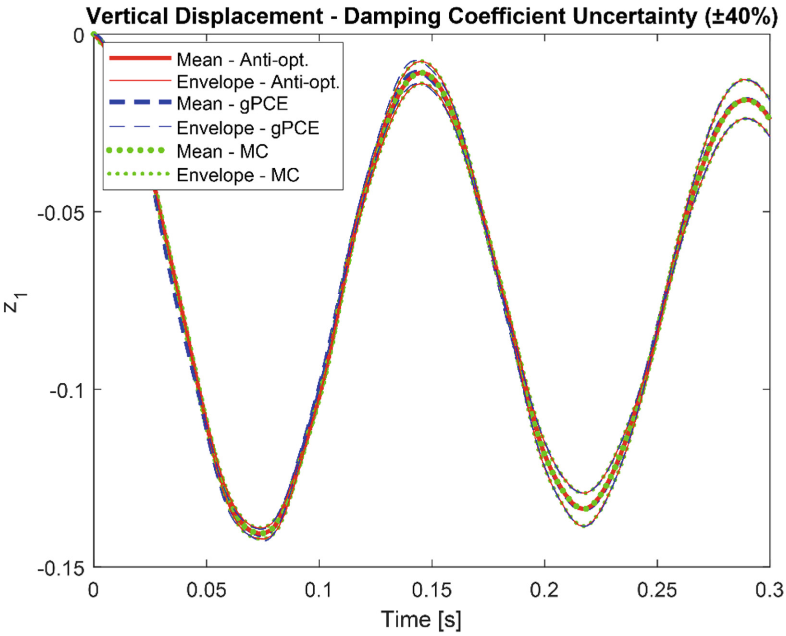


Fig. 6. Displacement envelope for ±40% uncertainty in damping coefficient.

3.4 Uncertainty in Fiber Orientation with Increasing Number of Laminae

In previous examples Monte Carlo and anti-optimization approaches generated similar results. In order to have a higher distinction between envelopes, a more challenging problem is devised. For this reason, the configuration used in this last example has been modified to accommodate a higher number of uncertain variables.

This last problem is a variation of the first one presented in this work, being a $[45_4/-45_4]$ graphite/epoxy composite with ±5° uncertainty in each lamina (uniform distribution). Figure 7 shows the results obtained by each approach.

Compared to the previous examples, the increase in the number of uncertain variables successfully highlighted the differences between approaches. In this problem, a level 3 polynomial expansion was used in gPCE (1408 evaluations), resulting in a more accurate solution compared to the level 2 expansion used in previous examples. Expansions up to level 6 were tested and did not yield improvements in precision. These higher expansion levels required 9376, 54672, and 287408 evaluations for levels 4, 5, and 6, respectively.

The anti-optimization approach resulted in a broader envelope and thus a better solution, as shown in Fig. 7. The number of evaluations for Monte Carlo and anti-optimization was 12600. For the first valley, $t = 0.8$ s, the percentual increase in maximum vertical displacement due to uncertainties found by gPCE, Monte Carlo, and anti-optimization, was respectively 14.8%, 19.6% e 23.6% (evaluated as described in example 3.1).

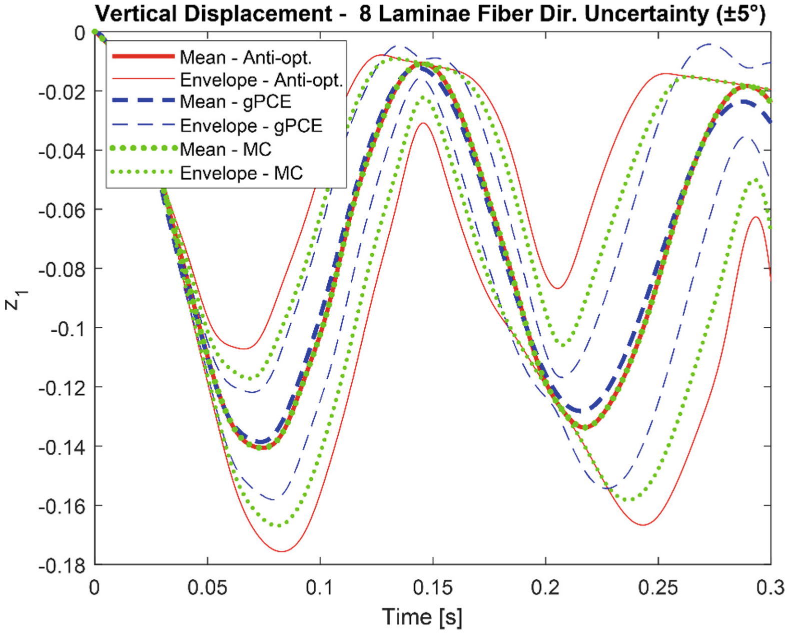


Fig. 7. Displacement envelope for ±5° uncertainty in fiber orientation at each of the 8 laminae.

4 Conclusion

In this work a 0.15 m x 0.05 m graphite/epoxy composite was simulated in the time domain using finite element analysis. This composite structure was used in 4 different uncertainty propagation problems, where gPCE, Monte Carlo, and anti-optimization were used to generate the vertical displacement envelopes.

The results show that gPCE is an extremely powerful tool, able to solve the problems with a minimum number function evaluations. Despite being precise in the simpler problems, gPCE was not able to compete satisfactorily against the accuracy of the other two methods in the last and more complex example. The narrower envelope resulting of this approach in the mentioned test translates into imprecisions that go against safety.

Anti-optimization, on the other hand, showed to be a reliable alternative for more complex problems, although costing much more function evaluations than gPCE. The ability to provide the exact combination used to generate each solution of the envelope is a solid advantage for the anti-optimization, which guarantees that every envelope solution is an actual combination of the uncertain variables.

Acknowledgments. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. Funding was also provided by National Council for Scientific and Technological Development CNPq (Grant no. 141260–2021-1). The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Adhikari, S.N., Dey, S.: Uncertainty Quantification in Laminated Composites: A Meta-Model Based Approach. CRC Press, Boca Raton (2021)
2. Dey, S., Mukhopadhyay, T., Adhikari, S.: Stochastic free vibration analysis of angle-PLY composite plates – a RS-HDMR approach. Compos. Struct. **122**, 526–536 (2015). https://doi.org/10.1016/j.compstruct.2014.09.057
3. Santana, P.B., Grotti, E., Gomes, H.M.: An efficient anti-optimization approach for uncertainty analysis in composite laminates. Mater. Res. **24**(Suppl 2) (2021). https://doi.org/10.1590/1980-5373-mr-2021-0334
4. Adali, S., Lene, F., Duvaut, G., Chiaruttini, V.: Optimization of laminated composites under buckling uncertainties via anti-optimization. In: 9th AIAA/ISSMO Symposium on Multidisciplinary Analysis and Optimization (2002). https://doi.org/10.2514/6.2002-5417
5. Kim, T., Hwang, I.H., Sin, H.: Optimal design of composite laminate with uncertainty in loading and material properties considered. Appl. Mech. (2005). https://doi.org/10.1115/imece2005-79251
6. Lee, J., Haftka, R.T., Griffin, O.H., Watson, L.T., Sensmeier, M.D.: Detecting delaminations in a composite beam using anti-optimization. Struct. Optim. **8**(2–3), 93–100 (1994). https://doi.org/10.1007/bf01743304
7. Elishakoff, I., Ohsaki, M.: Optimization and Anti-optimization of Structures Under Uncertainty. Imperial College Press, London (2010)
8. Chen, M., Zhang, X., Shen, K., Pan, G.: Sparse polynomial chaos expansion for uncertainty quantification of composite cylindrical shell with geometrical and material uncertainty. J. Mar. Sci. Eng. **10**(5), 670 (2022). https://doi.org/10.3390/jmse10050670
9. García-Merino, J., Calvo-Jurado, C., García-Macías, E.: Polynomial Chaos expansion for uncertainty propagation analysis in numerical homogenization of 2D/3D periodic composite microstructures. Compos. Struct. **300**, 116130 (2022). https://doi.org/10.1016/j.compstruct.2022.116130
10. Thapa, M., Missoum, S.: Uncertainty quantification and global sensitivity analysis of composite wind turbine blades. Reliab. Eng. Syst. Saf. **222**, 108354 (2022). https://doi.org/10.1016/j.ress.2022.108354
11. Cacuci, D.G.: Sensitivity and Uncertainty Analysis. Chapman & Hall/CRC Press, Boca Raton (2003)
12. Hirsch, C., Wunsch, D., Szumbarski, J., Łaniewski-Wołłk, Ł, Pons-Prats, J.: Uncertainty Management for Robust Industrial Design in Aeronautics. Findings and Best Practice Collected Dring UMRIDA, a Collaborative Research Project (2013–2016) Funded by the European Union. Springer, Cham (2019).https://doi.org/10.1007/978-3-319-77767-2
13. Kærgaard, E.B.: Spectral Methods for Uncertainty Quantifcation (Unpublished master's thesis). Technical University of Denmark – DTU (2013)
14. Novak, L., Novak, D.: Polynomial Chaos expansion for surrogate modelling: theory and software. Beton- Und Stahlbetonbau **113**, 27–32 (2018). https://doi.org/10.1002/best.201800048
15. Smith, R.C.: Uncertainty Quantification: Theory, Implementation, and Applications. Society for Industrial and Applied Mathematics, Philadelphia (2014)
16. Ferreira, A.J.M.: MATLAB Codes for Finite Element Analysis. Springer, Cham (2020)
17. Kaw, A.K.: Mechanics of Composite Materials. CRC, Taylor & Francis, Boca Raton (2006)

On the Collaboration Between Bayesian and Hilbertian Approaches

Eduardo Souza de Cursi[1] and Adriano Fabro[2]

[1] Laboratoire de Mécanique de Rouen, INSA Rouen, Normandie Université, Rouen, France
souza@insa-rouen.fr
[2] Department of Mechanical Engineering, University of Brasilia, Brasília, Brazil
fabro@unb.br

Abstract. In this work, we explore the use of Uncertainty Quantification (UQ) techniques of representation in Bayes estimation and representation. UQ representation is a Hilbertian approach which furnishes distributions from experimental data in limited number. It can be used to generate priors to be used by Bayesian procedures. In a first use, we consider De Finetti's representation theorem with few data points and we show that the UQ methods can furnish interesting priors, able to reproduce the correct distributions when integrated in the De Finetti's representation theorem. In a second use, we consider Bayes estimation of the parameters of a distribution. Analogously to the preceding situation, a limited sample is used to generate a UQ representation of the parameters. Then, we use it as prior for the Bayesian procedure. The results show that the approach improves the quality of the estimation, when compared to the standard Bayesian procedure. The results are also compared to Fisher's procedure of estimation.

Keywords: Uncertainty Quantification · Bayesian Inference · Hilbert Expansions

1 Introduction

A current task in Statistics consists in the determination of distributions corresponding to observed data. For instance, we can have a sample formed by n variates from a quantity X:

$$\mathcal{X} = (X_1, \ldots, X_n) \tag{1}$$

and desire to determine the distribution of X. Between 1912 and 1922, R. A. Fisher introduced a method for the solution of this problem, nowadays known as Maximum of Likelihood Estimation (MLE) [1–4]. In Fisher's approach, the user must furnish a model for the distribution of X, defined by a Probability Density Function (PDF) $f(x, \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a vector of unknown parameters to be determined - f is referred as the *model*. The *Likelihood* associated to the sample is

$$L(\mathcal{X}, \boldsymbol{\theta}) = \prod_{i=1}^{n} f(X_i, \boldsymbol{\theta}) \tag{2}$$

J. E. S. De Cursi (Ed.): UQSM 2023, LNME, pp. 178–189, 2024.
https://doi.org/10.1007/978-3-031-47036-3_16

and the MLE furnishes an estimator $\widehat{\boldsymbol{\theta}}$ which maximizes the Likelihood:

$$\widehat{\boldsymbol{\theta}} = \operatorname{argmax}\{L(\mathcal{X}, \boldsymbol{\theta}) : \boldsymbol{\theta}\}. \tag{3}$$

By the same period, K. Pearson has shed new lights on the works of Thomas Bayes, Condorcet, and Laplace about inverse probability [5–7]. These works introduced a second method for the solution of the problem, which is nowadays known as Bayesian Inference (BI), and was considered as radically different from Fisher's approach. In BI, the parameters $\boldsymbol{\theta}$ are considered as random variables and the user must furnish a second model $g_{pr}(\boldsymbol{\theta})$ for their distribution, usually referred as the *Prior Distribution*. The Bayesian approach updates the Prior using the Bayes Formula:

$$f_{po}(\boldsymbol{\theta}|\mathcal{X}) = Af(\mathcal{X}|\boldsymbol{\theta})f_{pr}(\boldsymbol{\theta}). \tag{4}$$

f_{po} is called the *Posterior Distribution*. Here, A is a constant destined to normalize the product $f(\mathcal{X}|\boldsymbol{\theta})f_{pr}(\boldsymbol{\theta})$ to get a probability density (its integral must be equal to the unity). If the variates in $\mathcal{X}$ are independent, - id est, if $\mathcal{X}$ is a sample – then $f(\mathcal{X}|\boldsymbol{\theta})$ is the Likelihood:

$$f(\mathcal{X}|\boldsymbol{\theta}) = L(\mathcal{X}, \boldsymbol{\theta}). \tag{5}$$

Then, the estimation of the parameters $\boldsymbol{\theta}$ is made by minimizing the mean value of a loss function l:

$$\widehat{\boldsymbol{\theta}} = \operatorname{argmin}\{E(l(\boldsymbol{\theta}, \boldsymbol{\eta})) = \int l(\boldsymbol{\theta}, \boldsymbol{\eta})f_{po}(\boldsymbol{\theta})\boldsymbol{d\theta} : \boldsymbol{\eta}\}. \tag{6}$$

Many loss functions can be found in the literature [8], such as, for instance, the quadratic one:

$$l(\boldsymbol{\theta}, \boldsymbol{\eta}) = \|\boldsymbol{\theta} - \boldsymbol{\eta}\|^2, \tag{7}$$

or the generalized LINEXP [9, 10]:

$$l(\theta, \eta) = \sum\nolimits_i (exp(\phi(\eta_i, \theta_i)) - \phi(\eta_i, \theta_i) - 1), \tag{8}$$

Here, ϕ is a conveniently chosen function [10].

In the last years, Uncertainty Quantification (UQ) tools were proposed to determine the distribution of a random variable from a sample. For instance, the Hilbert Approach (HA), which can be considered as an extension of Polynomial Chaos Approximations (PCA). HA considers a Hilbert basis $\{\varphi_i : i \in \mathbb{N}\} \subset L^2(\Omega)$, where Ω is the domain of the possible values of X. If no supplementary information is available, Ω is estimated as the range of the observed values in the sample. Then, HA looks for a representation

$$X = \sum\nolimits_{i \in \mathbb{N}} x_i \varphi_i(U) \approx P_k X = \sum\nolimits_{i=0}^{k} x_i \varphi_i(U), \tag{9}$$

where U is a convenient random variable – if no information about the source of randomness is available, U can be an artificial variable, chosen by the user. In this last situation, the data can be ranged in an increasing order: $X_1 \leq X_2 \leq \cdots \leq X_n$, and $U_1 \leq U_2 \leq \cdots \leq U_n$, to generate a non-negative covariance between X and the artificial U.

In practice, only the approximation $P_k X$ is determined, by determining the coefficients $\boldsymbol{x} = (x_0, x_1, \ldots, x_k)^t$. We can find in the literature works dealing with such a representation, such as, for instance, collocation, which considers a sample $U = (U_1, \ldots, U_n)$ and solves the linear system:

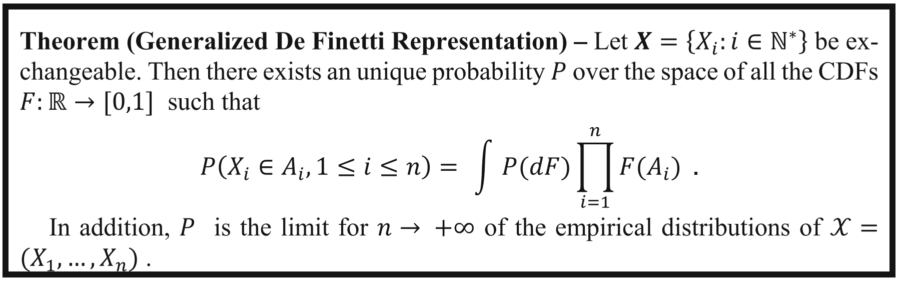
$$\boldsymbol{Mx} = \boldsymbol{N}, M_{ij} = \varphi_j(U_i), N_i = X_i, 1 \leq i \leq n, 0 \leq j \leq k. \tag{10}$$

In general, $k + 1 < n$ (the system is overdetermined) and (10) must be solved by a least squares approach. Once $P_k X$ is determined, we can generate a large sample of U, $\mathcal{U}_g = \left(U_1, \ldots, U_{ng}\right)$ and use it to generate a large sample $\mathcal{X}_g$ from X as $X_i = P_k X(U_i)$, $i = 1, \ldots, ng$. Then, $\mathcal{X}_g$ is used to estimate the CDF and PDF of X, by using the empirical CDF of the large sample. In the examples below, we use $ng = 1E5$.

HA can be easily combined to BI, by generating Priors or Models. In the sequel, we shall examine these combinations and compare their results to MLE and the single BI.

2 Collaboration Between Hilbert and De Finetti Representations

In 1930, Bruno De Finetti introduced the notion of exchangeability, which he considered preferable to the notion of independency [11, 12]. Recall that $\boldsymbol{X} = (X_1, \ldots, X_n)$ is exchangeable if and only if the distribution of $\boldsymbol{X}_{\boldsymbol{\Pi}} = \left(X_{\Pi(1)}, \ldots, X_{\Pi(n)}\right)$ coincides with the distribution of $\boldsymbol{X}$, for any permutation Π of $\{1, \ldots, n\}$. Analogously, $\boldsymbol{X} = \{X_i : i \in \mathbb{N}^*\}$ is exchangeable if and only if $\left(X_{i_1}, \ldots, X_{i_n}\right)$ is exchangeable, $\forall n \in \mathbb{N}^*$, $\{i_1 < i_2 < \ldots < i_n\} \in (\mathbb{N}^*)^n$. [13]. Among the properties of Exchangeable sequences, we have the Generalized De Finetti's Theorem [13, 14]:



Theorem (Generalized De Finetti Representation) – Let $\boldsymbol{X} = \{X_i : i \in \mathbb{N}^*\}$ be exchangeable. Then there exists an unique probability P over the space of all the CDFs $F: \mathbb{R} \rightarrow [0,1]$ such that

$$P(X_i \in A_i, 1 \leq i \leq n) = \int P(dF) \prod_{i=1}^{n} F(A_i) \ .$$

In addition, P is the limit for $n \rightarrow +\infty$ of the empirical distributions of $\mathcal{X} = (X_1, \ldots, X_n)$.

In practice, this theorem is applied using the version below:

Theorem (Practical De Finetti Representation) Let $\boldsymbol{X} = \{X_i : i \in \mathbb{N}^*\}$ be exchangeable, with marginal density $f(x, \boldsymbol{\theta})$. Then there exists an unique probability P such that the density $f_{(n)}$ of $\boldsymbol{X}_{(\boldsymbol{n})} = (X_1, \ldots, X_n)$ verifies

$$f_{(n)}(x_1, \ldots, x_n) = \int P(d\boldsymbol{\theta}) \prod_{i=1}^{n} f(x_i, \boldsymbol{\theta}) \ .$$

De Finetti Representations can be used for estimation. For instance, let us consider a sample formed by 10 variates from the exponential distribution *Exp* (3):

$$\mathcal{X} = (0.615, 5.543, 1.266, 1.044, 2.472, 3.862, 0.858, 0.521, 3.135, 7.737). \tag{11}$$

Assume that we do not know that this data comes from the Exponential Distribution and we suppose -erroneously – that the data comes from a Gaussian distribution $N(m, \sigma)$. Then, we consider $\boldsymbol{\theta} = (m, \sigma)$ and

$$f(x, \boldsymbol{\theta}) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}. \tag{12}$$

In this case, MLE furnishes the results shown in Table 1. The unbiased estimation for σ is $\widehat{\sigma}_{obs,unb} \approx 2.4$, close to the MLE. If we are interested also in the evaluation of $P(X \le i)$, $i = 1, 2, 3$, we can use the MLE estimates as parameters in the CDF associated to the model density (12) and determine the corresponding values. Alternatively, we can integrate the density $f_{(1)}$ to determine these probabilities – see the results in Table 1.

Now, assume that we believe that $m \in (2, 5)$ and $\sigma \in (2, 6)$. We can use the generalized De Finetti's theorem to generate the PDF $f_{(1)}(x)$. If the density associated to P is p, we have

$$f_{(1)}(x) = \int f(x, \boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}. \tag{13}$$

The CDF $F_{(1)}(x)$ can be generated by integrating $f_{(1)}$, and we have $P(X \le i) = F_{(1)}(i) = \int_{-\infty}^{i} f_{(1)}(x)dx$. In practice, the lower bound of the integral is taken as a convenient finite negative number – we use -10 as lower bound. We can also use $f_{(1)}$ to estimate

$$m_{(1)} \approx \int x f_{(1)}(x)dx, \sigma^2 \approx \int \left(x - m_{(1)}\right)^2 f_{(1)}(x)dx. \tag{14}$$

As an example, let us assume that our belief is that p corresponds to an uniform distribution – which ***is not*** the distribution corresponding to De Finetti's theorem. Then, we obtain the results in Table 1 and the distribution exhibited in Fig. 1.

However, in the Bayesian approach, we can modify our prior. For instance, let us use a prior more adapted to De Finetti's theorem - a gaussian distribution or a gaussian for m and a $\chi^2(9)$ for σ.The results appear in Table 1.

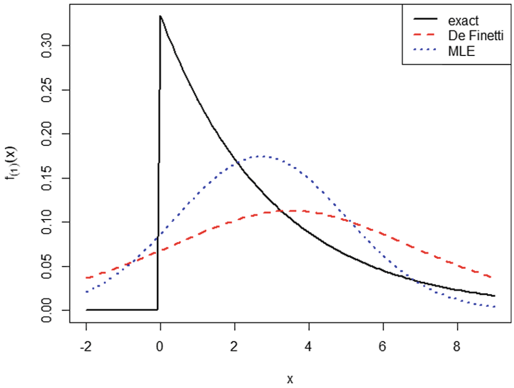


Fig. 1. Example of Application of De Finetti's theorem with an uniform belief for m and σ.

Table 1. Examples of Estimation results, with a gaussian model.

	$P(X \le 1)$		$P(X \le 2)$		$P(X \le 3)$		m	σ
	model	$f_{(1)}$	model	$f_{(1)}$	model	$f_{(1)}$		
Exact	0.2835		0.4866		0.6321		3	3
MLE	0.2280		0.3790		0.5513		2.7	2.3
De Finetti Uniform	0.2930	0.2731	0.3944	0.3676	0.5030	0.4750	3.0	3.6
De Finetti Gauss	0.3159	0.2964	0.4269	0.4060	0.5441	0.5269	2.6	3.3
De Finetti Gauss/χ^2	0.2984	0.2843	0.4132	0.3978	0.5360	0.5232	2.7	3.2

Table 2. Examples of Estimation results, with a gaussian model, using a confidence interval $\alpha = 0.05$.

	$P(X \le 1)$		$P(X \le 2)$		$P(X \le 3)$		m	σ
	model	$f_{(1)}$	model	$f_{(1)}$	model	$f_{(1)}$		
Exact	0.2835		0.4866		0.6321		3	3
De Finetti Uniform	0.2954	0.2820	0.4206	0.4085	0.5546	0.5478	2.6	3.0
De Finetti Gauss	0.2858	0.2709	0.4149	0.4022	0.5540	0.5490	2.6	2.9
De Finetti Gauss/χ^2	0.2969	0.2875	0.4209	0.4112	0.5533	0.5463	2.6	3.0

We can also use confidence intervals for m and σ instead of a "belief" interval, as in the preceding. For instance, using confidence intervals with $\alpha = 0.05$, we obtain the results in Table 2. Results for $\alpha = 0.01$ appear in Table 3.

A collaboration between the Hilbertian and Bayesian approaches can be implemented by using the HA to generate a prior distribution for the couple (m, σ): we start by generating samples from m and σ – for instance, we can use a bootstrap to generate

Table 3. Examples of Estimation results, with a gaussian model, using a confidence interval $\alpha = 0.01$.

	$P(X \leq 1)$		$P(X \leq 2)$		$P(X \leq 3)$		m	σ
	model	$f_{(1)}$	model	$f_{(1)}$	model	$f_{(1)}$		
Exact	0.2835		0.4866		0.6321		3	3
De Finetti Uniform	0.3338	0.3154	0.4468	0.4286	0.5644	0.5502	2.5	3.4
De Finetti Gauss	0.3094	0.2854	0.4313	0.4110	0.5602	0.5514	2.5	3.1
De Finetti Gauss/χ^2	0.3156	0.3015	0.4338	0.4196	0.5583	0.5479	2.5	3.2

samples $\mathcal{M}$ and $\mathcal{S}$ of nb variates from m and σ, respectively. These samples can be used to determine finite expansions $P_k m(U)$, $P_k \sigma(U)$, where U is a convenient random variable. Then, the expansion can be used to determine the PDFs p_m, p_σ, of m and σ, respectively (as indicated in the preceding, we generate $ng = 1E5$ variates from each variable to estimate their distributions). A possible prior is $p(m, \sigma) = p_m(m) p_\sigma(\sigma)$, which can be used in the Bayesian Approach.

As an alternative, we can use HA to generate a representation $P_k X(U)$ of X, and use this representation to generate samples from m and σ, which are used to determine $P_k m(U)$, $P_k \sigma(U)$. As previously indicated, we generate $ng = 1E5$ variates from each variable to estimate their distributions.

As an example, let us consider a variable $U \sim N(0, 1)$ and determine approximations with $nb = 500$, $k = 3$ and $\varphi_i(U) = s^{i-1}$, $s = (U - a)/(b - a)$, where $a = \text{minU}$, $b = \text{maxU}$ – these values are chosen to get $0 \leq s \leq 1$. The coefficients of the expansion were calculated by collocation [15–17]. This approach produces the results shown in Fig. 2 and Table 4. For $k = 5$, we obtained the results shown in Table 5.

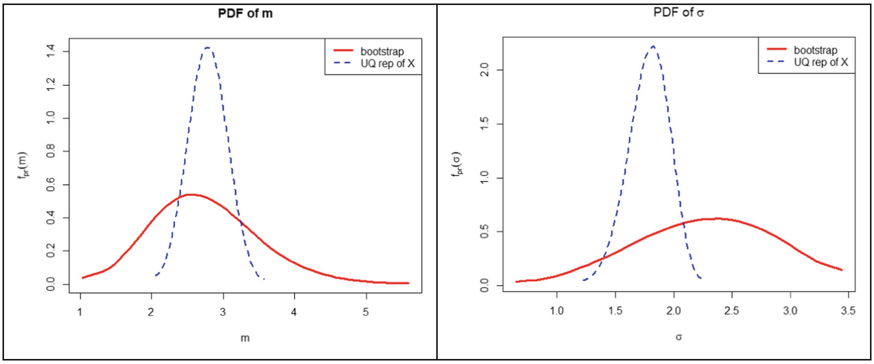


Fig. 2. Examples of Priors determined by HA.

Table 4. Examples of estimation results using a gaussian model, and a prior generated by HA, $k = 3, nb = 500$.

	$P(X \leq 1)$		$P(X \leq 2)$		$P(X \leq 3)$		m	σ
	model	$f_{(1)}$	model	$f_{(1)}$	model	$f_{(1)}$		
Exact	0.2835		0.4866		0.6321		3	3
From $P_k m, P_k \sigma$ (bootstrap)	0.2402	0.2263	0.3936	0.3840	0.5657	0.5701	2.6	2.3
From $P_k X$ (representation)	0.2292	0.2292	0.3877	0.3874	0.5679	0.5675	2.6	2.2

Table 5. Examples of estimation results using a gaussian model and a prior generated by HA, $k = 5, nb = 500$.

	$P(X \leq 1)$		$P(X \leq 2)$		$P(X \leq 3)$		m	σ
	model	$f_{(1)}$	model	$f_{(1)}$	model	$f_{(1)}$		
Exact	0.2835		0.4866		0.6321		3	3
From $P_k m, P_k \sigma$ (bootstrap)	0.2386	0.2264	0.3850	0.3759	0.5502	0.5523	2.7	2.4
From $P_k X$ (representation)	0.2380	0.2374	0.4788	0.4787	0.7278	0.7284	2.1	1.5

Of course, the results are improved if we know that the distribution of X is Exponential. In this case, we can use an exponential model and an exponential variable U: obtain the results in Table 6

$$f(x, m) = me^{-x/m}, U \sim Exp(m_{obs}), \tag{15}$$

where $\mathrm{m_{obs}}$ is the empirical mean. In this situation, we obtain the results in Table 6 and Fig. 3. Notice that the results generated by $P_k X(U)$ are close to the exact ones.

Table 6. Examples of Estimation results, with an exponential model, $k = 3, nb = 500$.

	$P(X \leq 1)$		$P(X \leq 2)$		$P(X \leq 3)$		m
	model	$f_{(1)}$	model	$f_{(1)}$	model	$f_{(1)}$	
Exact	0.2835		0.4866		0.6321		3
MLE	0.3090		0.5225		0.6701		2.7
De Finetti Uniform	0.3516	0.3569	0.5795	0.5999	0.7273	0.7127	2.2
From $P_k m$ (bootstrap)	0.3462	0.3355	0.5725	0.5499	0.7205	0.6988	2.9
From $P_k X$ (representation)	0.2895	0.2244	0.4951	0.3997	0.6413	0.5439	2.3

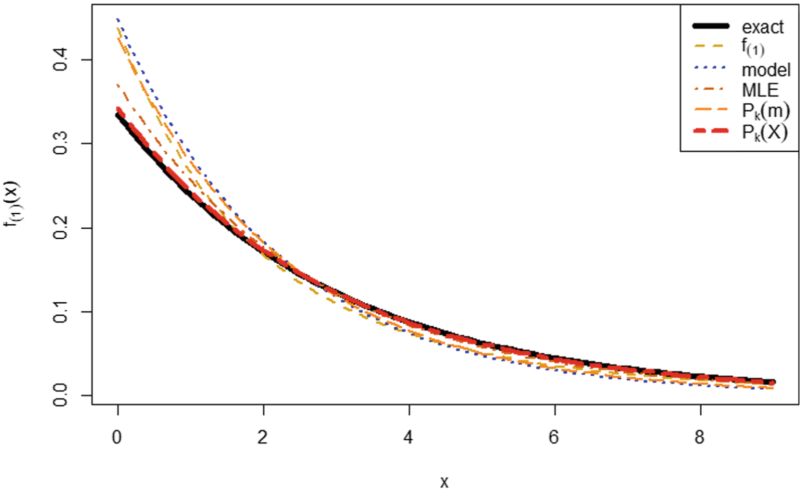


Fig. 3. Examples of resulting PDFs generated by the De Finetti's theorem with an exponential model. The PDF generated by using the prior generated by $P_k X$ is almost identical to the exact one.

3 Collaboration Between Hilbert and Bayesian Update

Applying De Finetti's theorem can be computationally expensive, since a large number of integrals has to be calculated. A simpler approach consists in use the Bayesian Update (4) to determine a Posterior Distribution f_{po} for the parameters $\boldsymbol{\theta}$. Then, we can use f_{po} and a loss function to estimate the parameters. Using the values determined, the model furnishes the probabilities.

Let us consider again the situation presented in the previous section. We consider the classical quadratic loss function (7) and GLINEXP with

$$\phi(\eta, \theta) = \frac{\eta}{\theta} - 1, \tag{16}$$

Analogously to the preceding section, we start by choosing a model for the distribution of X. In a second step, we generate a prior for m and σ, from samples $\mathcal{M}$ and $\mathcal{S}$ of nb variates from each variable. These samples are generated by bootstrap or from $P_k X(U)$ – whatever the choice, the samples are used to construct $P_k m(U)$, $P_k \sigma(U)$, and these expansions are used to generate the prior f_{pr}, and the posterior f_{po}, by the Bayesian Update (4).

For instance, let us consider the gaussian model (12) – which is erroneous, since the data comes from an exponential distribution. We can apply the method above, and we obtain results as shown in Table 7. The HA/BI approach was compared to the Uniform Prior and to two classical noninformative priors for Gaussian models: Jeffreys' Prior $1/\sigma^2$ [18] and the hierarchical Jeffreys Prior $1/\sigma$ [18].

Again, if the information about the distribution of X is available, we can use an exponential model for X. In this case, the results improve, as shown in Table 8. Notice that the results furnished by the Prior generated by $P_k X$ are almost exact in this situation.

Table 7. Examples of estimation results using a gaussian model.

Prior	Loss Function	$P(X \leq 1)$	$P(X \leq 2)$	$P(X \leq 3)$	m	σ
Exact		0.2835	0.4866	0.6321	3	3
Uniform, $\alpha = 0.05$	Quadratic (7)	0.2638	0.3969	0.5435	2.7	2.7
Uniform, $\alpha = 0.05$	GLINEXP (8)	0.2461	0.3662	0.5012	3.0	2.9
Jeffreys	Quadratic (7)	0.2444	0.3873	0.5476	2.7	2.5
Jeffreys	GLINEXP (8)	0.2282	0.3573	0.5052	3.0	2.6
Jeffreys (hierarchical)	Quadratic (7)	0.2463	0.3882	0.5472	2.7	2.5
Jeffreys (hierarchical)	GLINEXP (8)	0.2281	0.3547	0.5000	3.0	2.7
From $P_k m$, $P_k \sigma$ (bootstrap)	Quadratic (7)	0.2705	0.4079	0.5557	2.6	2.6
From $P_k m$, $P_k \sigma$ (bootstrap)	GLINEXP (8)	0.2695	0.3822	0.5661	2.6	2.6
From $P_k X$ (representation)	Quadratic (7)	0.2720	0.4067	0.5565	2.6	2.6
From $P_k X$ (representation)	GLINEXP (8)	0.2724	0.4093	0.5584	2.6	2.6

Table 8. of estimation results using an exponential model.

Prior	Loss Function	$P(X \leq 1)$	$P(X \leq 2)$	$P(X \leq 3)$	m
Exact		0.2835	0.4866	0.6321	3
Uniform (Jeffreys), $\alpha = 0.05$	Quadratic (7)	0.2875	0.4924	0.6384	3.0
Uniform(Jeffreys), $\alpha = 0.05$	GLINEXP (8)	0.2684	0.4648	0.6085	3.2
From $P_k m$, $P_k \sigma$ (bootstrap)	Quadratic (7)	0.2481	0.4346	0.5749	3.5
From $P_k m$, $P_k \sigma$ (bootstrap)	GLINEXP (8)	0.2758	0.4755	0.6201	3.1
From $P_k X$ (representation)	Quadratic (7)	0.2835	0.4866	0.6321	3.0
From $P_k X$ (representation)	GLINEXP (8)	0.2835	0.4866	0.6321	3.0

Here, the Jeffreys' Prior is uniform. In addition, no hierarchical approach can be applied, since we have a single parameter (the mean m).

4 Using the Hilbert Approach to Generate a Likelihood

As previously mentioned, HA can be used to generate a model $f(x, \theta)$, and a Likelihood $L(\mathcal{X}, \boldsymbol{\theta})$. To exemplify the approach, let us consider again the data (11) and let us generate a model for the situation where $\boldsymbol{\theta} = m$.

We start by reconsidering the expansion $P_k X$: in the preceding situations, it was determined by finding a Least Squares solution for (10), id est, by looking for the solution of

$$\boldsymbol{x}^* = \operatorname{argmin}\left\{\|\boldsymbol{Mx} - \boldsymbol{N}\| : \boldsymbol{x} \in \mathbb{R}^{k+1}\right\}. \tag{17}$$

However, this problem does not take into account the mean m, which is part of the model. A simple way to integrate m in the procedure consists in looking for a constrained minimum:

$$\boldsymbol{x}^* = \operatorname{argmin}\left\{\|\boldsymbol{Mx} - \boldsymbol{N}\| : \boldsymbol{x} \in \mathbb{R}^{k+1}, \overline{\boldsymbol{Mx}} = m\right\}, \overline{\boldsymbol{Mx}} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=0}^{k} M_{ij}x_j. \tag{18}$$

The problem (18) can be easily solved by introducing a Lagrange's Multiplier λ, associated to the restriction. The vector $\widetilde{\boldsymbol{x}}= (\boldsymbol{x}; \lambda)$ satisfies a linear system

$$\widetilde{\boldsymbol{M}}\widetilde{\boldsymbol{x}} = \widetilde{\boldsymbol{N}}, \tag{19}$$

analogous to (10). Once the system was solved, we can use the expansion $P_k X$ to generate the PDF of X, and use it to determine the Likelihood $L(\mathcal{X}, \boldsymbol{\theta}) = f(\mathcal{X}|\boldsymbol{\theta})$.

For instance, let us consider $k \in \{3, 5\}$ and a random variable $U \sim N(0, 1)$, $nb = 250$. Examples of results of the Likelihood generated appear in Fig. 4.

A first use for these Likelihoods is the determination of an MLE: we can determine the value $\widehat{m}$ corresponding to their maximum. Examples of results are shown in Tables 9 and 10.

A second use consists in Bayes Updating: we can generate a posterior distribution for m using (4). The prior can be chosen by the user or generated by HA, as in the preceding section. Examples of results, generated using the approaches by bootstrap and representation of$X$, are shown in Tables 9 and 10. The probabilities $P(X \leq x)$ are obtained by numerical integration of the model$f(x, \widehat{m})$:$P(X \leq x) = \int_{-\infty}^{x} f\left(x, \widehat{m}\right)dx$. In practice, the lower bound of integration is taken as the minimal value of X.

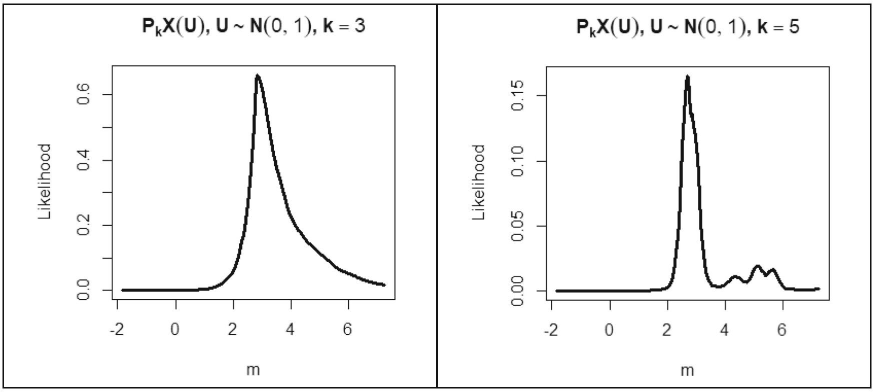


Fig. 4. Examples of Likelihood $L(\mathcal{X}, m)$ generated by HA. At left, $k = 3$; at right $k = 5$.

Table 9. Examples of estimation results using $U \sim N(0, 1)$ and $k = 3$

Prior	Loss Function	$P(X \leq 1)$	$P(X \leq 2)$	$P(X \leq 3)$	m
Exact		0.2835	0.4866	0.6321	3
MLE from the Likelihood generated by HA		0.2732	0.4516	0.5639	3.2
From $P_k m, P_k \sigma$ (bootstrap)	Quadratic (7)	0.3544	0.4836	0.5951	2.6
From $P_k m, P_k \sigma$ (bootstrap)	GLINEXP (8)	0.3033	0.4485	0.5648	2.9
From $P_k X$ (representation)	Quadratic (7)	0.3125	0.4542	0.5699	2.9
From $P_k X$ (representation)	GLINEXP (8)	0.3125	0.4542	0.5699	2.9

Table 10. Examples of estimation results using $U \sim N(0, 1)$ and $k = 5$

Prior	Loss Function	$P(X \leq 1)$	$P(X \leq 2)$	$P(X \leq 3)$	m
Exact		0.2835	0.4866	0.6321	3
MLE from the Likelihood generated by HA		0.2861	0.4666	0.5555	2.9
From $P_k m, P_k \sigma$ (bootstrap)	Quadratic (7)	0.3877	0.4947	0.5815	2.6
From $P_k m, P_k \sigma$ (bootstrap)	GLINEXP (8)	0.2877	0.4969	0.5561	2.9
From $P_k X$ (representation)	Quadratic (7)	0.2894	0.4756	0.5629	3.0
From $P_k X$ (representation)	GLINEXP (8)	0.2894	0.4756	0.5629	3.0

5 Concluding Remarks

We considered the use of the Hilbert Approach (HA) in Uncertainty Quantification (UQ) as an auxiliary method to the Bayesian Inference (BI). The Hilbert Approach was used for the generation of priors and Likelihoods. A first example was the use in De Finetti's representation theorem with a small number of data points. In the examples considered, HA furnished convenient priors and led to good results, even when a distribution different from the real one was used for the HA approximation. A second use considered the estimation of the parameters of a distribution. In a first application, HA was used to generate a model for the distribution of the parameters, using samples generated by two different techniques – bootstrap and representation of the variable itself. In both the cases, the distributions generated appeared to be convenient and led to good results. The last application was the use of HA to generate Likelihoods. In this situation, the performance was inferior for approximations of degree 3, but was more interesting for approximation involving a degree 5.

The experiments and results seem to indicate that the collaboration between HA and BI can be an interesting tool in estimation. Since the number of examples was limited up to this date, more evidence is requested for definitive conclusions. In addition, a single Hilbert basis was considered – the polynomial one, so that other basis must be investigated. These developments will be matter of future work.

References

1. Fisher, R.A.: On an absolute criterion for fitting frequency curves. Messenger of Mathematics, vol. 41, pp. 155–160 (1912). Republished at Statistical Science, vol. 12, no. 1, pp. 39–41 (1997)
2. Fisher, R.A.: On the "probable error" of a coefficient of correlation deduced from a small sample. Metron **1**, 3–32 (1921)
3. Fisher, R.A.: On the mathematical foundations of theoretical statistics. Philos. Trans. R. Soc. Lond. Ser. A, Contain. Pap. Math. Phys. Charact. **222**, 309–368 (1922)
4. Fisher, R.A.: The goodness of fit of regression formulae, and the distribution of regression coefficients. J. R. Statist. Soc. **85**(4), 597–612 (1922)
5. Pearson, K.: The fundamental problem of practical statistics. Biometrika **13**(1), 1–16 (1920)
6. de Laplace, PierreSimon : Mémoire sur la probabilité des Causes par les Evénements. Memoires de mathematique et de physique presentes a l'Academie royale des sciences, par divers savans, & lus dans ses assemblees, 6 621–656 (1774)
7. Condorcet (de Caritat, Jean-Antoine-Nicolas). Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix. Imprimerie Royale, Paris (1785)
8. Zhang, Y.-Y.: The Bayesian posterior estimators under six loss functions for unrestricted and restricted parameter spaces. In: Tang, N., (ed.) Bayesian Inference on Complicated Data, Chapter 7, pp. 89-106 (2020)
9. Chang, Y.C., Hung, W.L.: LINEX loss functions with applications to determining the optimum process parameters. Qual. Quant. **41**, 291–301 (2007)
10. Nassar, M., Alotaibi, R., Okasha, H., Wang, L.: Bayesian estimation using expected LINEX loss function: a novel approach with applications. Mathematics **10**(3), 436 (2022). https://doi.org/10.3390/math10030436
11. De Finetti, B.: Funzione Caratteristica di un Fenomeno Aleatorio, Memoria della Reale Accademia dei Lincei, vol. IV, fasc. 5, pp. 86-133 (1930)
12. De Finetti, B.: La prévision : ses lois logiques, ses sources subjectives. Annales de l'Institut Henri Poincaré, tome **7**(1), 1–68 (1937)
13. Bernardo, J.M., Smith, A.F.M.: Bayesian Theory. Wiley, Chichester (2000)
14. Hewitt, E., Savage, L.: Symmetric measures on Cartesian products. Trans. Am. Math. Soc. **80**, 470–501 (1955)
15. Souza de Cursi, E.: Uncertainty Quantification with R. Springer Verlag, Berlin (2023)
16. Souza de Cursi, E.: Uncertainty Quantification with Excel. Springer Verlag, Berlin (2022)
17. Sampaio, R., Souza de Cursi, E.: Uncertainty Quantification with Matlab. JISTE/Elsevier (2015)
18. Box, G., Tiao, G.: Bayesian Inference in Statistical Analysis. Wiley, Hoboken (1992)

A Data-Based Estimation of Power-Law Coefficients for Rainfall via Levenberg-Marquardt Algorithm: Results from the West African Testbed

Rubem V. Pacelli[1](✉), Nícolas de A. Moreira[1], Tarcisio F. Maciel[1],
Modeste Kacou[2], and Marielle Gosset[3]

[1] Federal University of Ceará, Fortaleza, Brazil
rubem070@alu.ufc.br
[2] University Félix Houphouët-Boigny, Abidjan, Côte d'Ivoire
[3] Institut de Recherche pour le Developpement, Marseille, France

Abstract. Rainfall monitoring is of paramount importance for many applications, such as meteorology, hydrology, and flood prevention. In order to circumvent the expensive deployment of weather radar networks, many articles propose rainfall estimation by using the received signal power of microwave links as they are sensitive to precipitation at certain frequencies. In this context, the International Telecommunication Union (ITU) provides a power-law relationship that allows the computation of the precipitation rate from the attenuation due to rainfall. This physics-based approach uses scattering calculation to derive the power-law coefficients, which depend only on the frequency. However, the practical use of this equation faces other important parameters, such as the link length and the distance from the bucket gauge to the microwave link. These factors may significantly affect the prediction. In this article, it is proposed a data-based alternative for the estimation of the power-law coefficients, where the Levenberg-Marquardt algorithm is used to adjust them using several data collected from different radio links in West Africa. The estimation quality is assessed in terms of its correlation with rain rate measurements from bucket gauges spread across the African testbed.

Keywords: rainfall estimation · signal attenuation · Levenberg-Marquardt algorithm

1 Introduction

In recent years, the world has been experiencing increasing scenarios of droughts and floods that threaten many countries economically and in many other different ways. Such extreme events might be a consequence of climate change that tends to worsen over time. An early-warning information system is a key part of counterbalancing this endangerment. For this purpose, rainfall monitoring plays a crucial role. The precipitation rate must be precisely measured with high spatial and temporal resolution in order to monitor an eminent extreme event.

J. E. S. De Cursi (Ed.): UQSM 2023, LNME, pp. 190–201, 2024.
https://doi.org/10.1007/978-3-031-47036-3_17

For developed countries, weather radar networks are a well-established solution for rainfall monitoring. Another usual approach is to use weather satellites to estimate the precipitation rate. Both solutions are capable of covering a broad area, which is particularly beneficial for large countries.

Unfortunately, having a dense deployment of weather radars is costly, and developing countries cannot afford it. Regarding weather satellite monitoring, some efforts have been made with the Tropical Rainfall Measuring Mission (TRMM), which consists of a constellation of satellites to monitor and study tropical and subtropical precipitation. This mission was a result of collaborative work between the United States and Japan. The TRMM has been an important data source for meteorological and hydrological activities worldwide [6]. Naturally, some emerging countries in such regions took advantage of the rainfall measurement. Notwithstanding, it is well-known that satellite rainfall monitoring still lacks accuracy, especially for high-resolution and real-time applications [5], and ground measurements to tweak the rainfall estimation are still required to adjust or to downscale satellite estimates.

Some papers have proposed alternatives to rainfall estimation, such as using satellite radio links already in operation and broadly spread worldwide. These satellite communication links usually operate on Ka or Ku bands, which are mainly corrupted by rainfall. Even though these base stations primarily focus on satellite services, it is possible to estimate the precipitation rate from rain-induced attenuation of the received signal [1,4,9].

In-situ rain gauge measurement is also a very cost-attractive and high-accuracy solution to monitor the rainfall [4]. However, this technique only provides a point-scale measurement, and a high density of gauges would be necessary to cover an urban area [8]. Another common usage is to compare the rain gauge measurement with the estimated values of a given method. In this case, the gauge is not a monitoring system but a reference set to assess the method's accuracy.

Similar techniques can also be applied to ground-to-ground microwave links [2,7,10,15]. These terrestrial radio networks have the advantage of providing measurements of close-to-ground links, which is beneficial for near-surface rainfall estimation. In scenarios where weather radars are not available operationally, commercial microwave links (CML) might be an alternative for measuring rainfall [7,10,15]. Since the International Telecommunication Union (ITU) provides a straightforward relationship between the attenuation due to rain and precipitation rate, it is possible to remove the path attenuation along the link (baseline level) and calculate the precipitation rate from the remaining attenuation, which is assumed to be caused by the rainfall. This physics-based approach utilizes scattering calculation to derive the power-law coefficients, which depend only on the frequency [11]. However, real applications also have other parameters that might influence the quality of the estimation, such as radio link length, distance from the precipitation area taken into consideration, etc.

This paper proposes an alternative data-driven estimation of the power-law coefficients, where the Levenberg-Marquardt algorithm is used to recursively adjust them in order to minimize the sum of the square of the error. The predicted rainfall time series of the proposed model is compared with a reference rain gauge (closest one), and the Pearson correlation between both curves is computed and used as figure of merit for precision. Moreover, the results are also compared with the time series predicted by using the standard ITU coefficients.

The main contributions of this article are:

1. Data treatment of the bucket gauge measurements and its mathematical analysis to estimate the rainfall via Levenberg-Marquardt algorithm.
2. Comparison of the estimated time series with the prediction from the ITU model.
3. Numerical analysis of a real dataset collected in Niamey, the capital of Niger.

The present article is organized as follows: In Sect. 2, the geographical context is detailed. Section 3 introduces the Levenberg-Marquardt algorithm and the data treatment for the present application. In Sect. 4 and 5, the results and conclusions are exposed, respectively.

2 Geographical Context

The present article makes use of a dataset collected in Niamey, the capital of Niger. The environment is characterized as semi-arid, with a rainfall rate between 500 mm/yr and 750 mm/yr, where most of this precipitation occurs between June and September. There is practically no rain in the remaining months (from October to April). Convective rains created by Mesoscale Convective Systems (MCS) comprise for $75\% - 80\%$ of the total rainfall [14].

The commercial microwave link data was originally obtained through a partnership established with the mobile telecommunication operator Orange (the network has now been bought by Zamani Com). The project entitled "Rain Cell Africa - Niger" is financed by the World Bank's Global Facility for Disaster Reduction and Recovery (WB/GFDRR) and aims to test the potential of CML-based rainfall estimation for urban hydrology in Africa. Indeed, previous results for a single radio link in Ouagadougou have indicated the feasibility of such an approach [2].

In order to cover this area, an instrumental setup that records the received power level and rainfall gauge measurement was built in Niamey. In 2017, this system continuously recorded the testbed for 6 months, approximately, and yielded a dataset from 135 microwave links and three bucket gauges. The microwave frequencies varies from 18 GHz up to 23 GHz, with a link length between 0.5 km and 5.5 km. Their measurements were recorded at a period of 15 min between the samples and with a resolution of 1 dB. The bucket gauges, on the other hand, recorded the precipitation in mm h^{-1} with a resolution of 0.5 mm but at the same rate. It is assumed no timing synchronism impairment between the samples from the CML and the bucket gauge.

The criterion used to associate the kth radio link to the ith tipping bucket gauge is the selection of the gauge whose distance from the center point of the communication link to it is minimum. This distance varies from 1 km to 6 km, approximately. Table 1 shows the number of radio links associated with each tipping gauge. The tipping gauge of number 1 has much more radio links associated with it, which naturally leads to more data availability.

Table 1. Distribution of radio links by frequency for each gauge: each row has the number of links that are associated with a given gauge.

Gauge number	18 GHz	19 GHz	22 GHz	23 GHz
1	4	4	44	45
2	2	1	10	8
3	3	3	3	8

3 Mathematical Analisys and Data Treatment

3.1 The Levenberg-Marquardt Algorithm

Let $y_i(n) \in \mathbb{R}$ be the nth measurement sample collected by the ith rain gauge (in mm/h) during one day. The signal $y_i(n)$ has no missing or zero-valued data, i.e., only the time series intervals with rainfall measurements are considered.

Let us further define $x_k(n) \in \mathbb{R}$ as the specific attenuation (in dB/km) attributable to rain along the kth radio link for the same day. Each signal $x_k(n) \in \mathbb{R}$ is associated to one and only one gauge $y_i(n)$. Our goal is to predict $y_i(n)$ by using the set $\{x_k(n) \mid k \in \mathcal{S}_i\}$, where $\mathcal{S}_i$ is the set of radio links associated to the gauge i.

A power-law relationship converts the specific attenuation into rain rate by using the following formula [15]:

$$\hat{y}_i(n) = \sqrt[w_{k,1}]{\frac{x_k(n)}{w_{k,0}}}, \tag{1}$$

where $\mathbf{w}_k = \left[w_{k,0}\ w_{k,1}\right]^\top$ and $(\cdot)^\top$ is the transpose operator.

The usual method is to apply Mie's solution to Maxwell's equations in order to evaluate the attenuation-rainfall relationship and define $\mathbf{w}_k$. This approach requires defining the environment temperature and the radio link operating frequency.

Another alternative is to treat (1) as an optimization problem, where each link-gauge pair has its own set of coefficients that minimizes a given objective function. For linear problems, the ordinary least-squares algorithm provides an analytical solution that minimizes the squared Euclidean distance of the error vector between the data and the curve-fit function. However, the power-law relationship is clearly nonlinear in its parameters, $\mathbf{w}_k$, and thus an iterative process shall be used in order to find the optimum solution. By treating the present problem as a nonlinear least-squares regression, the objective function can be defined as

$$\begin{aligned}\mathcal{E}(\mathbf{w}_k(n)) &= \mathbf{e}_i^\top(n)\mathbf{e}_i(n)\\ &= \mathbf{y}_i^\top(n)\mathbf{y}_i(n) - 2\mathbf{y}_i^\top(n)\hat{\mathbf{y}}_i(n) + \hat{\mathbf{y}}_i^\top(n)\hat{\mathbf{y}}_i(n),\end{aligned} \tag{2}$$

where $\mathbf{e}_i(n) = \mathbf{y}_i(n) - \hat{\mathbf{y}}_i(n)$ is the error vector, with $\mathbf{y}_i(n) = [y_i(1) y_i(2) \cdots y_i(n)]^\top$ and $\hat{\mathbf{y}}_i(n) = \left[\hat{y}_i(1)\, \hat{y}_i(2) \cdots \hat{y}_i(n)\right]^\top$. For each sample n, the optimization algorithm acts recursively on the parameter vector $\mathbf{w}_k(n)$ in order to minimize the cost function, $\mathcal{E}(\cdot)$.

The first-order Taylor series approximation of the function $\mathcal{E}(\cdot)$ at the instant $n+1$ is given by

$$\Delta\mathcal{E}(\mathbf{w}_k(n)) \simeq \boldsymbol{\Delta}\mathbf{w}_k^\top(n)\mathbf{g}(n), \tag{3}$$

where $\Delta\mathcal{E}(\mathbf{w}_k(n)) = \mathcal{E}(\mathbf{w}_k(n{+}1)) - \mathcal{E}(\mathbf{w}_k(n))$ is the difference of the cost function between the instants $n+1$ and n, $\boldsymbol{\Delta}\mathbf{w}_k(n) = \mathbf{w}_k(n+1) - \mathbf{w}_k(n)$ is the update vector to be calculated by the optimization method, and

$$\mathbf{g}(n) = \frac{\partial\mathcal{E}(\mathbf{w}_k(n))}{\partial\mathbf{w}_k(n)} \tag{4}$$

is the gradient vector.

By replacing (2) into (4), we have that [3]

$$\begin{aligned}\mathbf{g}(n) &= -2\frac{\partial\hat{\mathbf{y}}_i(n)}{\partial\mathbf{w}_k(n)}\mathbf{y}(n) + 2\frac{\partial\hat{\mathbf{y}}_i(n)}{\partial\mathbf{w}_k(n)}\hat{\mathbf{y}}_i(n) \\ &= -2\mathbf{J}^\top(n)\mathbf{e}_i(n),\end{aligned} \tag{5}$$

where

$$\mathbf{J}(n) = \frac{\partial\hat{\mathbf{y}}_i^\top(n)}{\partial\mathbf{w}_k(n)} \tag{6}$$

is the Jacobian matrix in denominator layout. The steepest descent algorithm updates the parameters vector in the opposite direction of the gradient vector. In other words, the vector

$$\boldsymbol{\Delta}\mathbf{w}_k(n) = \gamma\mathbf{J}^\top(n)\mathbf{e}_i(n) \tag{7}$$

points at the tangent line in which the downhill direction of $\mathcal{E}(\cdot)$ is maximum on the operating point $\mathbf{w}_k(n)$. In this equation, the value $\gamma \in \mathbb{R}$ is a step-learning hyperparameter that regulates the convergence speed [13]. Although the gradient descent method has the advantage of simplicity, such an estimator has only first-order local information about the error surface in its neighborhood. In order to increase the performance of the estimator, one can employ another algorithm, called Newton's method, which considers the quadratic approximation of the Taylor series, i.e.,

$$\Delta\mathcal{E}(\mathbf{w}_k(n)) \simeq \boldsymbol{\Delta}\mathbf{w}_k^\top(n)\mathbf{g}(n) + \frac{1}{2}\boldsymbol{\Delta}\mathbf{w}_k^\top(n)\mathbf{H}(n)\boldsymbol{\Delta}\mathbf{w}_k(n), \tag{8}$$

where

$$\mathbf{H}(n) = \frac{\partial^2\mathcal{E}(\mathbf{w}_k(n))}{\partial\mathbf{w}_k(n)\partial\mathbf{w}_k^\top(n)} \tag{9}$$

is the Hessian matrix at the instant n.

By differentiating (8) with respect to $\mathbf{w}_k(n)$ and setting its value to zero, we get the update vector that minimizes $\mathcal{E}(\cdot)$ quadratically, which is given by

$$\begin{aligned}\mathbf{g}(n) + \mathbf{H}(n)\mathbf{\Delta w}_k(n) &= \mathbf{0}\\ \mathbf{\Delta w}_k(n) &= -\mathbf{H}^{-1}(n)\mathbf{g}(n),\end{aligned}\tag{10}$$

where $\mathbf{0}$ is the zero vector. Replacing (5) into (10), ignoring the constant factor, and inserting the step-learning, yields

$$\mathbf{\Delta w}_k(n) = \gamma \mathbf{H}^{-1}(n)\mathbf{J}^{\top}(n)\mathbf{e}_i(n).\tag{11}$$

whereas the steepest descent seeks the tangent line in the most downhill direction, Newton's algorithm finds the tangent parabola that minimizes the cost function, which prompts to a faster convergence to the optimum value when compared to gradient-based methods. However, the biggest drawback is the computation of $\mathbf{H}^{-1}(n)$, which is usually costly.

One way out is to resort to quasi-Newton methods, where the inverse of the Hessian matrix is updated recursively and updated by low-rank matrices, without requiring inversion. Another solution is to obtain a nonrecursive approximation of $\mathbf{H}(n)$. For the optimization problems where the objective function is the sum of the squares of the error, the Gauss-Newton method can be used to accomplish such task. The biggest advantage is that the Hessian is approximated by a Gramian matrix that takes only first-order derivatives, in addition to being symmetric and positive definite, which consequently makes it invertible.

The base idea of the Gauss-Newton method is to linearize the dependence of $\hat{\mathbf{y}}_i(n)$ on a local operating point $\mathbf{w}$, i.e.,

$$\begin{aligned}\hat{\mathbf{y}}_i(n)|_{\mathbf{w}_k(n)+\mathbf{w}} &\simeq \hat{\mathbf{y}}_i(n) + \frac{\partial \hat{\mathbf{y}}_i^{\top}(n)}{\partial \mathbf{w}_k(n)}\mathbf{w}\\ &\simeq \hat{\mathbf{y}}_i(n) + \mathbf{J}(n)\mathbf{w},\end{aligned}\tag{12}$$

where $\hat{\mathbf{y}}_i(n)|_{\mathbf{w}_k(n)+\mathbf{w}}$ is the value of $\hat{\mathbf{y}}_i(n)$ when the coefficient vector is $\mathbf{w}_k(n) + \mathbf{w}$. By replacing (12) into (2), it follows that

$$\begin{aligned}\mathcal{E}(\mathbf{w}_k(n) + \mathbf{w}) =& \mathbf{y}_i^{\top}(n)\mathbf{y}_i(n) - 2\mathbf{y}_i^{\top}(n)(\hat{\mathbf{y}}_i(n) + \mathbf{J}(n)\mathbf{w})\\ &+ (\hat{\mathbf{y}}_i(n) + \mathbf{J}(n)\mathbf{w})^{\top}(\hat{\mathbf{y}}_i(n) + \mathbf{J}(n)\mathbf{w})\\ =& \mathbf{y}_i^{\top}(n)\mathbf{y}_i(n) + \hat{\mathbf{y}}_i^{\top}(n)\hat{\mathbf{y}}_i(n)\\ &- 2(\mathbf{y}_i(n) - \hat{\mathbf{y}}_i(n))^{\top}\mathbf{J}(n)\mathbf{w}\\ &- 2\mathbf{y}_i^{\top}(n)\hat{\mathbf{y}}_i(n) + \mathbf{w}^{\top}\mathbf{J}^{\top}(n)\mathbf{J}(n)\mathbf{w}\end{aligned}\tag{13}$$

Thus, differentiating (13) with respect to $\mathbf{w}$, and setting the result to zero, we obtain $\mathbf{w} = \mathbf{\Delta w}_k(n)$, i.e.,

$$-\mathbf{J}^{\top}(n)\mathbf{e}_i(n) + \mathbf{J}^{\top}(n)\mathbf{J}(n)\mathbf{\Delta w}_k(n) = \mathbf{0}.\tag{14}$$

Reorganizing the previous equation and inserting the step-learning, we have that

$$\mathbf{\Delta w}_k(n) = \gamma(\mathbf{J}^\top(n)\mathbf{J}(n))^{-1}\mathbf{J}^\top(n)\mathbf{e}_i(n). \tag{15}$$

By comparing (15) with (11), we notice that the Gauss-Newton method approximates the Hessian matrix, $\mathbf{H}(n)$, to $2\mathbf{J}^\top(n)\mathbf{J}(n)$ (the constant factor was dropped out), thus avoiding second-order derivatives.

The Levenberg-Marquardt (LM) method combines the two algorithms presented in this article: the steepest descent and a Newton-like algorithm. It tries to take advantage of the convergence guaranteed[1] by the gradient method, and the fast convergence of Newton's method. The update vector of the LM algorithm is given by

$$\mathbf{\Delta w}_k(n) = \gamma\left(\mathbf{J}^\top(n)\mathbf{J}(n) + \lambda(n)\mathbf{I}\right)^{-1}\mathbf{J}^\top(n)\mathbf{e}_i(n), \tag{16}$$

where $\mathbf{I}$ is the identity matrix. The goal of the hyperparameter $\lambda(n)$ is twofold: it performs the Tikhonov regularization, preventing $\mathbf{J}^\top(n)\mathbf{J}(n)$ from being ill-conditioned, and also controls how the algorithm behaves. The LM method leans toward Gauss-Newton for small values of $\lambda(n)$, while large values make it behave as the gradient descent. The initial coefficient vector, $\mathbf{w}_k(1)$, is likely far from the optimal point since it is randomly initialized. Hence, it is sensible that the LM algorithm initially behaves like the gradient descent once $\mathbf{J}^\top(1)\mathbf{J}(1)$ is probably a bad estimate (the Hessian matrix depends on the operating point of the coefficient vector when the cost function is nonquadratic). Insofar as the estimate of $\mathbf{H}(n)$ becomes trustworthy, the LM algorithm shall decrease $\lambda(n)$ toward zero, causing it to behave like the Gauss-Newton.

Finally, the coefficient vector adopted to estimate the rainfall is defined as

$$\mathbf{w}_k \triangleq \mathbf{w}_k(N+1) = \mathbf{w}_k(N) + \mathbf{\Delta w}_k(N), \tag{17}$$

where N is the number of samples in the training set.

3.2 Data Treatment and Analysis

The first step in the data treatment is to select days with rainfall events from the collected time series. In other words, the received power level, $\tilde{x}_k(m)$, and rainfall gauge measurement, $\tilde{y}_i(m)$, are decimated, producing $\tilde{x}_k(n)$ and $y_i(n)$, respectively. From the 6-month dataset, only 11 days with rainfall events are considered. A rainfall event is defined as a period in which the bucket gauge continuously measures nonzero values for, at minimum, 3 h and 45 min. Since the recording period is only 15 min, it leads to a dataset containing at minimum 15 samples for each rainfall event.

[1] Provided that γ is properly chosen.

Afterwards, the sampled received power level is converted to the specific attenuation, $x_k(n)$, by subtracting the baseline level from $\tilde{x}_k(n)$ and dividing it by the distance of the kth radio link. The baseline is determined by using the method recommended by Schleiss and Berne [12], which uses the moving window method to determine the variance.

In order to obtain a reliable analysis with the available dataset, it is performed a cross-validation where each fold comprises the samples obtained from a given rainfall event. The parameters were estimated using the training dataset, whereas the test dataset is used to assess the model performance. For this paper, it is adopted a ratio of $63\% - 37\%$ for the training and test datasets, respectively. It leads to 7 and 4 days for the training and test datasets, respectively. Algorithm 1 summarizes the procedure used in this work to process the data, estimate the parameters, and analyze the results. In this algorithm, the symbol ρ indicates the Pearson correlation coefficient between $y_i(n)$ and $\hat{y}_i(n)$, which is used as figure of merit.

Algorithm 1: Data processing for the kth radio link

```
Input: $x_k(m)$
Output: Pearson correlation of all folds
foreach Fold do                                        // Cross-validation
    forall Training set do
        $\tilde{x}_k(n) \leftarrow$ Decimate $\tilde{x}_k(m)$
        $y_i(n) \leftarrow$ Decimate $\tilde{y}_i(m)$
        $x_k(n) \leftarrow$ Get the specific attenuation from $\tilde{x}_k(n)$
        $\mathbf{\Delta w}_k(n) \leftarrow$ Equation (16)
        $\mathbf{w}_k(n+1) \leftarrow \mathbf{\Delta w}_k(n) + \mathbf{w}_k(n)$
    forall Test set do
        $\hat{y}_i(n) \leftarrow$ Equation (1)
    $\rho \leftarrow$ Compute the Pearson correlation between $y_i(n)$ and $\hat{y}_i(n)$
```

4 Results and Discussions

Considering that the system provides the minimum, mean, and maximum attenuation reached by each radio link, Fig. 1 shows the box plot obtained by the model for each situation. A kernel smoothing technique is applied to the set of Pearson correlations in order to estimate its distribution, which is also shown in this figure. For the sake of performance comparison, the results obtained in this work are contrasted to the performance obtained using the original ITU coefficients, under the same methodology of test datasets and cross-validation.

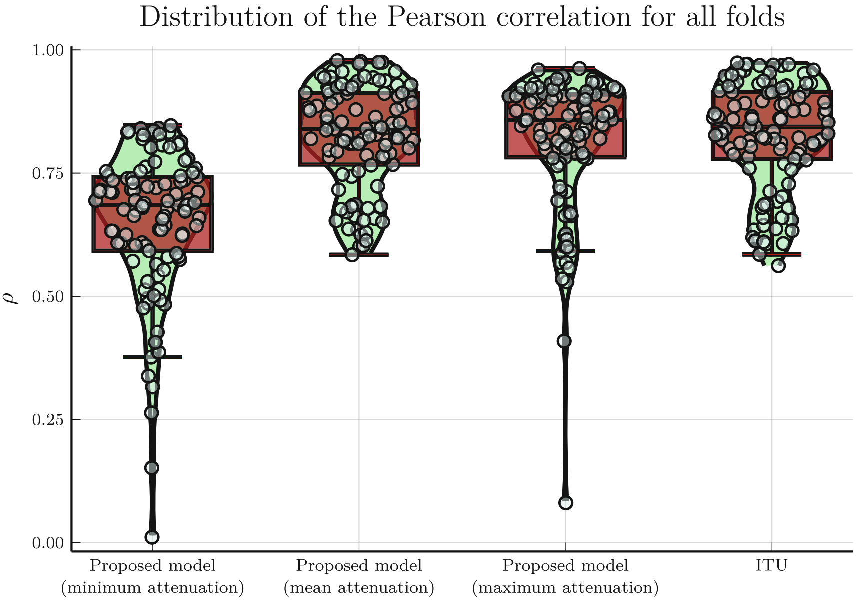
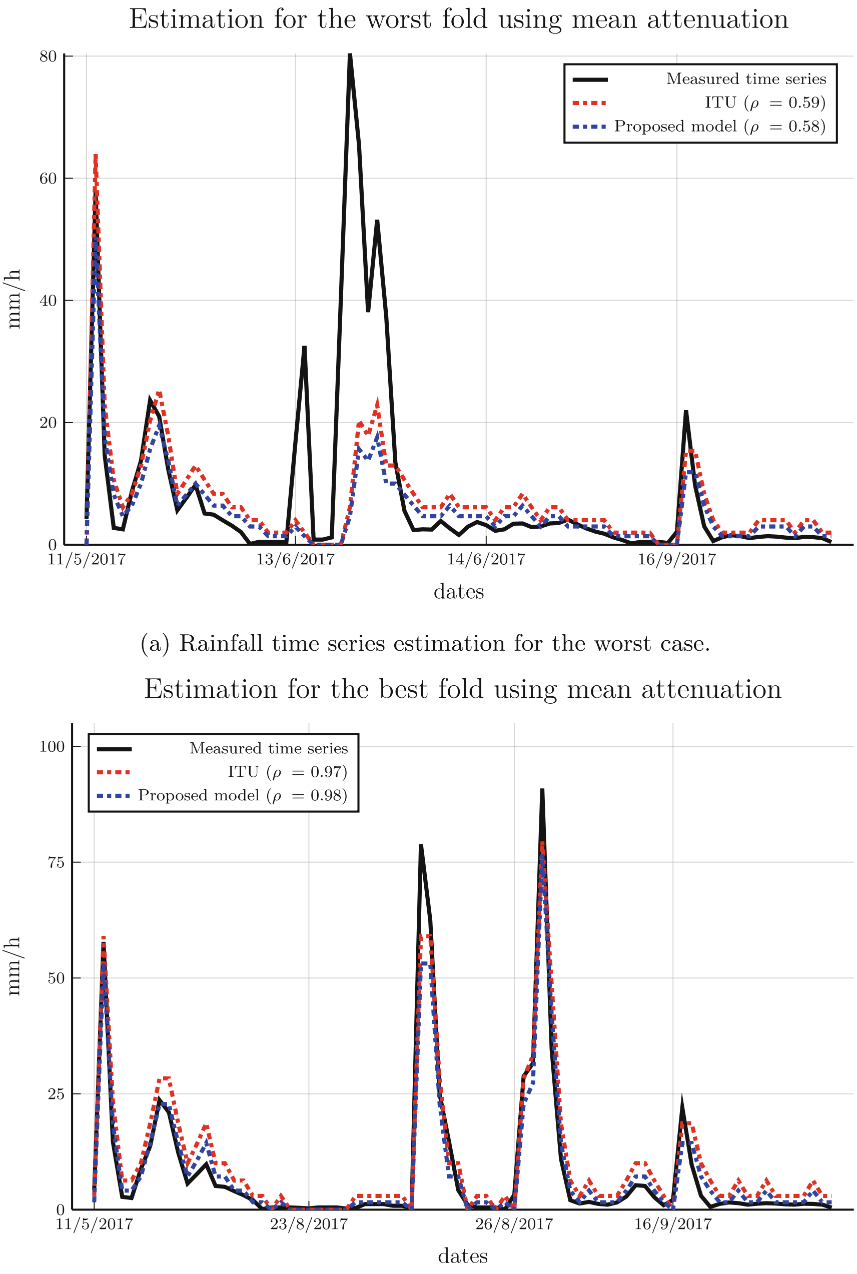


Fig. 1. Box plot of the rainfall estimation.

It is possible to notice that the proposed model presents a considerable variance with outliers when it is used the maximum or minimum attenuation. However, for the mean attenuation, the mean correlation is 82.45%, without considerable loss of performance for all folds. This result is slightly lower than the mean correlation obtained by the physics-based method. However, its performance in terms of variance, 1.10×10^{-2}, surpasses the results obtained when using the ITU model, which obtained 1.16×10^{-2}.

Figure 2 shows the time series estimation for the best and worst folding. Both estimations came from the gauge number 1, which has more radio links associated with it. Nevertheless, as shown in Fig. 2a, both methods failed in estimating the measured rainfall that occurred on June 13, 2017 for a given training-test dataset split.



(a) Rainfall time series estimation for the worst case.

(b) Rainfall time series estimation for the best case.

Fig. 2. Time series estimation of the best and worst folding case.

5 Conclusions and Future Work

In this article, we presented a new methodology to estimate the coefficient parameters of the rainfall via Levenberg-Marquardt algorithm. The available data was properly preprocessed before estimating the coefficient parameters. The cross-validation technique was applied in order to obtain a reliable estimation of performance, assessed in terms of the Pearson correlation coefficient.

Additionally, the estimated performance was compared with results when the original ITU coefficients are used, under the same conditions of the test datasets. The results show that both methodologies achieved similar results, where the present estimation technique presented a lower mean and variance.

In this work, the power-law relationship provided an estimation mapping that does not take into account the time correlation between the samples. Moreover, the raw data was decimated, and only intervals with reasonable rainfall events are considered. Future efforts might consider the correlation time of the radio link attenuation, and how it can be exploited to estimate precipitation.

References

1. Barthès, L., Mallet, C.: Rainfall measurement from the opportunistic use of an earth-space link in the ku band. Atmos. Meas. Tech. **6**(8), 2181–2193 (2013)
2. Doumounia, A., Gosset, M., Cazenave, F., Kacou, M., Zougmore, F.: Rainfall monitoring based on microwave links from cellular telecommunication networks: first results from a west african test bed. Geophys. Res. Lett. **41**(16), 6016–6022 (2014)
3. Gavin, H.P.: The Levenberg-Marquardt algorithm for nonlinear least squares curve-fitting problems. Department of Civil and Environmental Engineering, Duke University, 19 (2019)
4. Gharanjik, A., Bhavani Shankar, M.R., Zimmer, F., Ottersten, B.: Centralized rainfall estimation using carrier to noise of satellite communication links. IEEE J. Sel. Areas Commun. **36**(5), 1065–1073 (2018)
5. Han, C., Huo, J., Gao, Q., Guiyang, S., Wang, H.: Rainfall monitoring based on next-generation millimeter-wave backhaul technologies in a dense urban environment. Remote Sensing **12**(6), 1045 (2020)
6. Liu, Z., Ostrenga, D., Teng, W., Kempler, S.: Tropical rainfall measuring mission (TRMM) precipitation data and services for research and applications. Bull. Am. Meteor. Soc. **93**(9), 1317–1325 (2012)
7. Messer, H., Zinevich, A., Alpert, P.: Environmental monitoring by wireless communication networks. Science **312**(11), 713 (2006)
8. Mishra, K.V., Bhavani Shankar, M.R., Ottersten, B.: Deep rainrate estimation from highly attenuated downlink signals of ground-based communications satellite terminals. In: ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 9021–9025. IEEE (2020)
9. Mugnai, C., Sermi, F., Cuccoli, F., Facheris, L.: Rainfall estimation with a commercial tool for satellite internet in ka band: Model evolution and results. In: 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 890–893. IEEE (2015)
10. Overeem, A., Leijnse, H., Uijlenhoet, R.: Measuring urban rainfall using microwave links from commercial cellular communication networks. Water Resour. Res. (47) (2011)

11. ITUR Rec. ITU-R P. 838-3. specific attenuation model for rain for use in prediction methods. International Telecommunication Union-ITU, fevereiro, pages 838–3 (2005)
12. Schleiss, M., Berne, A.: Identification of dry and rainy periods using telecommunication microwave links. IEEE Geosci. Remote Sens. Lett. **7**(3), 611–615 (2010)
13. Strang, G.: Linear Algebra and Learning from Data, vol. 4. Wellesley-Cambridge Press, Cambridge (2019)
14. Turko, M., et al.: Rainfall measurement from mobile telecommunication network and potential benefit for urban hydrology in Africa: a simulation framework for uncertainty propagation analysis. Proc. Int. Assoc. Hydrol. Sci. **383**, 237–240 (2020)
15. Zinevich, A., Alpert, P., Messer, H.: Estimation of rainfall fields using commercial microwave communication networks of variable density. Adv. Water Resour. **31**, 1470–1480 (2008)

Stochastic Kriging-Based Optimization Applied in Direct Policy Search for Decision Problems in Infrastructure Planning

Cibelle Dias de Carvalho Dantas Maia(✉) and Rafael Holdorf Lopez

Department of Civil Engineering, Federal University of Santa Catarina,
João Pio Duarte Silva, 88040-900 Florianopolis, Brazil
{maia.cibelle,rafaelholdorf}@gmail.com

Abstract. In this paper, we apply a stochastic Kriging-based optimization algorithm to solve a generic infrastructure planning problem using direct policy search (DPS) as a heuristic approach. Such algorithms are particularly effective at handling high computational cost optimization, especially the sequential Kriging optimization (SKO). SKO has been proving to be well-suited to deal with noise or uncertainty problems, whereas assumes heterogeneous simulation noise and explores both intrinsic uncertainty inherent in a stochastic simulation and extrinsic uncertainty about the unknown response surface. Additionally, this paper employs a recent stochastic Kriging method that incorporates smoothed variance estimations through a deterministic Kriging metamodel. The problem evaluated is the DPS as a heuristic approach, this is a sequential decision problem-solving method that will be applied to a generic infrastructure planning problem under uncertainty. Its performance depends on system and cost model parameters. Previous research has employed Cross Entropy (CE) as a global optimization method for DPS, while this paper utilizes SKO as a stochastic Kriging-based optimization method and compares the results with those obtained by CE. The proposed approach demonstrates promising results and has the potential to advance the field of Kriging-based algorithms to solve engineering problems under uncertainties.

Keywords: optimization problem · stochastic Kriging · sequential Kriging optimization · direct policy search · infrastructure planning

1 Introduction

Predictive response surfaces through metamodels have been the strategy approach to high computational cost optimization problem-solving. The primary motivation for using metamodels in simulation optimization is to reduce the number of expensive fitness evaluations without degrading the quality of the obtained optimal solution.

J. E. S. De Cursi (Ed.): UQSM 2023, LNME, pp. 202–213, 2024.
https://doi.org/10.1007/978-3-031-47036-3_18

The basic idea is that the metamodel acts as an interpolating or regressor curve of support points that have information from the objective function and its constraints so that the results can be predicted without resorting to the use of the primary source (objective function) [4].

One of the most popular metamodels is Kriging, which has a long and successful tradition for modeling and optimizing deterministic computer simulations [14]. The great advantage of this metamodel is that it allows the quantification of the uncertainty of the response surface through the mean square error (MSE). An extension for the application in noisy problems is Stochastic Kriging (SK) proposed by Ankenman, Nelson and Staum in [1]. Additionally, a recent stochastic Kriging approach is proposed by Kaminski in [11] with smoothed variance evaluations using determinist Kriging as a variance surrogate.

A stochastic Kriging-based optimization is a powerful approach that combines stochastic surrogate modeling and optimization robust techniques to efficiently and accurately solve complex optimization problems with uncertain or noisy data. One of the methods, known as sequential Kriging optimization (SKO), was first introduced by Huang et al. in [8] and later adapted by Jalali, Van Nieuwenhuyse and Picheny in [9] to handle heterogeneous noise. This algorithm exploits both the information provided by the stochastic Kriging metamodel and the uncertainties about the noise of the stochastic problem, choosing as alternative the maximum augmented expected improvement (AEI) as an infill criterion, to iteratively add new points, improving the surrogate accuracy while at the same time seeking its global minimum.

So, in this paper is used SKO as optimization algorithm approaching a recent stochastic Kriging framework with smoothed variance evaluations to solve the direct policy search (DPS) with a heuristic applied to a generic infrastructure planning problem.

Direct policy search is one of the solution frameworks for solving the general decision problem under uncertainty. This method has been applied to a wide range of infrastructure planning problems, including transportation [6], energy [7], water resource management [5], and risk based inspection (RBI) planning ([2,13]).

In DPS with a heuristic, the parameters of a function mapping the system state to decisions are optimized rather than the decisions themselves. What results from optimization with DPS is therefore not a sequence of decisions, but a policy that one can operate [15].

Our problem involves applying the DPS optimization technique to an infrastructure planning problem. To find the heuristic parameters that maximize the expected total life-cycle utility under uncertainties, a global optimization method is necessary. Bismut et al. [3] approaches the cross entropy (CE) method for this case. In this context, this paper approaches SKO as an alternative method.

The optimization process through SKO involves employing a deterministic Kriging surrogate to approximate the variance of the error, constructing a predictive surface that provides smoothed noise variance estimates [11]. The variance of the error is subsequently included in the SK, resulting in a good noise filtering metamodel. To guide the addition of new points in the stochastic SKO framework, the AEI infill criterion is applied, utilizing information from the smoothed variance estimates.

This article is organized as follows: Sect. 2 provides a brief explanation of the Stochastic Kriging metamodel-based optimization adopted, introducing the stochastic Kriging framework and the optimization method SKO. Section 3 presents the problem, detailing the heuristic of the direct policy search used and the generic planning problem; Sect. 4 presents the analysis and results of the problems; and lastly, conclusions are presented in Sect. 5.

2 Stochastic Kriging Metamodel Based Optimization

The surrogate model, $\hat{y}$, is only an approximation of the true stochastic function $f(\mathbf{x}, \theta)$ we want to optimize, where θ is the vector of random parameters and $\mathbf{x}$ is the vector of design variables. The use of the Kriging metamodel in optimization problems is attractive because, not only can it give good predictions of complex landscapes, but it also provides a credible estimate of the possible error in these predictions. So, in the Kriging-based optimization algorithm, the error estimates make it possible to make tradeoffs between sampling where the current prediction is good (local exploitation) and sampling where there is high uncertainty in the function predictor value (global exploration), allowing searching the decision space efficiently [10].

Kriging-based optimization algorithms start by simulating a limited set of input combinations (referred to as initial sampling) and, to increase the response surface accuracy, iteratively select new input combinations to simulate by evaluating an infill criterion (IC), which reflects information from Kriging. Those updates points are called Infill Points (IPs). The response surface is then updated sequentially with information obtained from the newly simulated IPs. The procedure is repeated until the desired performance level is reached, and the estimated optimum is returned [16].

The remainder of this section briefly explains the stochastic Kriging surrogate framework adopted and the SKO structure, detailing the augmented expected improvement (AEI) as infill criteria applied for search and the replication strategy of the IPs.

2.1 Stochastic Kriging

An extension to the deterministic Kriging methodology to deal with stochastic simulation was proposed by Ankenman, Nelson and Staum in [1]. Their main contribution was to account for the sampling variability that is inherent to a stochastic simulation, in addition to the extrinsic error that comes from the metamodel approximation. Then, the stochastic Kriging (SK) prediction can be seen as:

$$\hat{y}(\mathbf{x}) = M(\mathbf{x}) + Z(\mathbf{x}) + \epsilon(\mathbf{x}). \tag{1}$$

where $M(\mathbf{x})$ is the usual average trend, $Z(\mathbf{x})$ accounts for the model uncertainty and is now referred to as extrinsic noise. And, the additional term $\epsilon(\mathbf{x})$, represents the intrinsic noise, accounts for the simulation uncertainty or variability. The intrinsic noise has a Gaussian distribution with zero mean and is independently and identically distributed (i.i.d.) across replications. The SK prediction and variance for a given point x^+ are, respectively:

$$\widehat{y}(x^+) = \hat{\mu} + \hat{\sigma}_Z^2 \mathbf{h}^T \left[\sum{}_Z + \sum{}_\epsilon\right]^{-1} (\overline{\mathbf{y}} - \widehat{\mu}\mathbf{1}). \tag{2}$$

$$\hat{s}^2(x^+) = \hat{\sigma}_Z^2 - (\hat{\sigma}_Z^2)^2 \mathbf{h}^T \left[\sum{}_Z + \sum{}_\epsilon\right]^{-1} \mathbf{h} + \frac{\delta^T \delta}{\mathbf{1}^T \left[\sum{}_Z + \sum{}_\epsilon\right]^{-1} \mathbf{1}}. \tag{3}$$

where $\hat{\mu}$ and $\hat{\sigma}_z^2$ are the mean and variance trend of the SK metamodel, respectively, $\mathbf{h}$ is the correlation vector, $\sum_Z$ is the covariance matrix of all the support points of Z, $\sum_\epsilon$ is the covariance matrix of ϵ, $\overline{\mathbf{y}}$ is the vector of the approximate mean value of the objective function at each design point $\overline{\mathbf{y}} = 1/n_t \sum_{j=1}^{n_t} y(x, \boldsymbol{\theta}_j)$, where n_t is the sample size and, lastly, $\delta(\mathbf{x}^+) = \mathbf{1} - \mathbf{1}^T \left[\sum_Z + \sum_\epsilon\right]^{-1} \hat{\sigma}_z^2 \mathbf{h}(\mathbf{x}^+)$.

To improve the quality of the metamodel, Kaminski in [11] employs deterministic Kriging metamodel prediction to approximate the problem variance in the stochastic Kriging. So, the covariance $\sum_\epsilon$ is a diagonal matrix $diag\{\hat{V}(x_1)/n_1, \ldots, \hat{V}(x_k)/n_t\}$, where $\hat{V}$ is given by a determinist Kriging metamodel built on the sample variances s_i^2 obtained at design point x_i, $i = 1, 2, \ldots, t$, where t is the training sample size. So, the SK prediction for a variance in a given point is:

$$\widehat{V}(x^+) = \hat{\mu}_V + \hat{\sigma}_{Z_V}^2 \mathbf{h}^T \left[\sum{}_{Z_V} + \sum{}_\eta\right]^{-1} \mathbf{h}. \tag{4}$$

where $\mathbf{h}$ is the correlation vector, $\hat{\mu}_V$ and $\hat{\sigma}_{z_V}^2$ are the mean and variance trend of the variance on the Kriging metamodel, respectively, $\sum_{Z_V}$ is the covariance matrix of all the support points of variance extrinsic noise Z_V, $\sum_\eta$ is the covariance matrix of η. The covariance $\sum_\eta$ is a diagonal matrix $diag\{\hat{V}_{s^2}(x_1), \ldots, \hat{V}_{s^2}(x_k)\}$, where the estimator of variance is $s^2 = \sum_{i=1}^n (y_i - \overline{y}^2)/(n-1)$ and the variance of this estimator is approximately $\hat{V}_{s^2} = 2(s^2)/(n-1)$. For more stochastic Kriging framework information, see [1,11,17] and [18].

2.2 Sequential Kriging Optimization - SKO

An effective strategy to incorporate the advantages of the noise filtering from SK metamodel provided by Kaminski in [11] is to incorporate it into the infill criterion. The Augmented Expected Improvement (AEI), provided by Huang et al. in [8], is an extension for stochastic function evaluation of the Expected Improvement (EI) criterion presented by Jones, Schonlau, and Welch in [10] for the deterministic case. Therefore, in the formulation, uncertainties about the noise of the stochastic function and the prediction by the Kriging metamodel are taken into account. So, for algorithm optimization, we use the sequential Kriging optimization (SKO) approach by Jalali, Van Nieuwenhuyse and Picheny in [9] where it employs AEI as an infill criterion. The next infill point is given by:

$$AEI(x^{+}) = E[max(y_{min} - \hat{y}, 0)] \left(1 - \frac{\hat{\sigma}_{\epsilon}^{2}(x^{+})}{\sqrt{\hat{s}^{2}(x^{+}) + \hat{\sigma}_{\epsilon}^{2}(x^{+})}}\right) \tag{5}$$

where $\hat{y}$ is SK predictor, y_{min} is the Kriging prediction at the current effective best solution, i.e., the point with minimum among the simulated point. $\hat{s}^2$ is the variance of the metamodel expected value, Eq. 3, and $\hat{\sigma}_{\epsilon}^{2} = \hat{V}(x_i)/n_i$ is the variance of the function noise. For more framework details, see [9] and [8]. The maximum of the expected improvement $E[max(y_{min} - \hat{y}, 0)]$ is obtained as discussed by Jones, Schonlau and Welch in [10].

The SKO optimization involves three main steps: first, to select the data of independent variables in the design space using Latin Hypercube as experimental design techniques and to obtain the function values of the selected data; then, to generate the response surfaces of the objective functions; and finally, to carry out the optimization processes through AEI using information from the response surface model, the variance of the metamodel expected value and the function noise, and, finally, find the optimal results.

3 Direct Policy Search to the Infrastructure Planning Problem

The problem investigated, direct policy search (DPS), is a strategy for solving sequential decision problems that addresses heuristics as a search strategy for the global optimum within a reduced solution space. DPS is often chosen for its flexibility and intuitive principles. The optimal strategy for the decision problem is:

$$\mathcal{S}^{*} = \arg\max_{\mathbf{x} \in \mathscr{S}} (\mathbf{E}[y(\mathbf{x}, \boldsymbol{\theta})]) \tag{6}$$

where $\mathscr{S}$ is the space of all possible strategies, and $\mathbf{E}[y(\mathbf{x}, \boldsymbol{\theta})]$, in our problem, is the expected total life-cycle utility associated with a strategy $\mathbf{x}$.

3.1 Infrastructure Planning Problem

We investigate a generic infrastructure planning problem, as described Bismut and Straub in [3], which involves increasing the system's capacity in an optimized manner so that demand is met at the lowest implementation cost. Therefore, each year t, the system's capacity a_t must cover the demand θ_t, which will increase over the discrete time interval $[1, 2, 3, ..., T]$. The initial system capacity a_1 is fixed by the operator and can be increased at any time for a cost. The upgrade costs are given by:

$$U = y(\mathbf{x}, \boldsymbol{\theta}) = \sum_{t=1}^{T} U_{C,t}\left(\mathbf{x}\right) + U_{R,t}\left(\mathbf{x}, \boldsymbol{\theta}\right), \tag{7}$$

in which

$$U_{C,1}\left(\mathbf{x}\right) = -c_a a_1 \tag{8}$$

$$U_{C,t}\left(\mathbf{x}\right) = (a_t - a_{t-1}) c_a \gamma_t \tag{9}$$

$$U_{R,t}\left(\mathbf{x}, \boldsymbol{\theta}\right) = \varPhi\left(\frac{a_t - \theta_t}{\alpha}\right) c_F \gamma_{t-1} \tag{10}$$

where $c_a = 1$ is the upgrading cost factor, $c_F = 10$ is the penalty factor, $\varPhi$ is the standard normal cumulative distribution of the capacity, $\varPhi \sim N(a_t, \theta_t)$, $\alpha = 0.1$ is the tolerance and, lastly, γ_t is the discount factor $\gamma_t = \frac{1}{(1+r)^t}$ with $r = 0.02$ and θ_t is the system demand defined in Table 1.

The system incurs a penalty, $U_{R,t}$, when demand is not met within a certain margin. Thus, the expected total cost, U, of the system will be given by the portion of the upgrade cost, $U_{C,t}$, plus the cost of the excess penalty, U_{R_t}. So, the demand growth dependent on random quantities and the objective function of the SDP is its expected value:

$$f(\mathbf{x}) = E[U] = E\left[\sum_{t=1}^{T} U_{C,t}\left(\mathbf{x}\right) + U_{R,t}\left(\mathbf{x}, \boldsymbol{\theta}\right)\right], \tag{11}$$

The parameters used for the DPS are given in Table 1. Given the uncertain nature of demand, the initial demand θ_1 is modeled as a normal random variable. T is the design horizon, while Z_t denotes the noisy observation of demand at time.

Table 1. Model parameters.

Variable	Type	Mean	Std.Dev
θ_1	Normal distr	$\mu_{ini} = 1.0$	$\sigma_{ini} = 0.5$
θ_t	function	$\sigma_{t-1} + \tau$	–
τ	Normal distr	$\mu_\tau = 0.02$	$\sigma_\tau = 0.05$
Z_τ	Normal distr	θ_t	$\sigma_\epsilon = 0.1$
T	Deterministic	100 [anos]	$\sigma_\epsilon = 0.1$

Bismut and Straub (2019)

After setting the initial capacity a_1, the system is subject to the demand of the first year θ_1, where Z_1 is the noisy observation of this demand. Then, the capacity for the next year (a_2) is calculated, and so on, up to the time horizon T. The capacity can only increase over time and must be restricted to six stages, $t = 0, 1, 2, 3, 4, 5, 6$.

3.2 Heuristics Investigated

It is adopted the following heuristic, presented by Bismut and Straub in [3], to update the system's capacity finding. According to this approach, when the current observation Z_t is within a specific margin of the available capacity a_t, the system's capacity is increased by at least Δa, where the size of this margin is determined by the factor k. This infrastructure planning problem is presented in Fig. 1. To maximize the expected total life-cycle utility, the Sequential Kriging Optimization (SKO) method is used to optimize the parameters a_1, Δa, and k.

```
in heuristic parameters $\{a_1 \geq 0, \Delta a \geq 0, k\}$
time horizon T, observations of demand $Z_1, \ldots, Z_T$
out vector $a_t$ of capacities
$t \leftarrow 1$;
while $t < T$ do
    if $a_t - Z_t < k.a_t$
        $a_{th} \leftarrow k.a_t + Z_t$;
        $a_{t+1} \leftarrow min(max(a_t + \Delta a), 6)$;
    end
    $t \leftarrow t + 1$
end
return $a_t$
```

Fig. 1. Heuristic to upgrade the capacity based on the observed value of demand.

4 Comparison of Results

In this section, we investigated the performance of the sequential Kriging optimization (SKO) algorithm with smoothed noise variance applied to a generic infrastructure planning problem solved by direct policy search. Our focus is on comparing the results considering continuous design variables, which were achieved by utilizing the cross-entropy (CE) optimization method, as approached in both Bismut and Straub in [3] and Lopez, Bismut and Straub in [12]. The problem consists of:

$$\begin{aligned} \text{Find:} \quad & \mathbf{x}^* \\ \text{that minimizes:} \quad & f(\mathbf{x}) = E\left[U\right] \\ \text{subject to:} \quad & k \in [0.5, 0.25] \\ & a_0 \in [1, 3] \\ & \Delta a \in [1, 4] \end{aligned}$$

Figure 2 illustrates the convergence of the expected value of the total cost U in relation to the sample size of the noisy parameter. It can be observed that the expected value start stabilizes after approximately 1600 samples. Furthermore, we observe that the expected cost, measured at the 10th and 90th percentiles, stabilizes with larger sample sizes.

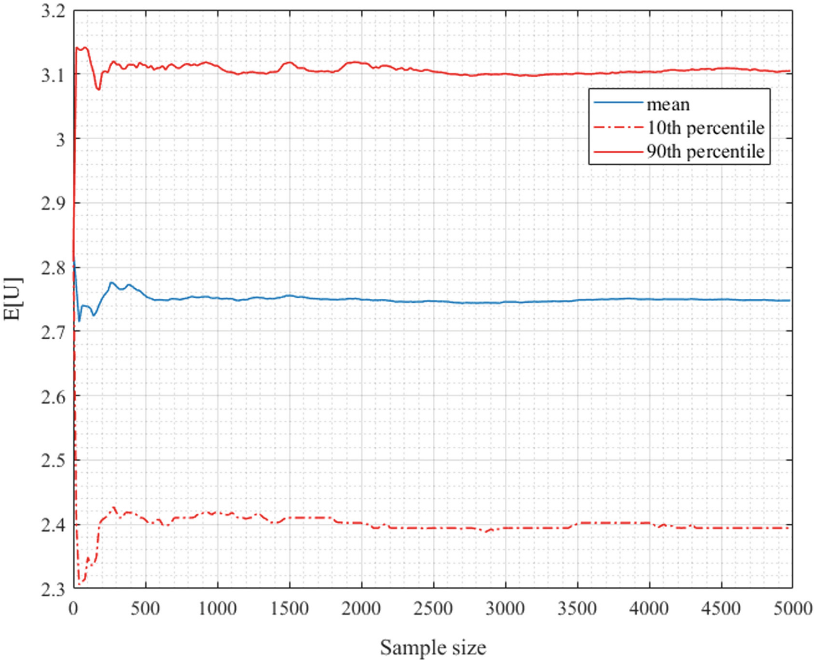


Fig. 2. Expected total cost according sample size.

There will be two approaches to the analysis of the problem. Initially, the direct policy search problem was simplified by considering a one-dimension problem with only the parameter k as a design variable. Subsequently, the problem was optimized for all three parameters: $a_0, \Delta a, k$. We will evaluate both cases by varying the sampling size (n_t) of the stochastic parameter of the DPS problem.

In the SKO optimization algorithm approach, we utilized specific parameters. These included setting n_0 to be equal to $10 \times d$, where n_0 represents the number of elements in the initial sample space of the metamodel, while d is the dimension of the problem. Moreover, to distribute these elements, we utilized the Latin Hypercube [10]. Furthermore, the stop criterion for this process was established when the maximum number of infill points equaled twenty.

4.1 One Dimension Problem

In this case, we considered a one-dimensional problem with only the parameter k as a design variable. As results obtained by Bismut and Straub in [3], we will set the others initial parameters as $a_0 = 2$ and $\Delta a = 1$. Subsequently, the performance of SKO will be evaluated by varying the sample size of the stochastic parameter (n_t). The result is presented in Table 2, with the values representing the analysis for fifty iterations of the optimization process.

Table 2. Optimum values.

Optimization algorithm	Sample size	Design variable	E[U]	E[U]
	n_t	k	Mean	Std
CE	-	0.1236	2.79	–
SKO	1000	0.1123	2.769	0.0011
SKO	500	0.1129	2.769	0.0012
SKO	100	0.1148	2.771	0.0032

According to Table 2, the CE algorithm obtained a minimum value of 2.79, while the SKO algorithm outperformed it by attaining a better minimum value of 2.771, considering a small sampling of the stochastic parameter $n_t = 100$, and presenting a small standard deviation of 0.0032 of the results. Consequently, the design variable k found by the CE and SKO algorithms were different, with values of k of 0.1236 and 0.1148, respectively. Moreover, it is worth noting that the total cost tended to stabilize after 1600 samples, as shown in Fig. 2. This analysis further supports the superior performance of the SKO algorithm, which provided a better global minimum value and maintained low variance even with small samples of the stochastic parameter.

4.2 Three Dimension Problem

For this case, we considered a three-dimensional problem with the parameters a_0, Δa and k as continuous design variable from the DPS, i.e., the generic infrastructure planning problem. Then, the resulting optimization problem is posed for different sample size for both algorithms, CE and SKO. The results obtained from these algorithms are presented in Table 3, where the mean and standard deviation of the expected total cost from the CE algorithm is available in Lopez, Bismut and Straub in [12]. The performance of SKO was analyzed over fifty iterations of the optimization process.

From Table 3, the results obtained by the SKO largely outperformed CE. The CE algorithm exhibited average minimum values of 2.7747 and 2.864 for sample sizes of 10000 and 5000, respectively. The standard deviation for the CE algorithm was higher for the sample size of 5000, indicating greater variability in the obtained results. In contrast, the SKO algorithm performed better, showing relatively low standard deviation for smaller sample sizes. Specifically, focusing on the results of the smallest sample sizes of 1000 and 500, the minimum mean values were 2.7768 and 2.7812, with standard deviations of 0.0186 and 0.0211,

Table 3. Optimum values for Case 02.

Optimization algorithm	Sample size n_t	E[U] Mean	E[U] Std
CE	10000	2.7747	0.0079
CE	5000	2.864	0.2031
SKO	5000	2.7719	0.0129
SKO	1000	2.7768	0.0186
SKO	500	2.7812	0.0211

respectively. Therefore, SKO demonstrated superior performance by finding better results for problems with much smaller sample sizes, making it an effective alternative to the problem. The parameter values obtained by SKO are presented in the Table 4.

Table 4. Optimum values from SKO.

Sample size	Design variable	Total cost
5000	$a_0 = 1.902,\ \Delta a = 1,\ k = 0.1233$	2.7719
1000	$a_0 = 1.875,\ \Delta a = 1,\ k = 0.1296$	2.7768
500	$a_0 = 1.870,\ \Delta a = 1,\ k = 0.1309$	2.775

These results demonstrate the efficiency of the SKO optimization algorithm addressing the stochastic Kriging metamodel with smoothed variance involving DPS for solving a general planning problem. The obtained results were more precise and computationally more economical, requiring a smaller number of stochastic parameter samples. As a result, this strongly supports the use of the SKO algorithm as an efficient and practical tool for solving similar optimization problems in various fields.

5 Conclusion

This paper compares the effectiveness of the stochastic Kriging-based optimization algorithm (SKO) with the results obtained using the cross-entropy method in a generic sequential decision problem for infrastructure planning. Additionally, the study approach smoothed variance estimations in both the stochastic Kriging surrogate and the infill criterion in an optimization framework, which uses the Augmented Expected Improvement (AEI). The results of the comparison indicate that SKO is a promising optimization algorithm for this type of problem.

The problem being analyzed involves the optimization of three parameters - the initial system capacity (a_0), the capacity increment (Δa), and a capacity correction factor (k) - using a heuristic as the solution strategy. Initially, the problem was approached as one-dimensional, with the design variable limited to only the parameter k. Later, the analysis was extended to include all three parameters as continuous design variables. In both cases, the SKO algorithm demonstrated superior performance by producing more accurate minimum values and showing convergence in the results even for a small sampling of the stochastic parameter.

In conclusion, the use of SKO for direct policy search with a heuristic achieved excellent results, especially due to incorporating noise information during the optimization process. In addition to obtaining more accurate results, the SKO approach reduces computational cost, since it decreases the number of problem evaluations while improving the accuracy of the noisy model. The use of Kriging-based algorithms for optimizing modeled systems through stochastic simulation, especially with heterogeneous noise, is relatively new and has great research potential.

Notes and Comments. The authors are grateful for the support by CAPES through PROAP (Graduate Support Program).

References

1. Ankenman, B., Nelson, B.L., Staum, J.: stochastic kriging for simulation metamodeling. OperationsResearch **58**(2), 371–382 (2010)
2. Bismut, E.; Straub, D. Adaptive direct policy search for inspection and maintenance planning in structural systems. In: Life Cycle Analysis and Assessment in Civil Engineering: Towards an Integrated Vision, CRC Press, pp. 3103–3110 (2018)
3. Bismut, E., Straub, D.: Direct policy search as an alternative to POMDP for sequential decision problems in infrastructure planning. In: 13th International Conference on Applications of Statistics and Probability in Civil Engineering, ICASP13 (2019)
4. Forrester, A.I., Keane, A.J.: Recent advances in surrogate-based optimization. Progress Aerospace Sci. **45**(1–3), 50–79 (2009)
5. Giuliani, M., et al.: Curses, tradeoffs, and scalable management: advancing evolutionary multiobjective direct policy search to improve water reservoir operations. J. Water Resources Planning Manage. **142**(2), 04015050 (2016)
6. Golla, A., et al.: Direct policy search for multiobjective optimization of the sizing and operation of citizen energy communities. In: HICSS, pp. 1–10 (2021)
7. Gupta, A., et al.: Exploring a direct policy search framework for multiobjective optimization of a microgrid energy management system (2020)
8. Huang, D., Allen, T.T., Notz, W.I., Zeng, N.: Global optimization of stochastic black-box systems via sequential kriging meta-models. J. Global Optim. **34**(3), 441–466 (2006)
9. Jalali, H., Nieuwenhuyse, V.: Inneke and Picheny, victor: comparison of kriging-based algorithms for simulation optimization with heterogeneous noise. Eur. J. Oper. Res. **261**(1), 279–301 (2017)
10. Donald, R.J., Matthias, S., William, J.W.: Efficient global optimization of expensive black-box functions. J. Global Optim. **13**(4), 455 (1998)

11. Kaminski, B.: A method for the updating of stochastic Kriging metamodels. Eur. J. Oper. Res. **247**(3), 859–866 (2015)
12. Lopez, R.H., Bismut, E., Straub, D.: Stochastic efficient global optimization with high noise variance and mixed design variables. J. Brazilian Soc. Mech. Sci. Eng. **45**(1), 7 (2023)
13. Luque, J., Straub, D.: Risk-based optimal inspection strategies for structural systems using dynamic Bayesian networks. Struct. Saf. **76**, 68–80 (2019)
14. Picheny, V., Wagner, T., Ginsbourger, D.: A benchmark of Kriging-based infill criteria for noisy optimization. Struct. Multidiscip. Optim. **48**, 607–626 (2013)
15. Quinn, J.D., Reed, P.M., Keller, K.: Direct policy search for robust multi-objective management of deeply uncertain socio-ecological tipping points. Environ. Model. Softw. **92**, 125–141 (2017)
16. Rojas-Gonzales, S., Nieuwenhuyse, V.: Inneke: a survey on kriging-based infill algorithms for multiobjective simulation optimization. Comput. Oper. Res. **116**, 104869 (2020)
17. Sacks, J., et al.: Design and analysis of computer experiments. Stat. Sci. **4**(4), 409–423 (1989)
18. Wang, W., Chen, X.: The effects of estimation of heteroscedasticity on stochastic kriging. In Winter Simulation Conference (WSC), vol. 2016, pp. 326–337. IEEE (2016)

Uncertainty Quantification for Climate Precipitation Prediction by Decision Tree

Vinicius S. Monego, Juliana A. Anochi, and Haroldo F. de Campos Velho(✉)

National Institute for Space Research (INPE), São José Dos Campos, SP, Brazil
haroldo.camposvelho@inpe.br

Abstract. Numerical weather and climate prediction have been addressed by numerical methods. This approach has been under permanent development. In order to estimate the degree of confidence on a prediction, an ensemble prediction has been adopted. Recently, machine learning algorithms have been employed for many applications. Here, the con- fidence interval for the precipitation climate prediction is addressed by a decision tree algorithm, by using the Light Gradient Boosting Machine (LightGBM) framework. The best hyperparameters for the LightGBM models were determined by the Optuna hyperparameter optimization framework, which uses a Bayesian approach to calculate an optimal hyperparameter set. Numerical experiments were carried out over South America. LightGBM is a supervised machine-learning technique. A period from January-1980 up to December-2017 was em- ployed for the learning phase, and the years 2018 and 2019 were used for testing, showing very good results.

Keywords: Climate prediction · precipitation · decision tree · uncertainty quantification

1 Introduction

Precipitation is a very difficult meteorological variable to be predicted, due to its space and time high variability. But, the rainfall is a key issue for society water planning, embracing all human activities. For Brazil, precipitation has another relevant importance, because almost 70% of energy comes from hydroelectric plants [11,13].

Operational centers for weather and climate forecasting address the prediction by solving a set of mathematical differential equations, where numerical methods are employed [8,16]. This approach is under continuous development [4]. Another issue is to verify how good is the prediction, quantifying the uncertainties (*predictability*), showing which regions the forecasting has a better/worse performance. A scheme for computing the predictability is the ensemble prediction [9]. Indeed, ensemble prediction is the standard procedure apllied to the operational centers (NCEP: National Center for Environemnt Prediction – USA, ECMWF: European Centre for Medium-range Weather Forecasts – EU, INPE:

J. E. S. De Cursi (Ed.): UQSM 2023, LNME, pp. 214–224, 2024.
https://doi.org/10.1007/978-3-031-47036-3_19

National Institute for Space Research, Brazil). Weather forecasting is a computer intensive task, requiring supercomputers. For applying ensemble prediction, the ensemble members have lower resolution than the deterministic prediction.

Here, the predictability is estimated by using a machine learning algorithm. However, instead of using artificial neural networks, as employed by Anochi and co-authors [2], a version of *decision tree* algorithm is applied: the LightGBM (*Light Gradient Boosting Machine*) algorithm [10]. The methodology with LightGBM is similar to that developed by Anochi et al. [2], where a time series is partitioned – here three months are used to compute mean and variance. Variances are our uncertainty estimations. The variance maps indicate regions with lower and/or higher predictability.

The decision tree algorithm has been already applied to meteorological issues. Anwar et al. [3] used a gradient-boosting approach for rainfall prediction; Ukkonen and Makela [15] did a comparison among four machine learning algorithms (including a decision tree) for thunderstorm events, Freitas and co-authors [7] employed a version of the decision tree algprithm to predict deep convection over the Rio de Janeiro metropolitan area, and decision tree has also be applied for rain area delineation [12].

Two decision trees are designed from the available observation data. The first one is our predictor for monthly climate precipitation forecasting. The second decision tree is configured to estimate the predictability – prediction variance (error) for each grid point. The paper's innovation is to compute the prediction uncertainty by a machine learning algorithm in a separate fashion – here, a gradient decision tree approach.

The next section does a brief description of the LightGBM decision tree scheme. The used data and numerical experiment are commented on in Sect. 3. Results are shown in Sect. 4. Finally, Sect. 5 addresses conclusions and final remarks.

2 LightGBM: A Decision Tree Algorithm

LightGBM is a supervised machine learning algorithm based on a decision tree (DT) strategy. It works by building up a framework from a basic (initial) configuration of a decision tree – this first DT is considered as a *weak learner*. For each construction step of the DT, another decision tree is added to the former, reducing the residue from the previous DT architecture. So, the current DT is understood as a *stronger learner* than a previous one.

Let F_m be our LightGBM model defined by:

$$\hat{y}_m = F_m(x) = F_{m-1}(x) + \gamma_m(x)\, h(m) \ , \tag{1}$$

where $h_m(x)$ is a DT added to the previous one $F_{m-1}(x)$, $\gamma_m(x)$ is an empirical weight function, and x is the set of attributes (inputs). The model $F_m(x)$ is DT from $m-$th step with addition of function h as a result to minimize the objective function:

$$L(y_i, \gamma) = \sum_{i=1}^{n} [y_i - \hat{y}_{m,i}]^2 + \lambda_1 \, L_1[X] + \lambda_2 \, L_2[X] \tag{2}$$

where λ_1 and λ_2 are regularization parameters, L_j $(j = 1, 2)$ regularization operators, and $X = [x_1, \; x_2, \dots \; , \; x_n]^{\mathrm{T}}$.

There are several parameters for configuring the LightGBM model. Table 1 shows the hyperparameters for this framework. The hyperparameter values have influence at the LightGBM performance.

Table 1. LightGDM hyperparameters.

Hyperparameter	Description
learning_rate	Weight contribution for every tree
max_leaves	Max. number of leaves in each tree
n_estimators	Max. number of boosted trees
reg_alpha	regularization parameter: γ_1
reg_lambda	regularization parameter: γ_2
subsample	Fraction of subsampled rows
colsample_bytree	Fraction of subsampled columns
min_child_weight	Min. number of data points needed in a leaf node
min_child_samples	Min. sum of weights required in a child

The best hyperparameter set for the LightGBM is computed by using the Optuna optimizer [1]. Optuna uses a Bayesian scheme to compute the expected improvement (EI). Its search domain for our application is shown in Table 2.

Table 2. Hyperparameters search space.

Hyperparameter	Lower threshold	Upper threshold
learning_rate	1E-2	0.1
num_leaves	20	256
n_estimators	50	1000
reg_alpha	1E-2	1
reg_lambda	1E-2	1
subsample	0.5	1.0
colsample_bytree	0.5	1.0
min_child_weight	1E-3	0.7
min_child_samples	10	1000

3 Data and Experiment Description

The area for the monthly precipitation prediction is the South America region – see Fig. 1. This continent has several climate zones: equatorial zone (with hot weather over all year, with dry and wet seasons – with intense precipitation), tropical zone (hot weather over most part of the year), subtropical and temperate zones (with four well-defined seasons, with cold winter).

The LightGBM model is configured by using data from the Global Precipitation Climatology Project (GPCP) Version 2.3: Monthly Analysis Product. Execution for the LightGBM models was carried out with processor Intel-i5 dual-core. The GPCP dataset has a space horizontal resolution of 2.5 degree grid. More details for this dataset are available from the NOAA web-page[1] The period for our experiments are in the interval from January 1980 up to December 2019. The GPCP precipitation data is employed as reference values for training and testing to the LightGBM model.

The dataset was split into three subsets: effective training, cross-validation, and testing:

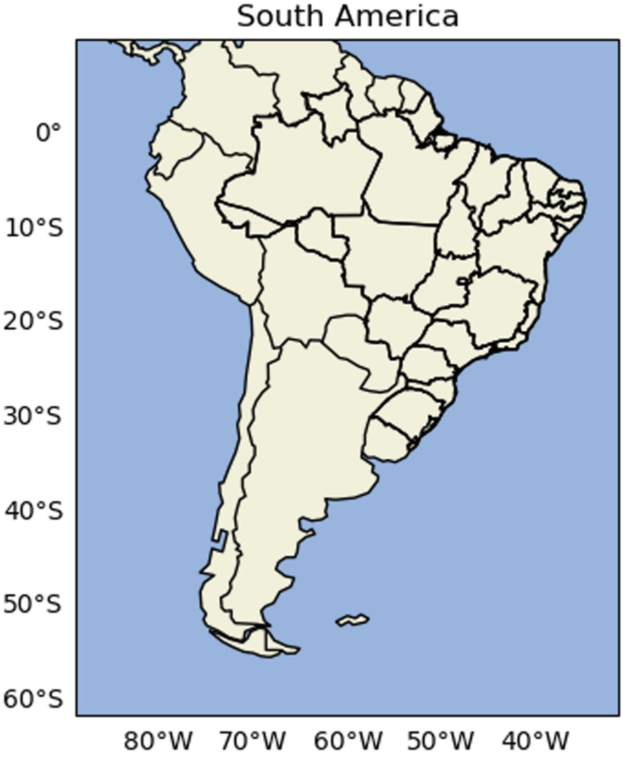


Fig. 1. Study area.

[1] See the link: https://psl.noaa.gov/data/gridded/data.gpcp.html..

Table 3. Data description.

Variable	Units	Level
Surface Pressure	millibars	surface
Air Temperature	degC	surface
Air Temperature	degC	850hPa
Specific Humidity	grams/kg	850hPa
Meridional wind	m/s	850hPa
Zonal wind	m/s	500hPa
Zonal wind	m/s	850hPa
Precipitation	mm	surface

- Training and cross-validation subsets: January-1980 up to January-2018 randomly split into 75% and 25% of the set, respectively;
- Testing (prediction) subset corresponds to the period from February-2018 up to December-2019.

The input attributes to the LightGBM model are the same as defined by Anochi et al. [2]: month, latitude, longitude, surface pressure, air temperature, specific humidity, meridional and zonal wind components, and precipitation of the current month – summarized in Table 3. The output is the predicted precipitation for the next month.

4 Results

The variance to be predicted is defined as the centered rolling variance of the error $E \equiv (P_{predicted} - P_{observed})^2$ for every 3 months, as described in the equation below:

$$\mathrm{Var}(E_k) = \frac{1}{3} \sum_{i=k-1}^{k+1} \left[E_{k,i} - \mu_k\right]^2, \quad k = 1, 2, \ldots, 12 \tag{3}$$

where $\mu_k = (1/3) \sum_{j=k-1}^{k+1} [E_{k.j}]$, being $E_{k,i}$ a set of 3 consecutive months. As seen in Sect. 3, 40 years (Jan/1980–Dec/2019) were the period for configuring and testing the LightGBM predictor. The missing values for the first (January 1980) and last (December 2017) are set to the median of the differences of December and January for 37 years available.

The evaluation for the prediction performance to the LightGBM model was the mean error (ME) and root mean squared error (RMSE):

$$\mathrm{ME} = \frac{1}{N} \sum_{i=1}^{N} \left[(P_{predicted})_i - (P_{observed})_i\right] \tag{4}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N} \sum_{k=1}^{N} \left[d_k - y_k\right]^2} \tag{5}$$

where N is the number of entries in the dataset, i.e., the number of grid points in the domain, d_k denotes the target values, and y_k are the predicted outputs.

For Model-1 (precipitation prediction), the best hyperparameters calculated by Optuna are shown in Table 4. Table 5 shows the mean (ME) and root-mean-squared error (RMSE) for Model 1.

For comparison, the error results for climate precipitation prediction obtained with the system based on differential equations – from the 3D **B**razilian **A**tmospheric global circulation **M**odel (BAM) [6] – are shown in Table 6. BAM was developed by the National Institute for Space Research (INPE, Brazil) – INPE-AGCM. The BAM is operationally executed for weather and climate (monthly/seasonal) predictions, as well as for climate change scenarios too. Climate prediction with BAM spend around 2 h in the INPE's supercomputer (Cray XT-6: 2 CPU Opteron 12-cores per node, with total of 30528 cores). From Tables 5 and 6, prediction with LightGDM presented worse values for ME quantity than the BAM, for most of the months. However, LightGBM showed a systematic better result than BAM for all months of the 2019 year. From Tables 5 and 6, prediction with LightGDM presented a worse values for ME quantity than the BAM, for most of the months. However, LightGBM showed a sistematic better result than BAM for all months of the 2019 year.

Table 4. LightGBM Model-1: monthly precipitation – optimal hyperparameters.

Hyperparameter	Optimal value
learning_rate	7.5E-2
num_leaves	244
n_estimators	619
reg_alpha	1.4E-2
reg_lambda	1.4E-2
subsample	0.63
colsample_bytree	0.86
min_child_weight	8.6E-3
min_child_samples	34

Table 5. LightGBM: Error table for precipitation.

Year	Metric	Jan	Feb	Mar	Apr	May	Jun
2019	ME	0.73	0.71	0.38	0.72	0.56	0.60
2019	RMSE	1.75	1.73	1.36	1.69	1.63	1.29
Year	Metric	Jul	Aug	Sep	Oct	Nov	Dec
2019	ME	0.69	0.56	0.71	0.65	0.41	0.90
2019	RMSE	1.32	1.23	1.66	1.35	1.53	2.36

Model-2 is used to quantify uncertainty for the climate precipitation prediction, For this LightGBM model, the best hyperparameters computed by Optuna code are shown in Table 7. Table 8 shows the mean and root-mean-squared error for Model 2.

Figure 2(a) shows the precipitation recorded in January 2019 from the GPCP dataset; Fig. 2(b) shows the forecast for the month of January from LightGBM Model-1. Figure 2(c) the uncertainty associated with observed precipitation, using the second model strategy (Model-2), which was calculated from the measured error in the precipitation forecast (Model-1), and Fig. 2(d) shows the uncertainty quantification in climate precipitation prediction by LightGBM Model-1.

Analyzing the results of the precipitation forecast from the LightGBM model, it is evident that the model was able to capture of the most intense precipitation patterns. However, LightGBM model was unable to capture the intense precipitation region over a zone embracing part of Uruguay, northeast of Argentina, and Rio Grande do Sul state (Brazil). The uncertainty was correctly predicted over the cited region – see Fig. 2(d). Ferraz et al. [5] already mentioned about the difficulty to carry out a climate prediction on that region, because models have a bias to reproduce, over that region, similar patterns found for the major

Table 6. BAM (INPE-AGCM): Error table for precipitation.

Year	Metric	Jan	Feb	Mar	Apr	May	Jun
2019	ME	0.47	0.72	−0.25	−0.42	−0.31	0.41
2019	RMSE	3.45	3.38	3.09	2.89	2.46	2.51
Year	Metric	Jul	Aug	Sep	Oct	Nov	Dec
2019	ME	0.51	0.81	0.65	0.19	−0.10	0.004
2019	RMSE	2.22	2.16	1.98	2.32	2.20	2.98

Table 7. LightGBM Model-2 (variance estimation): optimal hyperparameters.

Hyperparameter	Optimal value
learning_rate	5.2E-2
num_leaves	191
n_estimators	704
reg_alpha	0.15
reg_lambda	0.49
subsample	0.98
colsample_bytree	0.94
min_child_weight	0.38
min_child_samples	24

part of the South America territory. Machine learning algorithm, decision tree in our case, could not also overcome the reported difficulty.

Figure 3-(a) shows precipitation observed in July 2019 from the GPCP dataset; Fig. 3-(b) presents the forecast for July 2019 by LightGBM Model-1. Figure 3-(c) shows the uncertainty associated with observed precipitation, and Fig. 3-(d) shows the uncertainty quantification in precipitation prediction by LightGBM Model-2. The LightGBM model is able to identify the precipitation in the extreme north of SA related to the presence of the ITCZ (Intertropical Convergence Zone), where it was able to correctly predict the intensity of precipitation in western Colombia, Venezuela, and Guyana.

Table 8. LightGBM: Error table for uncertainty.

Year	Metric	Jan	Feb	Mar	Apr	May	Jun
2019	ME	0.31	0.34	0.003	−0.12	−0.008	0.11
2019	RMSE	3.82	2.39	3.44	2.91	2.96	2.27
Year	Metric	Jul	Aug	Sep	Oct	Nov	Dec
2019	ME	0.42	0.17	0.34	0.18	−0.74	−0.48
2019	RMSE	1.92	2.20	2.05	2.92	5.37	4.12

The application of a neural network for climate precipitation prediction presented a better estimation than forecasting using differential equations [2]. However, Monego and co-authors [14] showed that a version of the decision tree algorithm can also improve the sesonal climate precipitation prediction performed by the neural network. Our results also show a good performance for decision tree approach for monthly climate predition.

5 Final Remarks

Two models using LightGBM decision tree algorithm were applied to monthly precipitation forecasting and to estimate the variance of the prediction error – allowing to compute the confidence interval for the prediction. The model for prediction uncertainty quantification presented a good performance related to the position of the greater prediction errors. The error intensity (estimated variance) was sistematically smaller than the exact values.

The CPU-time for the prediction (LightGBM Model-1) and uncertainty quantification (LightGBM Model-2) on an Intel-i5 CPU was, in average, 156 μs ± 35.9 μs and 69.6 μs ± 1.6 μs per execution, respectively. Likewise, the training time for Model-1 and Model-2 were, respectively, 8.04 s ± 186 ms and 7.67 s ± 167 ms on the same hardware and software environment.

Preliminary results applying a decision tree approach for forecasting uncertainty quantification are shown in this paper. However, the application of machine learning algorithms for predictability allows predictability with finer

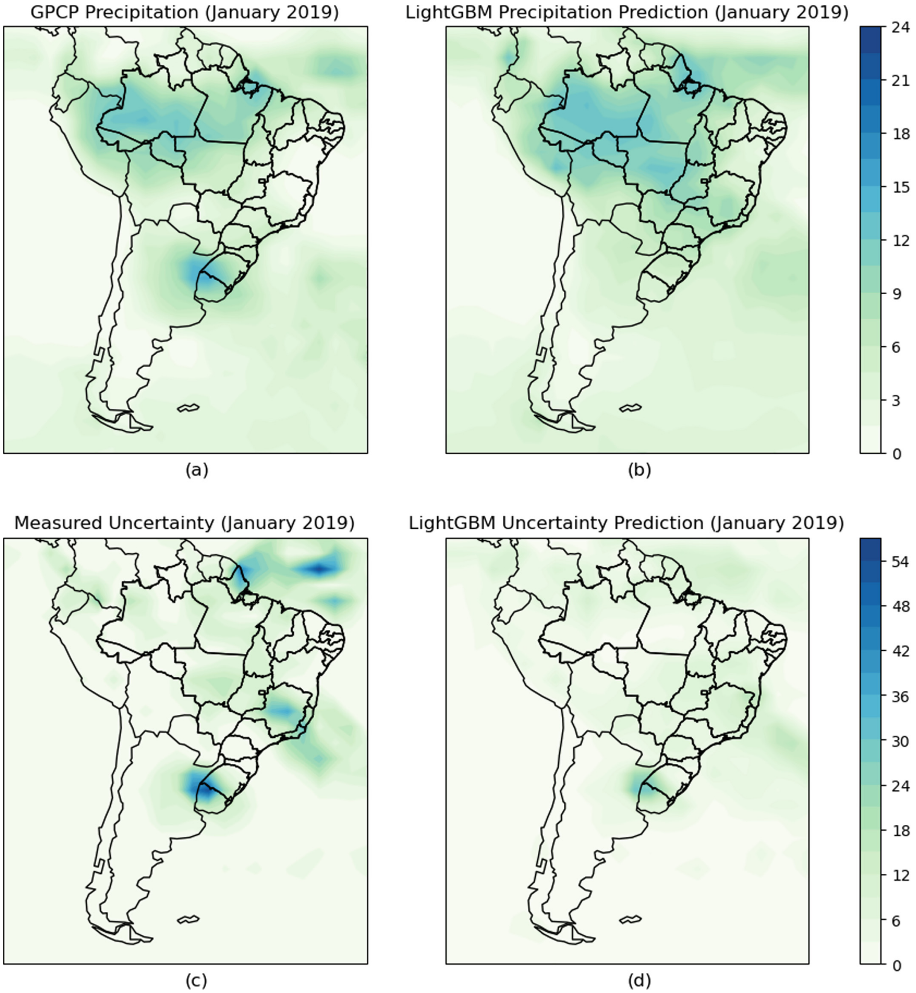

GPCP Precipitation (January 2019)
LightGBM Precipitation Prediction (January 2019)
24
21
18
15
12
9
6
3
0
(a)
(b)
Measured Uncertainty (January 2019)
LightGBM Uncertainty Prediction (January 2019)
54
48
42
36
30
24
18
12
6
0
(c)
(d)

Fig. 2. Monthly precipitation (mm/day) over South America. (a) January precipitation from GPCP. (b) January prediction from LightGBM model. (c) Uncertainty associated with observed precipitation. (d) Uncertainty quantification by LightGBM.

resolution for the prediction, instead of using a coarser resolution model as employed by the ensemble prediction.

Finally, this paper shows the application of the machine learning algorithm for climate precipitation prediction, addressing the forecasting uncertainty too. The precipitation prediction using a laptop spending with micro-seconds was better than 2 h of the supercomputer. It is important to point out that better precipitation prediction means better planning for energy production – in particular for Brazil, agriculture, and monitoring and preparing for natural disasters.

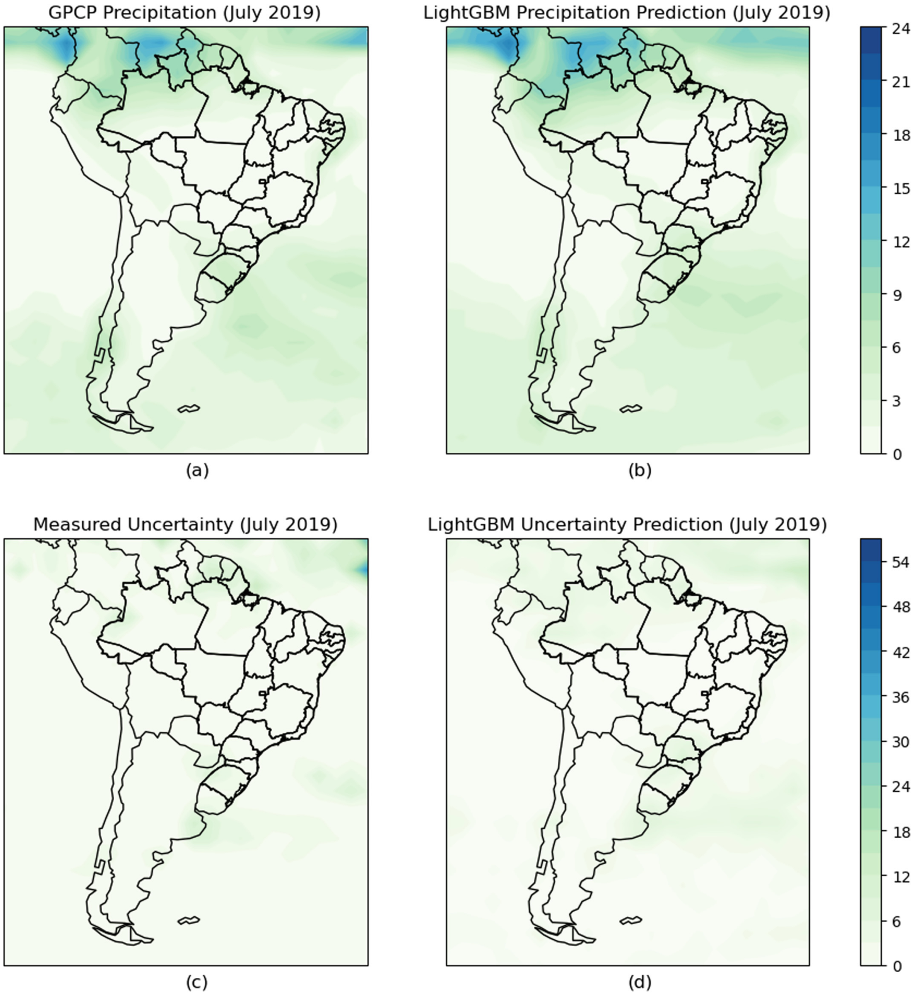

GPCP Precipitation (July 2019)
LightGBM Precipitation Prediction (July 2019)
24
21
18
15
12
9
6
3
0
(a)
(b)
Measured Uncertainty (July 2019)
LightGBM Uncertainty Prediction (July 2019)
54
48
42
36
30
24
18
12
6
0
(c)
(d)

Fig. 3. Monthly precipitation (mm/day) over South America. (a) July precipitation from GPCP. (b) July prediction from LightGBM model. (c) Uncertainty estimated from observed precipitation. (d) Uncertainty prediction by LightGBM.

Acknowledgments. Authors wish to thank the National Institute for Space Research. Author VSM thanks the Coordination for the Improvement of Higher Education Personnel (CAPES, Brazil), and author HFCV also thanks the National Council for Scientific and Technological Development (CNPq, Brazil) for the research grant (CNPq: 312924/2017-8).

References

1. Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M.: Optuna: a next-generation hyperparameter optimization framework (2019). https://arxiv.org/abs/1907.10902
2. Anochi, J.A., Hernández Torres, R., Campos Velho, H.F.: Climate precipitation prediction with uncertainty quantification by self-configuring neural network. In: De Cursi, J.E.S. (ed.) Uncertainties 2020. LNME, pp. 242–253. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-53669-5_18
3. Anwar, M.T., Winarno, E., Hadikurniawati, W., Novita, M.: Rainfall prediction using extreme gradient boosting. In: 2nd Annual Conference of Science and Technology (ANCOSET 2020), 28 November, Malang, Indonesia (2020)
4. Bauer, P., Thorpe, A., Brunet, G.: The quiet revolution of numerical weather prediction. Nature **4**(525), 47–55 (2015)
5. Ferraz, S.E.T., Souto, R.P., Dias, P.L.S., Campos Velho, H.F., Ruivo, H.M.: Analysis for precipitation climate prediction on South of Brazil, pp. 500–596. Ciência e Natura, Special Issue (2013)
6. Figueroa, S.N., et al.: The Brazilian global atmospheric model (BAM): performance for tropical rainfall forecasting and sensitivity to convective scheme and horizontal resolution. Weather Forecast. **31**(5), 1547–1572 (2016)
7. Freitas, J.H.V., França, G.B., Menezes, W.F.: Deep convection forecasting using decision tree in Rio de Janeiro metropolitan area. Anuário do Instituto de Geociências (UFRJ. Brazil) **42**(1), 127–134 (2019)
8. Haltiner, G.J., Williams, R.T.: Numerical Prediction and Dynamic Meteorology. Wiley, Hoboken (1980)
9. Kalnay, E.: Atmospheric Modeling. Data Assimilation and Predictability, Cambridge University Press, Cambridge (2002)
10. LightGBM project: https://github.com/Microsoft/LightGBM
11. Lima, G.C., Toledo, A.L.L., Bourikas, L.: The role of national energy policies and life cycle emissions of PV systems in reducing global net emissions of greenhouse gases. Energies **14**(4), 961 (2021)
12. Ma, L., Zhang, G., Lu, E.: Using the gradient boosting decision tree to improve the delineation of hourly rain areas during the summer from Aavanced Himawari imager data. J. Hydrometeorology **19**(5), 761–776 (2018)
13. Mendonça, A.K.S., Bornia, A.C.: Wind speed analysis based on the logarithmic wind shear model: a case study for some Brazilian cities. Res. Soc. Dev. **9**(7), e298973984 (2020)
14. Monego, V.S., Anochi, J.A., Campos Velho, H.F.: South America seasonal precipitation prediction by gradient-boosting machine-learning approach. Atmosphere **13**(2), 243 (2022). https://doi.org/10.3390/atmos13020243
15. Ukkonen, P., Makela, A.: Evaluation of machine learning classifiers for predicting deep convection. J. Adv. Model. Earth Syst. **11**(6), 180–1784 (2019)
16. Washington, W.M., Parkinson, C.L.: An Introduction to Three-Dimensional Climate Modeling, University Science Books (1986)

Road Accidents Forecasting: An Uncertainty Quantification Model for Pre-disaster Management in Moroccan Context

Hajar Raillani[1,2], Lamia Hammadi[1,2], Abdessamad El Ballouti[1], Vlad Stefan Barbu[3], Babacar Mbaye Ndiaye[4], and Eduardo Souza de Cursi[2](✉)

[1] Laboratory of Engineering Sciences for Energy, National School of Applied Sciences, UCD, El Jadida, Morocco

[2] Laboratory of Mechanic of Normandy LMN, INSA Rouen Normandy, Rouen, France
esc.insa.rouen@gmail.com

[3] Laboratory of Mathematics Raphaël Salem LMRS, University of Rouen - Normandy, Rouen, France

[4] Laboratory of Mathematics of Decision and Numerical Analysis, University of Cheikh Anta Diop, Dakar, Senegal

Abstract. Uncertainty quantification has become a major interest for researchers nowadays, particularly in the field of risk analysis and optimization under uncertainties. Uncertainty is an essential parameter to take into consideration in time series forecasting. In this field we aim to develop mathematical models based on uncertainty quantification tools for road accidents forecasting as a part of the pre-disaster management phase and also provide an anticipative visualization of the most sensitive zones to accidents in Morocco. To achieve this goal, we use the Interpolation-based approximation method for resolution in order to describe and analyze the road traffic accidents by defining the cumulative distribution functions (CDFs) of road accidental deaths and injuries. The obtained CDFs show that the distribution of road accidental deaths and injuries in Morocco varies according to seasons i.e., High season and Low season. These models can be used for making predictions of the future occurrence and human impact of road traffic accidents as a part of the pre-disaster management phase which complete and validate our disaster risk management approach as a decision-making tool dedicated to governments and humanitarian organizations. This work deals with humanitarian logistical field and aims to use the developed models for probabilistic calculation of the road traffic accidents behavior which helps in the preparation of the logistical fabric for the future events.

Keywords: Uncertainty Quantification · Disaster · Road Accidents · CDF · Collocation · Visualization · Morocco

1 Introduction

Disaster modeling is oriented toward using mathematical models to analyze and predict the potential effects of natural or man-made disasters. These developed models can be used to assess and mitigate the risk and the vulnerability on populations, buildings,

J. E. S. De Cursi (Ed.): UQSM 2023, LNME, pp. 225–244, 2024.
https://doi.org/10.1007/978-3-031-47036-3_20

infrastructure, and the environment, and to improve disaster risk reduction, preparedness, response and rehabilitation and recovery plans. Disaster modeling consists of three main concepts which are: hazard assessment, exposure assessment, and vulnerability assessment. These three components constitute the basis for disaster modeling that can be used in disaster risk management system. Disaster modeling can also be used to simulate and test different disaster hypothesis and scenarios and then make an evaluation of the effectiveness of different approaches of interventions and mitigation [1].

There are several tools and techniques dedicated to disaster modeling, such as information systems, machine learning technics, or uncertainty quantification-based models among others. The field of disaster modelling and forecasting is in a continuous evolution, and researchers are always about to develop new technics and improve new models to better understand the potential occurrence and behavior of disasters and then prepare for it.

Uncertainty quantification (UQ) is an important aspect of disaster modeling, it is a field of study that focuses on analyzing, quantifying, and managing uncertainty in mathematical models and simulations. The goal of UQ is to identify and analyze the sources of uncertainty in each system, and then trying to assess their impact on the general behavior of the system. This is achieved by developing mathematical models of the system studied, and then using statistical methods to analyze those models and estimate the uncertainty in their predictions. While quantifying uncertainty in mathematical models, different types of uncertainties can be identified which are; Aleatory uncertainty, which is a type of uncertainty that come from natural variations like measurement errors or environmental changes [2]. The other type of uncertainty is epistemic uncertainty, which is caused by incomplete knowledge or information about the system to model, like the lack of data or knowledge about specific parameters [3–5]. There are several UQ tools and techniques available to deal with these uncertainties, such as sensitivity analysis, statistical inference, Bayesian inference [6, 7], Monte Carlo simulations [4, 5], and surrogate modeling, among others. These methods can help identify the most significant sources of uncertainty in a model, estimate their impact on the predictions of model, and finally improve the accuracy and reliability of the predictions of the developed model. So, it's very important to understand the different types of uncertainties and choose the appropriate UQ tools and techniques to make more accurate predictions.

UQ has many applications in science and engineering, including the optimization in engineering problems field, the prediction of natural and technological disasters such as earthquakes, floods and accidents [8–10], and the assessment of financial risks linked to investment filed and so other decision-making. This work involves the application of the UQ model in the field of disaster modeling and the use of interpolation-based approximations to determine the cumulative density function (CDF) of the traffic accident state variables in Morocco. The Cumulative Distribution Function (CDF) can obtain various statistics about a state variable, such as: probability of an events, mean or variance. If necessary, the probability density function (PDF) can be derived by numerical differentiation of the CDF. In the current study, however, there is no need to determine the PDF, as our focus is primarily on the statistics and probabilities that can be obtained from the CDF.

In this paper, we aim to develop an uncertainty quantification-based model for road accidents in Morocco. We begin by providing a brief overview of the essence of uncertainty quantification in Sect. 2. Next, we make a numerical analysis to data of road accidents in Morocco in Sect. 3. We then present our results obtained by the Collocation method (COL) and discuss the relevance of our findings for road accidents forecasting and probabilities calculation in Sect. 4. Finally, we finish with a discussion of the benefits and limits of our research, as well as recommendations for further investigation. in Sect. 5.

2 Uncertainty Quantification-Based Model

2.1 Model Situation

Uncertainty quantification (UQ) models are based on the representation of random variables and the identification of their probability distributions (cumulative or density functions). The problem is about determining a representation for a random variable U using another random variable Z:

Consider two random vectors U and Z, possibly of different dimensions, defined on the probability space (Ω, P). Let define S as a set of functions of Z, and let PU be an element of S that best approximates U on S. In other words, PU is the function in S that is closest to U. The goal is to obtain a representation or approximation of U that satisfies $U - PU \approx 0$, indicating that U can be effectively replaced by PU for practical purposes. However, in general, it is not possible to achieve an exact representation, so the aim is to find an approximation of U that is sufficiently close to PU to make a practical difference, according to our objectives [11].

In UQ, we consider a system with an input variable denoted by $X \epsilon \mathbb{R}^k$, an internal parameter $W \epsilon \mathbb{R}^k$ and an output denoted by $U \epsilon \mathbb{R}^k$. The system's parameters are all impacted by uncertainties; Thus, they are put together as a vector of uncertain variables $Z = (X, W) \epsilon \mathbb{R}^k$.. The principal of UQ models is based on the approximation linked to the joint distribution of the couple (Z, U), where Z represents the uncertainties on the system and the variable and U refers to the response of the system [9, 11].

Now, we admit that U =U(Z) in a Hilbertian space, so we choose an adequate Hilbert basis H $= \{\Phi_i\} i \epsilon \mathbb{N}^*$, and we give the following representation to the components of $u_i = \left(u_{i1}, \ldots, u_{ik_Z}\right)$: [11, 12]

$$U = \sum_{i \in \mathbb{N}^*} u_i \Phi_i(z), u = \left(u_{i1}, \ldots, u_{ik_Z}\right) \in \mathbb{R}^{k_Z} \tag{1}$$

We have two random vectors, U and Z, defined on a probability space (Ω, P), where Ω is a subspace consisting of functions of Z. Our goal is to find an element PU from Ω that provides the best possible approximation of the variable U on Ω [11, 12]:

$$U = PU = \sum_{i \in \mathbb{N}^*} u_i \Phi_i(z) \tag{2}$$

The expression for the evolution of the system response, denoted by PU, involves unknown coefficients represented by u_i.

2.2 Interpolation-Based Approximation (Collocation)

The Interpolation-based Approximation (COL) method employs a different type of sample instead of $\{(Z_i, U_i) : i = 1, \ldots, u_k\}$ and deals with the real value on the sample, so, it is about to determine [8, 11]:

$$PU(Z) = \sum_{i \in \mathbb{N}^*} u_i \Phi_i(z) \tag{3}$$

Then:

$$\Phi(Z_i)_{\mathbb{Z}} = U_i \tag{4}$$

Such that:

$$PU(Z_i) = U_i, i = 1, \ldots, u_s \tag{5}$$

where, $U = (u_1, \ldots, u_k)^t$ is the solution corresponding to a linear system that produces u_s equations involving k unknowns.:

$$RU = M \tag{6}$$

$$\mathcal{R}_{ij} = \Phi_i(Z_i), \mathcal{M}_i = U_i(1 \leq i \leq u_s, 1 \leq i \leq k) \tag{7}$$

We look for the approximation of U by a polynomial function of Z, so the idea consists of generating a sample (Zi, Ui) for conveniently chosen values of Z. Then, we may interpolate the values of U [8, 11].

3 Numerical Analysis to Road Accidents in Morocco

3.1 Data Analysis

In this study, we used a dataset of monthly road accident data of Morocco in the period between January 2008 and May 2022 provided by The National Road Safety Agency (NARSA) of Morocco [13]. Our analysis aimed to identify patterns and trends in road accidents, deaths, and injuries in order to make predictions and give a decision-making tool to develop road safety. This analysis shows that road accidents frequency, deaths, and injuries have been very high in Morocco over the past decade, with a significant decrease observed in the period between 2020 and 2021 (during Covid 19 period) (See Fig. 1, 2 and 3).

Our analysis indicated that road accidents frequency, deaths, and injuries show a clear seasonal pattern, with higher numbers observed during the summer months (peaks occur at approximately the same month each year which is August) and lower numbers during the winter months. We also identified the trends, such as overall increase road accidents count, deaths, and injuries over the past decade (see Table 1).

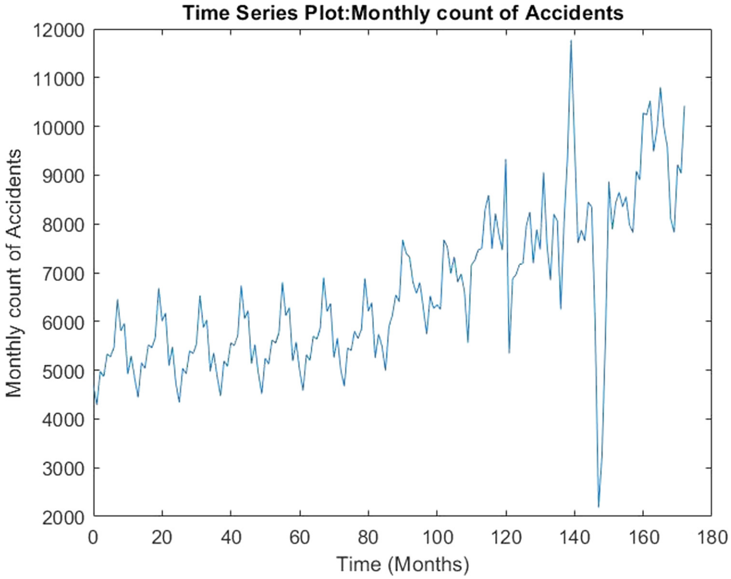


Fig. 1. Monthly count of road accidents in Morocco from January 2008 to May 2022

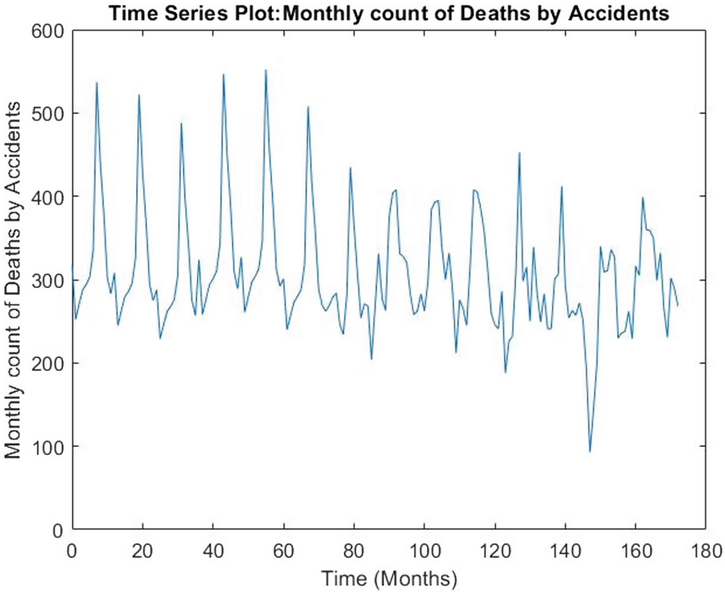


Fig. 2. Monthly count of deaths by accidents in Morocco from January 2008 to May 2022

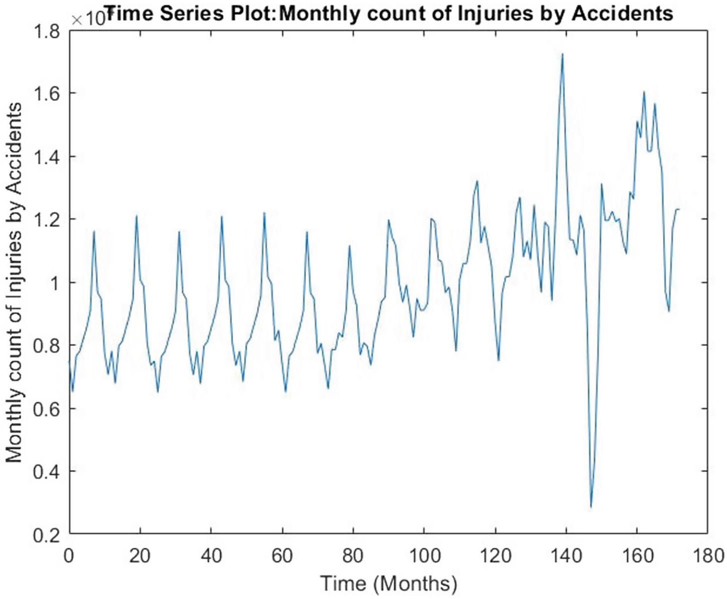


Fig. 3. Monthly count of injuries by accidents in Morocco from January 2008 to May 2022

Table 1. Morocco summary statistics per month.

Values	Cases	Deaths	Injuries	Cases p_day	Deaths p_day	Injuries p_day
Mean	6608.774	307.167	9821.289	6608.774	307.167	9821.289
Std	1608.516	71.195	2242.077	1608.516	71.195	2242.077
Min	2181.000	93.000	2838.000	2181.000	93.000	2838.000
25%	5400.000	263.000	8077.000	5400.000	263.000	8077.000
50%	6245.000	294.000	9477.000	6245.000	294.000	9477.000
75%	7649.000	332.000	11351.000	7649.000	332.000	11351.000
Max	11772.000	552.000	17251.000	11772.000	552.000	17251.000

The average increase in road accidents frequency per day is 6608 and in deaths is 307. Additionally, the number of injured per day is 9821 which aggravating the problem of emergency service capacity in Morocco.

3.2 Fatality Rate

In this study, we analyzed the fatality rate of road accidents in Morocco over a period of 15 years, from 2008 to 2022. Our study focused on calculating and analyzing the fatality rate of road accidents during this period. We calculated the fatality rate by dividing the deaths number resulting from road accidents by the total population and multiplying the result by 100,000 according to the following formula [14]:

$$FatalityRate = \frac{NumberofDeathsfromRoadAccidents}{TotalPopulation} * 100,000 \tag{8}$$

We then drow the curve of the fatality rate over the years to analyze its evolution and find any trends or patterns in the time series of the components under the study. The Fig. 4 represents the fatality rate calculated for the period between January 2008 and May 2022:

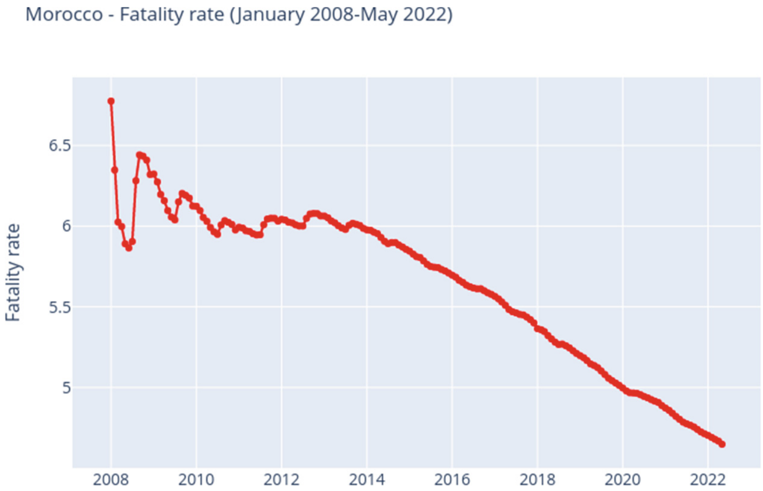


Fig. 4. Fatality rate of road accidents in Morocco from January 2008 to May 2022

Our analysis shows that the fatality rate of road accidents in Morocco has been significantly high over the past 15 years, with a significant decrease after 2016. Adding to this, we have noticed that the fatality rate was particularly high in summer season compared to winter season and among national or regional roads and highways.

3.3 Seasonality Test

The variables used in this study (Accidents count, deaths, and Injuries count) presents the behavior of seasonal time series because the values of the variable under the study show a regular pattern over a fixed time interval which is the month in this case. To determine if a series shows seasonality, we should conduct a seasonality test, which can be important because seasonality might help us in time series analysis and prediction.

To understand of the seasonality of the time series, we need to separate them into their trend, seasonal, and residual components [15–18]. The seasonal component indicates the series recurrent patterns, while the trend component represents the series' long-term behavior. The residual component refers to the fraction of the data that is not covered by the trend and seasonal components.

In this study, we used nonlinear regression to test the seasonality in time series of data considered, based the following steps developed in previous research [19, 15, 16, 18]:

- Firstly, we identified the time series data that we wanted to analyze for seasonality [20]. (See Figs. 4, 5, 6 and 7).
- Next, we chose a nonlinear regression model that can find the seasonal patterns in the data. In our application, we used trigonometric equations and then fit the chosen nonlinear regression model to the time series data [19, 15] and then we estimated the parameters of the model using the least squares method [18].

	Models Equations
Accidents count	F(x) = 18871,5–13612,6*Cos(2*Pi()*31*x)
Deaths	G(x) = 199,219 + 129,345*Cos(2*Pi()*31*x)
Injuries	H(x) = 7522,42 + 62907*Sin(2*Pi()*1,00007*x)

Finally, we tested the model effectiveness employing graphical diagnostics which are residual plots and time series plots [15–17] (Fig. 5, 6, 7, 8, 9 and 10).

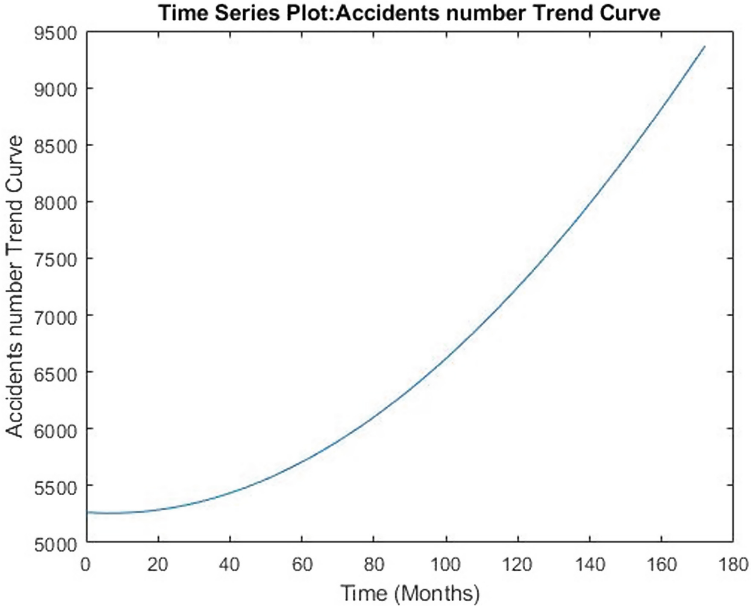


Fig. 5. Morocco: Residual components of road accidents count from January 2008 to May 2022

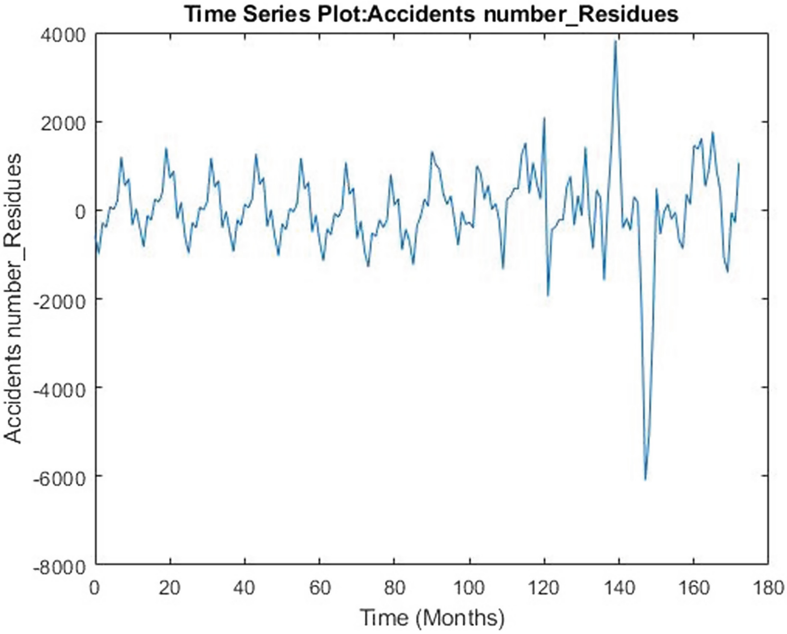


Fig. 6. Morocco: Trend curve of road accidents count from January 2008 to May 2022

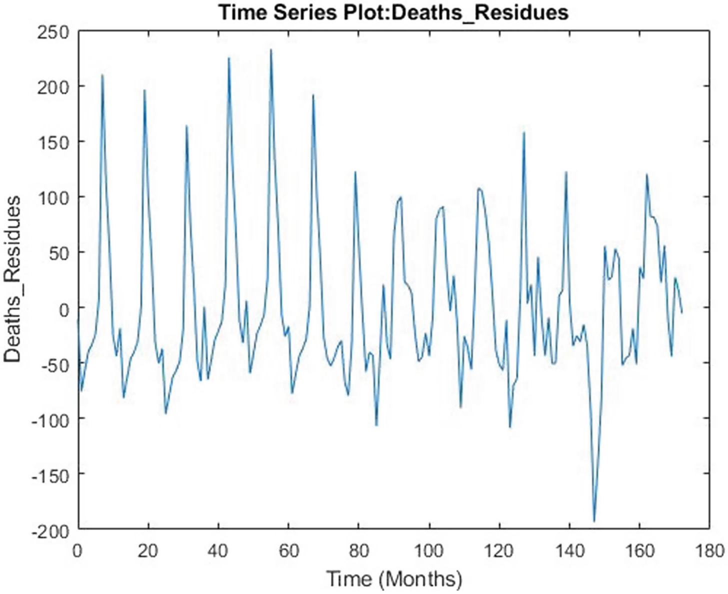


Fig. 7. Morocco: Residual components of deaths by road accidents from January 2008 to May 2022

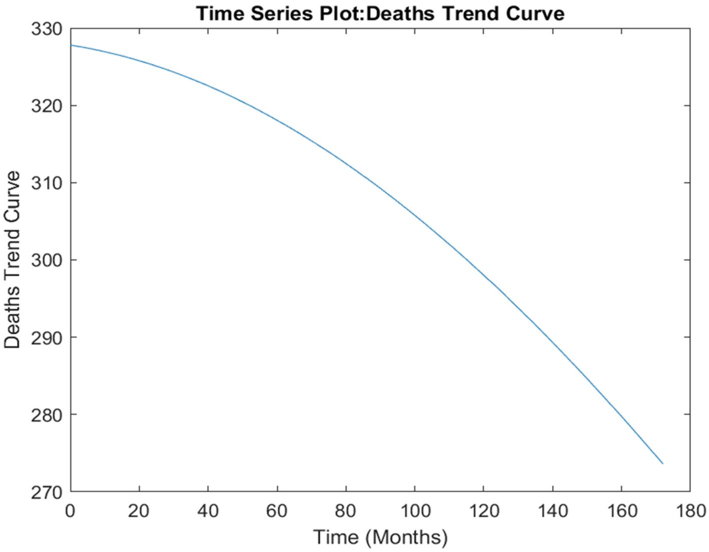


Fig. 8. Morocco: Trend curve of deaths by road accidents from January 2008 to May 2022

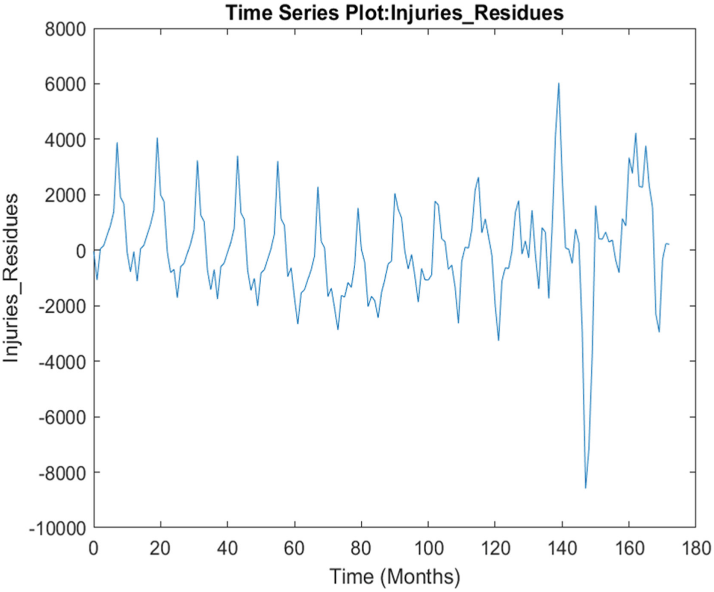


Fig. 9. Morocco: Residual components of Injuries by road accidents from January 2008 to May 2022

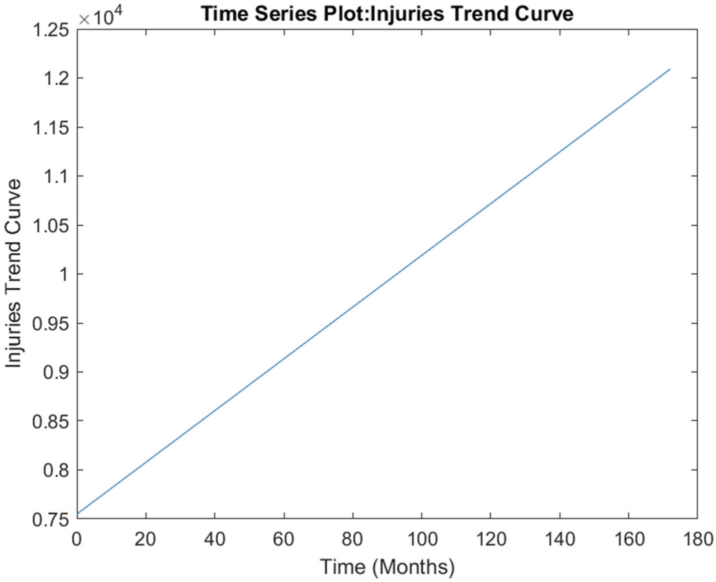


Fig. 10. Morocco: Trend curve of injuries by road accidents from January 2008 to May 2022

Finally, we can conclude that the time series data shows seasonality, and we can use the model to make forecasting of future values of the time series, taking into account the seasonal patterns. For time series modelling we consider two seasons: High season (May, June, Jully, August, September, and October) and Low season (January, February, March, April, November, and December) (see Fig. 9).

3.4 Data Visualization

To better understand the distribution of road accidents in Morocco, we created a visualization that gives the locations of the human impact of accidents over the country. By analyzing this map, we can identify and detect regions that have a higher or lower incidence of accidents and identify potential hotspots that may require more attention from government and organizations. This work can be an important decision-making tool for improving road safety and reducing the number of injuries, fatalities, and economical damage due to road accidents in Morocco (see Fig. 11).

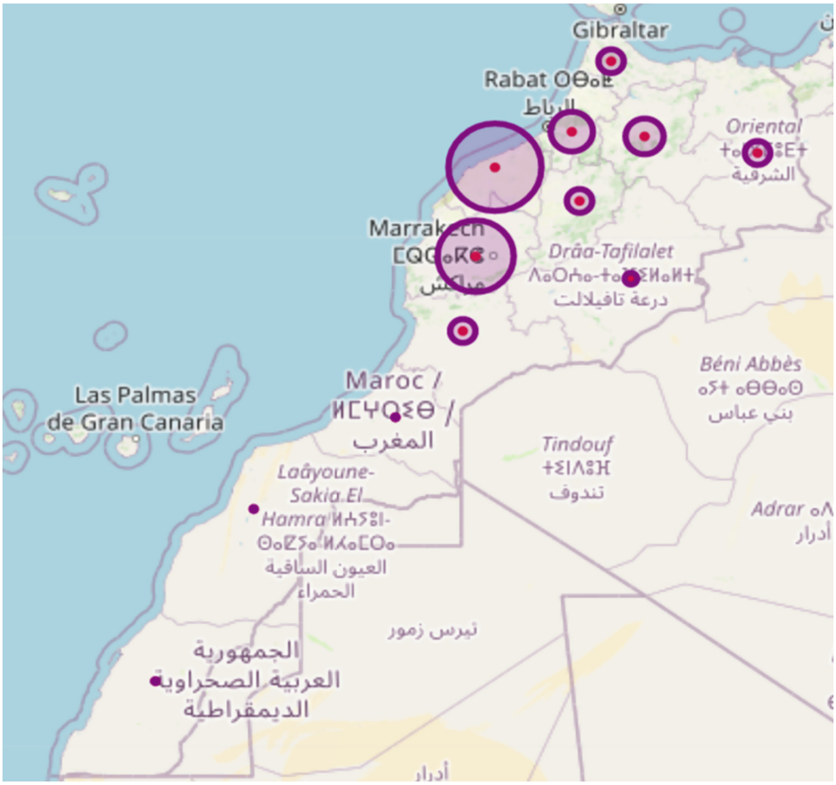


Fig. 11. Morocco 2017: Visualization map of the location of deaths and injuries by road accidents: the red hotspots represent deaths concentration, and the blue hotspots represent injuries component.

The visualization map gives a clear and informative representation of the location of road accident deaths and injuries over Morocco in 2017. The regions with the highest number of deaths and injuries are concentrated in the western and central regions of the country, particularly around urban areas with high traffic volumes, such as Casablanca Rabat, and Marrakech.

4 Results and Discussion

The results section presents the key findings of our study based on the data analysis and we give the findings using the collocation method (COL) including statistical quantities and probabilities derived from the CDF.

4.1 Root Mean Square Error (RMSE)

Root Mean Square Error (RMSE) is a commonly used parameter to measure the performance of a predictive model. It quantifies the discrepancy between the predicted and observed values of a variable and provides a way to quantify the precision of a predictive model [21].

To compute the RMSE, one needs to take the square root of the mean of the squared differences between the predicted and observed values. Mathematically, it can be expressed as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{n}} \sum_{\mathrm{i}=1}^{\mathrm{n}} (\mathrm{y_i} - \mathrm{z_i})^2} \tag{9}$$

where n is the total number of observations, $\mathrm{y_i}$ is the predicted value for the i^{th} observation, and $\mathrm{z_i}$ is the observed value for the i^{th} observation.

The root mean square error (RMSE) is a useful tool to assess the accuracy of a predictive model. The RMSE takes into consideration the size of errors in the predictions of the model. A low RMSE suggests that the model's forecasts are highly accurate and in close agreement with the observed values, while a high RMSE means that the model's forecasts are further from the observed values and, therefore, less accurate [21, 22].

4.2 Results by Collocation

The collocation (COL) method was applied to the dataset, and based on the simulations, we obtain the cumulative distribution functions (CDFs) of different components (Accidents number, Deaths, and Injuries) in high and low season. The resulting CDFs are presented in Figs. 12, 13, 14, 15, 16 and 17. These figures provide a visual representation of the probability distributions of the variables under study for the high and low seasons:

High Season. Collocation results for accidents count deaths and injuries components**.**

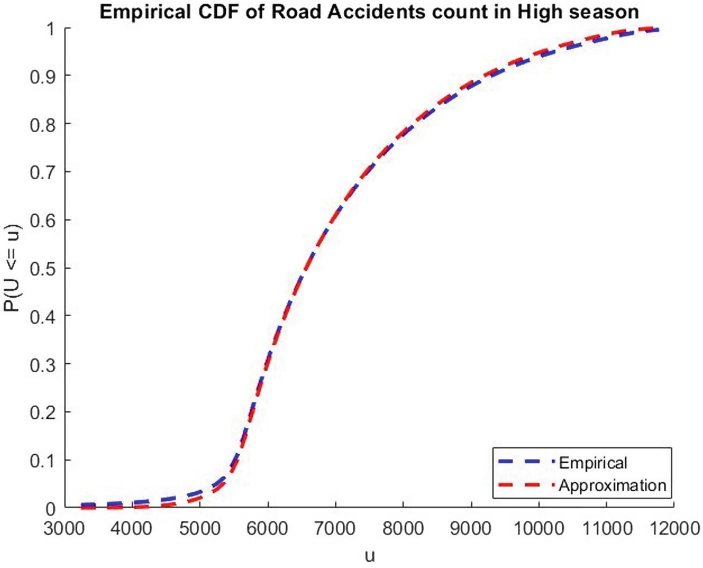


Fig. 12. Empirical CDF of deaths by road accidents in high season obtained by collocation method.

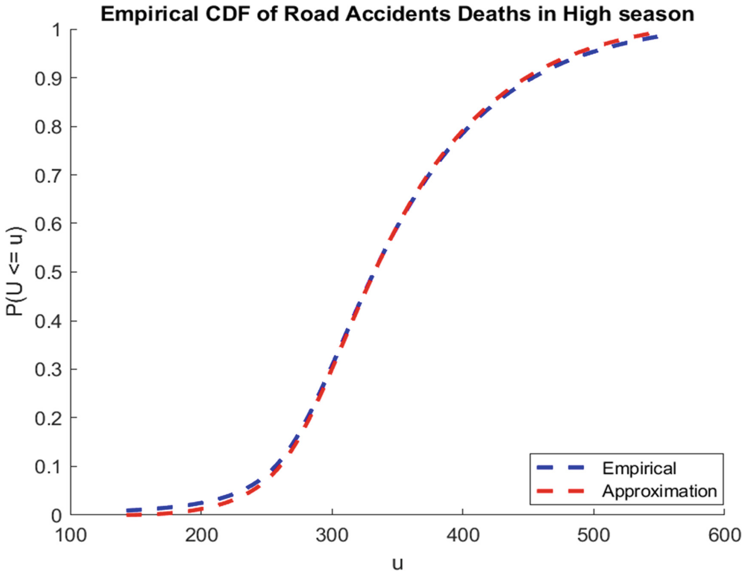


Fig. 13. Empirical CDF of deaths by road accidents in high season obtained by collocation method.

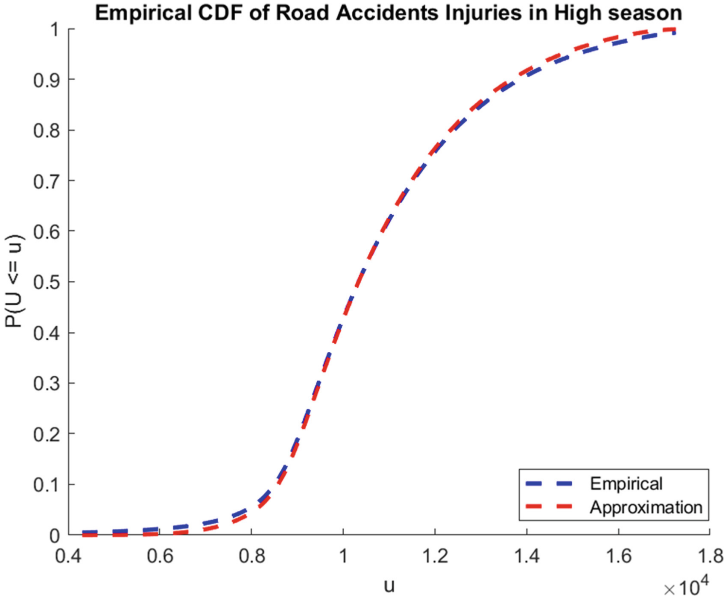


Fig. 14. Empirical CDF of injuries by road accidents in high season obtained by collocation method.

The interpolation results in the high season of deaths and injuries components indicate that the variable (u), which represents the response of the system, can be approximated by a fourth-degree polynomial function of variable (z), referring to the uncertainties affecting the system with a normal distribution. Referring to the RMSE (see Table 3), we can infer that the Collocation method gives an accurate cumulative distribution function.

Table 2. RMSE and characteristic polynomial of accidents count, deaths and injuries components in high season.

	Characteristic Polynomial	RMSE
Accidents count	$PU = 41.53 + 100{,}25U – 292.25U^2 + 486.24U^3 – 242.20U^4$	0.0078
Deaths	$PU = 1.75 + 10{,}98U – 39.93U^2 + 70.56U^3 – 37.85U^4$	0.0087
Injurie	$PU = 62.95 + 194.72U – 610.71U^2 + 1058.94U^3 – 540.54U^4$	0.0086

Low Season. Collocation results for Accidents count, deaths and injuries components.

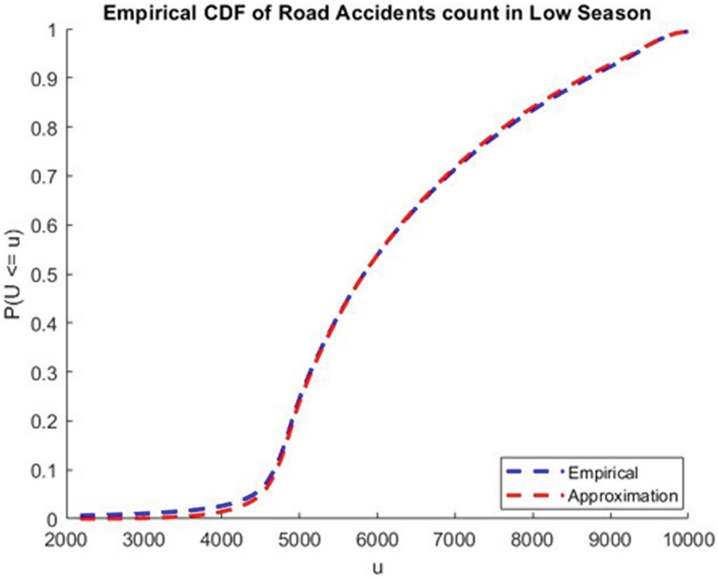


Fig. 15. Empirical CDF of road accidents count in low season obtained by collocation method.

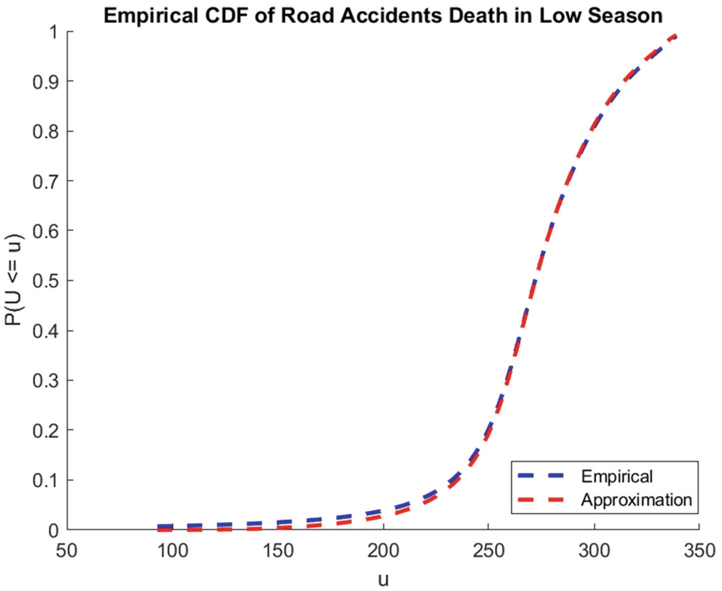


Fig. 16. Empirical CDF of deaths by road accidents in low season obtained by collocation method.

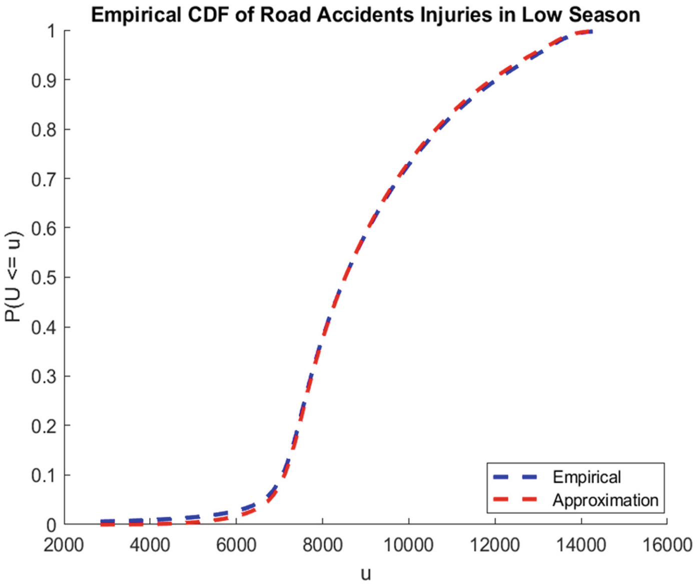


Fig. 17. Empirical CDF of injuries by road accidents in low season obtained by collocation (COL) method.

The interpolation results in the low season of deaths and injuries components indicate that the variable (u), which represents the response of the system, can be approximated by a fourth-degree polynomial function of variable (z), referring to the uncertainties affecting the system with a normal distribution. Referring to the RMSE (see Table 2), we can infer that the Collocation method gives an accurate cumulative distribution function (Table 3).

Table 3. RMSE and characteristic polynomial of accidents count, deaths and injuries components in low season.

	Characteristic Polynomial	RMSE
Accidents count	$PU = 26.55 + 213.39U – 820.53U^2 + 1.42U^3 – 749.21U^4$	0.0067
Deaths	$PU = 1.32 + 8.52U – 22.98U^2 + 29.12U^3 – 12.64U^4$	0.0081
Injuries	$PU = 62.95 + 194.73U - 610.71U^2 + 1058.94U^3 – 540.54U^4$	0.0085

4.3 Probabilities Calculation

The CDFs allow for the identification and calculation of the probabilities of a given event, such as the likelihood of a specific outcome occurring in a particular region.

In our application, one of the main objectives is to quantify the human impact degree of road accidents in Morocco [8]. To achieve this goal, we have considered different human impact classes as shown in the Table 4. This classification is based on the guidelines provided by the Ministry of Transport of Morocco [23], which provides a framework for evaluating the severity of road accidents based on different parameters such as the number of deaths, injuries, and economical damages (See Table 4).

Table 4. Gravity classes of human impact of accidents.

Gravity scale	Related Human Impact
Minor	0 to 10 deaths and/or 0 to 50 injuries
Moderate	11 to 50 deaths and/or 50 to 100 injuries
Major	More than 51 deaths and/or more than 100 injuries

The table presented below provides a detailed summary of the probabilities that were obtained using the collocation method See Table 5:

Table 5. Related probabilities obtained by collocation method (COL).

Gravity scale	Related Probabilities (Deaths)	Related Probabilities (Injuries)
Minor	0	0
Moderate	0	0
Major	1	1

This probabilities can be very important and useful for decision-makers, because it can be used to improve risk management strategies and resource allocation efforts while the preparation of the humanitarian interventions in emergency situations [24]. In general, the use of the collocation method has allowed us to develop a good understanding of the probability distributions and trends of the variables under consideration, providing valuable insights for future analysis and research. As an example, the model can be used to make predictions about the next occurrence of the accidents by considering road sections in the simulation of the model and can also provide the potential impact degree of the accidents. By predicting where and how accidents are likely to occur and behave, it is possible to implement focused interventions and strategies to improve road safety and mitigate the accidents count, injuries, and deaths.

To summarize, road accident modeling and monitoring is a key area of research and practice, because it helps in developing effective interventions and strategic plans to improve road safety, reduce road accidents frequency and mitigate their impact on citizens and society in general.

5 Conclusion

Road accidents modelling and monitoring constitutes an important field of research that aims to improve road safety, reduce road accidents frequency, and mitigate their impact and fatalities on the roads. Road accidents have a harmful economic damage and social consequences and represent a big challenge to societies, as they are responsible for numerous deaths, disabilities, and properties damages.

Our approach of modelling and monitoring road accidents was about to use an uncertainty quantification-based model and data analysis techniques to identify patterns, trends of the different components associated to road accidents. These resulted models can help to identify the potential impact of road accidents, high-risk regions, populations exposed, and this information can help for the design of specific preparation and mitigation plans with the aim of improving road safety to the users.

In conclusion, disaster modelling with uncertainty quantification-based models provides a useful tool for monitoring and predicting the potential impact of natural or human-made disasters [8]. These models consider the uncertainty associated to the system such events and provide a means for quantifying the range of potential outcomes.

By including uncertainty while disaster modelling, decision-makers can better understand the risks and vulnerabilities associated with different hypothesis and scenarios and can then develop more effective preparedness, mitigation and response strategies in their disaster management system [25]. Furthermore, uncertainty quantification-based models can help decision-makers to evaluate the effectiveness of different mitigation measures and intervention plans, allowing for more informed and cost-effective decisions.

Although challenges that disaster modelling represents, such as limited data availability, model accuracy, and computational complexity, the use of uncertainty quantification-based models is a significant progress for research in enhancing our capacity to plan and respond to disasters.

References

1. Birkmann, J., et al.: Framing vulnerability, risk and societal responses: the move framework. Nat. Hazards J. Int. Soc. Prev. Mitig. Nat. Hazards **67**(2), 193–211 (2013)
2. The Helmholtz UQ Community 2020, Types of Uncertainty — Uncertainty Quantification. https://dictionary.helmholtz-uq.de/content/types.html. Accessed 11 July (2022)
3. Oberkampf, W.L., Trucano, T.G.: Verification and validation in computational fluid dynamics. Prog. Aerosp. Sci. **38**(3), 209–272 (2002)
4. Saltelli, A., Chan, K., Scott, E.M.: Sensitivity analysis. Wiley, Hoboken (2004)
5. Smith, R.C.: Uncertainty quantification: theory, implementation, and applications, pp. XVIII + 382. SIAM (2013). ISBN: 978–1–611973–21–1
6. Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B.: Bayesian Data Analysis, 3rd edn. Chapman and Hall/CRC, New York (2015). https://doi.org/10.1201/b16018
7. Kass, R.E., Raftery, A.E.: Bayes Factors. J. Am. Stat. Assoc. **90**(430), 773–795 (1995). https://doi.org/10.1080/01621459.1995.10476572
8. Raillani, H., Hammadi, L., El Ballouti, A., Barbu, V., Cursi, E.: Uncertainty quantification for disaster modelling: flooding as a case study. Stoch. Environ. Res. Risk Assess., 1–12 (2023). https://doi.org/10.1007/s00477-023-02419-y
9. Hammadi, L., et al.: Uncertainty quantification for epidemic risk management: case of SARS-CoV-2 in Morocco. Int. J. Environ. Res. Public Health **20**, 4102 (2023). https://doi.org/10.3390/ijerph20054102
10. Qian, W., Zhang, D., Zhao, Y., Zheng, K., Yu, J.: Uncertainty quantification for traffic forecasting: a unified approach (2022). https://doi.org/10.48550/arXiv.2208.05875
11. de Cursi, E.S., Sampaio, R.: Uncertainty Quantification and Stochastic Modeling with Matlab. Numerical Methods in Engineering. ISTE Press, London (2015)
12. Souza De Cursi, E.: Uncertainty Quantification and Stochastic Modelling with EXCEL. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-77757-9. ISBN: 978-3-030-77756-2, e-ISBN 978-3-030-77757-9
13. Etudes et statistiques | Agence Nationale de la Sécurité Routière (2023). https://www.narsa.ma/fr/etudes-et-statistiques. Accessed 01 Apr 2023
14. National Highway Traffic Safety Administration. Traffic safety facts: 2017 data. National Highway Traffic Safety Administration, Washington, DC (2017)
15. Brockwell, P.J., Davis, R.A.: Introduction to Time Series and Forecasting, 2nd ed. Springer, New York (2002). in Springer texts in statistics
16. Hyndman, R.J., Athanasopoulos, G.: Forecasting: principles and practice. OTexts (2018). https://otexts.com/fpp2/
17. Liang, Y., Gillett, N.P., Monahan, A.H.: Climate model projections of 21st century global warming constrained using the observed warming trend. Geophys. Res. Lett. **47**(12), e2019GL086757 (2020). https://doi.org/10.1029/2019GL086757

18. Shumway, R.H., Stoffer, D.S.: Time Series Analysis and Its Applications: With R Examples. Springer International Publishing, Cham (2017) https://doi.org/10.1007/978-3-319-52452-8. in Springer Texts in Statistics
19. Box, G.E.P., Jenkins, G.M.: Time Series Analysis: Forecasting and Control. Wiley, Hoboken (2015)
20. Chatfield, C.: The Analysis of Time Series: An Introduction with R. CRC Press, Boca Raton (2019)
21. Ndiaye, B.M., Balde, M.A.M.T., Seck, D.: Visualization and machine learning for forecasting of COVID-19 in Senegal. arXiv: https://doi.org/10.48550/arXiv.2008.03135 (2020)
22. Willmott, C.J., Matsuura, K.: Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Clim. Res. **30**(1), 79–82 (2005)
23. OCDE 2018, CONNAITRE ET EVALUER LES RISQUES DE CATASTROPHES NATURELLES AU MAROC, Royaume du Maroc ministère de l'Intérieur, Direction du développement et de la coopération DDC (2018)
24. Raillani, H., Hammadi, L., Samed, M.M.A., Ballouti, A.E., Barbu, V.S., Cursi, E.S.D.: Disaster assessment in humanitarian relief supply chain: application to the Moroccan context. In: 2022 11th International Symposium on Signal, Image, Video and Communications (ISIVC), pp. 1–6 (2022). https://doi.org/10.1109/ISIVC54825.2022.9800727
25. Raillani, H., Hammadi, L., Samed, M., El Ballouti, A., Barbu, V.: Humanitarian logistics in the disaster relief supply chain: state of the art. WIT Trans. Eng. Sci. **129**, 181–193 (2020). https://doi.org/10.2495/RISK200161

Process Capability Indices for Dairy Product's Temperature Control in Dynamic Vehicle Routing

Khadija Ait Mamoun[1,3](✉), Lamia Hammadi[1,3], Abdessamad El Ballouti[1], Antonio G. N. Novaes[2], and Eduardo Souza De Cursi[3]

[1] Laboratory of Engineering Sciences for Energy, National School of Applied Sciences ENSAJ, UCD, El Jadida, Morocco
aitmamoun.k@ucd.ac.ma
[2] Department of Civil Engineering, Encamp - State University of Campinas, São Paulo, Brazil
[3] LMN, National Institute of Applied Sciences INSA of Rouen, Saint-Etienne-du-Rouvray, Rouen, France

Abstract. During a delivery process, and in the global transportation network chain, milk and dairy products are considered as sensible and so a higher requirement must be imposed. This paper addresses a vehicle routing problem and propose an optimization model that consider the temperature as a source of uncertainty that has an impact on dairy products. Temperature is maintained and controlled within specified interval and limits, using some sensors introduced inside the vehicles. The process capability indices are introduced to measure the capability of the process, especially thermal characteristics. Dynamic Vehicle Routing (DVR) is presented in this work, optimizing both of the distance traveled and product's temperature. The objective is to deliver products to different BIM stores in El Jadida city, and find the optimal route while maintaining the dairy product Temperature in their optimal values. We propose then a developed algorithm using the meta-heuristic Simulated Annealing (SA) algorithm. Numerical results show the optimized route sequence and also the optimized product's temperature along the route.

Keywords: Dairy product's temperature · Dynamic Vehicle routing (DVR) · Simulated Annealing · Process Capability Indices · Uncertainty

1 Introduction

Please In recent years, there has been a significant increase in the demand for high-quality products. Consumers are more aware of the benefits of consuming fresh and high-quality products. This has led to a growing need to ensure that refrigerated and frozen products are kept at the right temperature to maintain their quality and freshness. Cold chain distribution is a process that involves the transportation of temperature-sensitive products under controlled conditions. This type of distribution is essential for products such as fresh produce, meat, dairy products, and pharmaceuticals. Compared

J. E. S. De Cursi (Ed.): UQSM 2023, LNME, pp. 245–258, 2024.
https://doi.org/10.1007/978-3-031-47036-3_21

to regular distribution, cold chain distribution requires strict temperature control to preserve food quality. Temperature fluctuations can result in spoilage, loss of nutrients, and decreased shelf life, which can impact product quality and safety. Companies responsible for delivering products under cold chain conditions must ensure that their products are delivered to customers in different locations at minimal delivery cost while maintaining the food's quality and reducing product damage. This involves investing in appropriate refrigeration technology, monitoring temperature throughout the transportation process, and training personnel to handle and transport products safely. Maintaining the cold chain can be challenging and costly, but it is necessary to ensure that customers receive high-quality products that meet their expectations. It is also important for companies to adopt sustainable practices that minimize energy consumption and reduce the carbon footprint of their operations. Therefore, the growing demand for high-quality products has made it necessary to prioritize the maintenance of refrigerated and frozen products at the right temperature. Cold chain distribution is essential to ensure that products are delivered to customers in good quality and at minimal cost. Companies must invest in appropriate technology and adopt sustainable practices to maintain the cold chain effectively. In the field of operational research, the Vehicle Routing Problem (VRP) is a well-known problem that deals with the optimal routing of vehicles to visit a set of customers while minimizing the total distance traveled or the total cost incurred. This problem has many practical applications, including in the field of cold chain distribution, where the efficient and effective routing of vehicles is critical to maintaining the quality and freshness of temperature-sensitive products. Using VRP algorithms, companies can optimize their delivery routes and schedules, which can result in significant cost savings while ensuring that products are delivered on time and in good condition. The VRP can also help companies to reduce their carbon footprint by minimizing the number of vehicles needed and the distance traveled, leading to a more sustainable and environmentally friendly cold chain distribution. [1] have presented a model that takes real-time outside temperature into account, serve good quality of food to customers while reducing the distribution cost. Temperature control is critical for the quality and safety of temperature-sensitive products, especially during the distribution stage. When the refrigerator door is opened and closed frequently during distribution, it can lead to temperature fluctuations and damage to the products. Frozen products should be maintained between 4 °C to -1 °C to prevent them from thawing and spoiling. Refrigerated products, such as fresh produce and meat, should be kept between 2 °C to 6 °C to maintain their quality and reduce the risk of bacterial growth. Milk and dairy products are particularly sensitive to temperature, and they should be maintained between 2 °C to 7 °C to prevent spoilage and maintain their freshness. Hence, in cold chain not only minimize the total transportation cost but also keep products in good quality and high safety [2]. To keep product fresh, temperature constraint should be introduced during optimization to consider the temperature variation during the delivery process. The result would impact to reduce products case and increase customer satisfaction. The cold chain is considered as a complex network with cost efficiency, product quality, and carbon emission, environmental impacts, and cost. Hence, one of the most important challenge of the cold chain is to find the balance between cost and product quality. In addition to the cost reduction, dealing with requirements regarding product quality is challenging

[3]. In the same context [4] have developed the Simulated Annealing algorithm to get the optimal distance travelled, respecting the quality level expressed by the Capability Indices (PCI) to distribute perishable food. To keep sensible products fresh, temperature criteria must be introduced in the route optimization of the delivery process in order to control the optimal temperature values, especially during frequent door opening and hot weathers. In addition, short distance delivery, refrigerated and frozen products could be impacted by the high number of doors opening, where there is the heat ingress from outside air ([5] and [6]). There is a growing recognition of the significance of maintaining appropriate temperature conditions in various domains, including transportation, storage, and manufacturing processes. Researchers have emphasized the impact of temperature on crucial parameters such as product integrity, microbial growth, chemical reactions, and sensory attributes. For instance, [7] conducted a comprehensive study on the influence of temperature on microbial growth in food, developing a predictive model for different temperature conditions. Nevertheless, to optimize transportation for temperature-sensitive products, previous studies have primarily focused on minimizing transportation cost without explicitly considering temperature control during the delivery process. For example, [8] proposed an algorithm for fresh meat distribution in Athens, Greece, but did not explicitly analyze temperature variations during the delivery process. The thermal behavior of products was not considered a critical factor in the specific conditions and parameters outlined in their research. Similarly, [9] compared optimization techniques for pharmaceutical distribution in West Jakarta but did not incorporate temperature control as a crucial aspect. In our work, we aim to address the gap in previous studies by explicitly considering temperature variations during the delivery process of dairy products. We recognize the importance of temperature control in preserving the quality and safety of perishable goods, such as dairy products. By incorporating temperature as a critical factor, we intend to provide valuable insights into the optimization of delivery routes that not only minimize transportation time but also ensure appropriate temperature conditions throughout the distribution process. In comparison to previous studies, our work presents a novel approach that considers temperature control during the delivery process of dairy products. By integrating temperature constraints into the vehicle routing problem, we have optimized the route delivery of dairy products while maintaining the temperature within predefined limits.

In this study, we consider a delivery process of multiple products including milk and dairy products that have to be delivered with a minimize cost while keeping products in their optimal temperature value and hence in good quality. To do that, Simulated Annealing algorithm is developed considering the temperature variation during distributing products to some BIM stores in El Jadida city. Sensors have been introduced into the delivery trucks in order to control the temperature values inside the vehicle refrigerator. This paper is recognized as follows: Sect. 2 describes the cold chain transportation, Sect. 3 defines the process capability indices, Sect. 4 presents the mathematical modelling of the Dynamic vehicle Routing and of the Simulated Annealing (SA), Sect. 5 presents the problem description including the context study and computational results, Sect. 6 contains conclusion and perspectives.

2 Literature Review

This section presents an overview of the cold chain management, and the role of refrigerated transports to ensure a good quality of milk and dairy products from the point of temperature to its destination.

2.1 Cold Chain Logistics

The cold chain logistics (CCL) is full of complexity. This complexity is due to the fact that the cold chain involves the transportation of temperature-sensitive products through thermal methods. Cold chain logistics (CCL) can be affected by different factors including temperature variation especially during hot weathers, or during a frequent door opening and a risk to have damaged products is then strongly present. This impact is noticed in general as physical, chemical and biological changes in products. [10] have showed that temperature variation can impact directly the temperature-sensitive product along the cold chain causing quality losses. Hence, temperature can be seen as the most important factor affecting the deterioration rate and postharvest lifetime [11]. Among the different part of cold chain logistics, [12] and [13] have mentioned that transportation and storage are two principles parts of the cold chain. Thus, distribution process should be optimized reducing logistics cost and in the same time avoiding product's waste. [14] considered that quality lost should be included in distribution process of perishable food. In this study, and in order to optimize distribution process while maintaining a good quality of sensitive-product, a temperature control have been introduced, this temperature is related to the time of door opening when servicing a customer. [4] have introduced the temperature criteria during a distribution process of perishable food respecting the quality level, and using the Process Capability Indices (PCI). In general, temperature variation of sensitive products is impacted by the time and the frequency of door opening when servicing the customer, especially in short distances and hot weather. In this context, temperature constraint can be considered as a source of uncertainty that should be added during optimization of the route sequences. Compared with traditional distribution logistics, requirements on product quality in cold chain logistics is a real priority, so ensuring the freshness of temperature-sensitive products is the major problem to be solved in cold chain logistics. To solve this kind of problems, vehicle routing problem (VRP) is used by researchers through literature [15], and [1]. [16] and [17] have studied the normal temperature logistics under stochastic demand. Figure 1 presents the definition of the cold chain logistics.

As shown in Fig. 1, for a good control of the cold chain, and to maintain ambient conditions, temperature variation should be kept in optimal values within the limits of acceptability. In practical terms, the control of an optimal temperature throughout the distribution process is one the most sensible tasks, especially in short distance delivery when temperature-sensitive products can be subjected to many doors opening. As temperature considered as a source of uncertainty, Dynamic Vehicle Routing (DVR) is introduced adding the temperature constraint in the initial objective function. Hence, this work intended to minimize travel cost, and in the same time keeping temperature of products in optimal values.

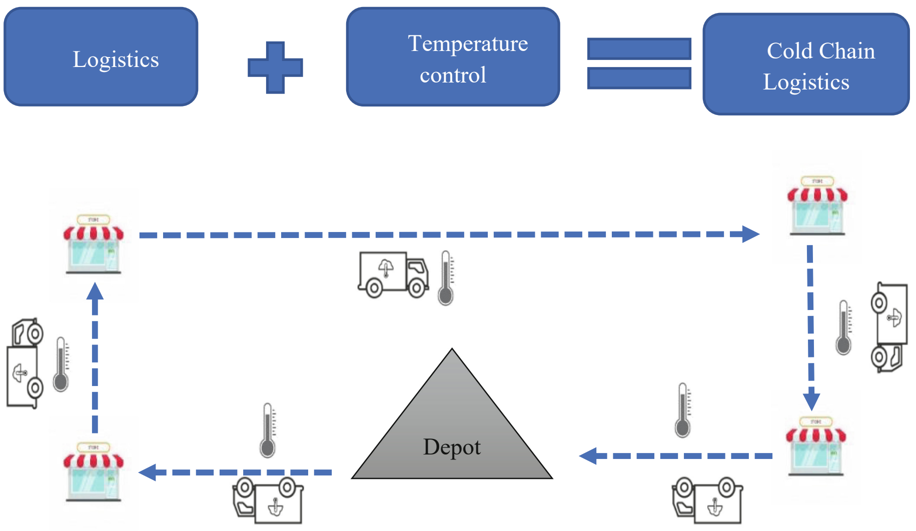


Fig. 1. Cold chain logistics.

2.2 Temperature Performance Analysis

Several models have been introduced temperature aspects in transportation process including ambient conditions, door opening, loading/unloading of products, travel time and others. Some of researchers focused on the effects of transportation temperature on microbial growth and its influence and impact on food safety. [18] and [6] studied the effects of door openings in refrigerated cargo transport. Refrigerated products can be subjected to up to fifty door openings per transportation run [5]. In short distance-transportation of temperature-sensitive products, as in the case of this work, it is important to control the temperature of the product within defined limits. Dynamic nature of the problem is due to the fact, that temperature of sensitive products can be change during distribution process because of the ambient conditions (for example hot weathers), and the frequent door openings where there is heat ingress from outside. These factors converge to produce a complex system, where a good optimization algorithm should be implemented to obtain a balance between optimal cost and optimal product temperature and so to get good quality of products. The classical distribution process starts with a vehicle loaded at the depot and traveling to a series of customers. During the process, product's temperature is assumed to be optimal anywhere in the vehicle along the journey. In this work, to control temperature inside vehicles, sensors have been introduced to measure product's temperature. Sensors indicate input temperature when the driver arrives at the customer, and indicate also the output temperature when the driver leaves the customer. In one hand, when the driver arrives at the customer, temperature is impacted by the "time of door opening" in order to take products from the refrigerator and unload them from the truck. In the other hand, when the driver leaves the customer, temperature is impacted by the travel time: the longer the distance between customers, the more time have products for cooling. Figure 2 explains temperature variation during a delivery process for temperature-sensitive products.

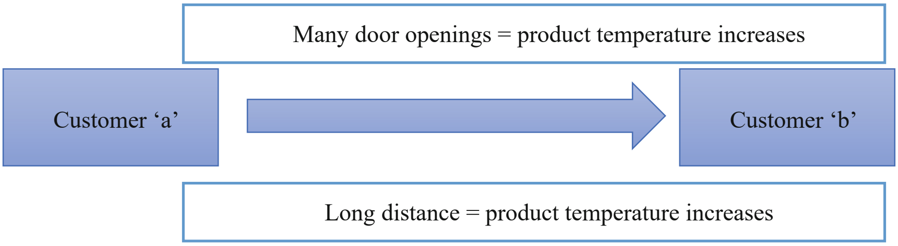

Many door openings = product temperature increases
Customer 'a'
Customer 'b'
Long distance = product temperature increases

Fig. 2. Temperature variation during delivery process.

As we can see in Fig. 2, temperature variation is affected by both travel time and door opening. These factors have been considering to implement the optimization algorithm. The aim is to minimize the travel cost within maintaining temperature in the optimal values measured by the Process Capability Indices (PCI).

3 Process Capability Indices to Assess Temperature Data

Process Capability indices (CPI) is defined as the ability of the process in achieving whether or not the mean of a measurement. It compares the characteristics of a certain process to its engineering specifications [19]. PCI is highly used as a part of statistical control of quality process and productivity. In this study, we used PCI to measure temperature performance inside the vehicle along the distribution journey for refrigerated products. Hence, temperature is the variable considered in this case. For normal distribution, four capability indices can be used: C_p, C_{pk}, C_{pm}, C_{pmk} [20]. The lesser the standard deviation, the greater the capability indices. We consider μ as the mean, and σ as the standard deviation. That is:

$$C_p = \frac{USL-LSL}{6\sigma} \tag{1}$$

where USL is the upper limit and LSL is the lower limit.

- If $C_p > 1$, then the temperature fits within the specification limits.
- If $C_p < 1$, it means that the process does not meet with the specifications.
- If $1 \leq C_p < 1.33$, it indicates that maybe the process meets with the specifications, but more attention should be taken.
- If $C_p \geq 1.33$, it indicates that the process is fully capable.

In general, the process mean is not assumed to be centered between the specified boundaries. To control these situations the index C_{pk} is defined:

$$C_{pk} = min\left\{\frac{USL-\mu}{3\sigma}, \frac{\mu-LSL}{3\sigma}\right\} \tag{2}$$

In addition, C_{pm} can considers the distance between the mean and the target value:

$$C_{pm} = \frac{USL-LSL}{6\sqrt{\sigma^2+(\mu-T)^2}} \tag{3}$$

where T is the target value.

In our application, we will not define a target value for temperature, it should only respect the specified limits, $USL = 7°C$ and $LSL = 2°C$. Thus, the index C_{pk} is employed.

4 Dynamic Optimization Model

In this section, we present the mathematical modeling of the objective function where the temperature constraint is added. Also, we present the algorithm used in this paper, Simulated Annealing (SA).

4.1 Mathematical Model

We used DVRP in this work to optimize both of cost travel and temperature variation, and a vehicle capacity is respected. Temperature is considered as a source of uncertainty and hence a source of dynamism. We consider "n" customers to be served from "i" to "j", and m vehicles. c_{ij} is the travel cost, x_{ij} is the binary variable that equal 1 if vehicle k goes from customer i to customer j, and x_{ij} equal 0 otherwise. Each vehicle has a maximal capacity Q and each customer is associated with a demand q_i of goods to be delivered. Hence the objective function contains constraint capacity that defines the classical Capacitated Vehicle Routing Problem (CVRP) [21], and contains also the penalty function of the temperature.

The objective function can be presented as follows:

$$Min \sum_{k=1}^{m} \sum_{i=0}^{n} \sum_{j=0}^{n} c_{ij}x_{ij}^{k} + \beta \sum_{i=0}^{n}\left[S\left(T_{(i)in}\right) + S\left(T_{(i)out}\right)\right] \tag{4}$$

where $\left[S\left(T_{(i)in}\right) + S\left(T_{(i)out}\right)\right]$ is the temperature penalty function and β is the penalty coefficient, that is:

$$S(T) = P^{+}(T) + P^{-}(T) \tag{5}$$

where

$$P^{+}(T) = \left(T - T_{upper}\right)^{+} = \begin{cases} 0, & T \leq T_{upper} \\ T - T_{upper}, & T > T_{upper} \end{cases} \tag{6}$$

And

$$P^{-}(T) = (T - T_{lower})^{-} = \begin{cases} 0, & T \geq T_{lower} \\ T - T_{lower}, & T < T_{lower} \end{cases} \tag{7}$$

Subject to

$$\sum_{i=0}^{n} \sum_{k=1}^{m} x_{ijk} = 1, \;\; j = 1, \ldots\ldots\ldots.n \tag{8}$$

$$\sum\nolimits_{i=0}^{n} x_{ipk} - \sum\nolimits_{j=0}^{n} x_{pjk} = 0,\ k = 1, \ldots\ldots.m;\ p = 1, \ldots\ldots.n \tag{9}$$

$$\sum\nolimits_{i=1}^{n} (q_i \sum\nolimits_{j=0}^{n} x_{ijk}) \leq Q,\ k = 1\ldots\ \ldots\ \ldots m \tag{10}$$

$$\sum\nolimits_{j=1}^{n} x_{0jk} = 1,\ \ k = 1\ldots\ldots\ldots m \tag{11}$$

Constraint (4) defines the objective function, constraints (5), (6) and (7) define the temperature penalty function, constraint (8) means that every customer should be visited exactly once, constraint (9) shows that every vehicle should depart from the customer visited by this vehicle, constraint (10) means that every vehicle should not exceed the maximal capacity Q, and constraint (11) means that every vehicle must be used exactly once.

4.2 Optimization Algorithm

Simulated Annealing (SA) is a useful meta-heuristic to solve combinatorial optimization problems. It was introduced by [22], it is an approach bases on annealing process of solids. Annealing process is based on heating a material followed by slow cooling procedure to obtain strong crystalline structure. The basic principle of SA is to move from a current solution to a random neighbor in each iteration. If the cost of neighboring solution is less than the one of the current solution, then the neighboring solution is accepted; otherwise, it is accepted or rejected with the probability $p = e^{-\frac{\Delta C}{T}}$. The probability of accepting inferior solutions is a function of the temperature T, and the change in cost between the neighboring solution and the current solution, ΔC. The temperature is decreased during the optimization process and so the probability of accepting a worse solution decreases. First, the temperature T is large and an inferior solution has a high probability of being accepted. During this step, the algorithm acts as a random search to find a promising region in the solution space. The temperature decreases as long as the optimization process progresses, and there is a lower probability of accepting an inferior solution. In general, meta-heuristics algorithms are used to solve large problems called NP-Hard. For example, [23] have used Simulated Annealing to solve large instances for a problem of no-idle open shops scheduling.

5 Case Study of Dairy Products

This section presents the context setting of this paper and numerical results concerning a delivery process of refrigerated products considering the temperature constraint.

5.1 Problem Description and Settings

The case study of this paper concerns delivering refrigerated products from the depot to 10 BIM stores in El Jadida City. The aim is to deliver these temperature-sensitive products in an optimal distance traveled while maintaining product's temperature in optimal values. In this context, we introduce the Dynamic Vehicle Routing to solve this problem using the meta-heuristic Simulated Annealing (SA).

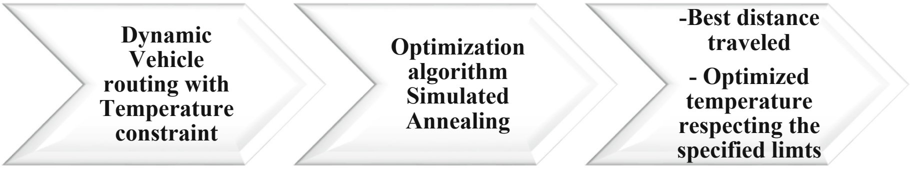

Dynamic Vehicle routing with Temperature constraint
Optimization algorithm Simulated Annealing
-Best distance traveled
- Optimized temperature respecting the specified limts

Fig. 3. Optimization process.

Figure 3 explains the optimization process of this study, introducing the Dynamic Vehicle Routing. The objective function is introducing both of minimizing the distance traveled and optimizing product's temperature. The optimized result is given by the developed algorithm Simulated Annealing on MATLAB R2018a.

We apply the following data:

- Customer number: 10 plus one depot
- Truck number: 3
- Store demand: 0, 30, 40, 20, 30, 50, 25, 10, 20, 30, 50
- Truck capacity: 110 (assume that all trucks have the same capacity = 110 box)

The matrix distance is calculated based on the customer's location (longitude and latitude from Google map). Temperature variation is calculated considering the travel time as a random variable, and the time of refrigerator door opening is based on experimental data. Such that the temperature decreases if the travel time is large, and temperature increases if we have frequent door opening, or when the time of refrigerator door opening is higher.

5.2 Computational Results and Discussion

In this work, we present route optimization and in the same time controlling temperature of milk and dairy products that must be maintained in an interval of 2 °C to 7 °C, for a delivery process to 10 BIM stores in El Jadida city. The optimization algorithm used in this work is the Simulated Annealing (SA), and Fig. 4 presents the optimized route sequence designed and (Fig. 5) presents the best cost Iteration:

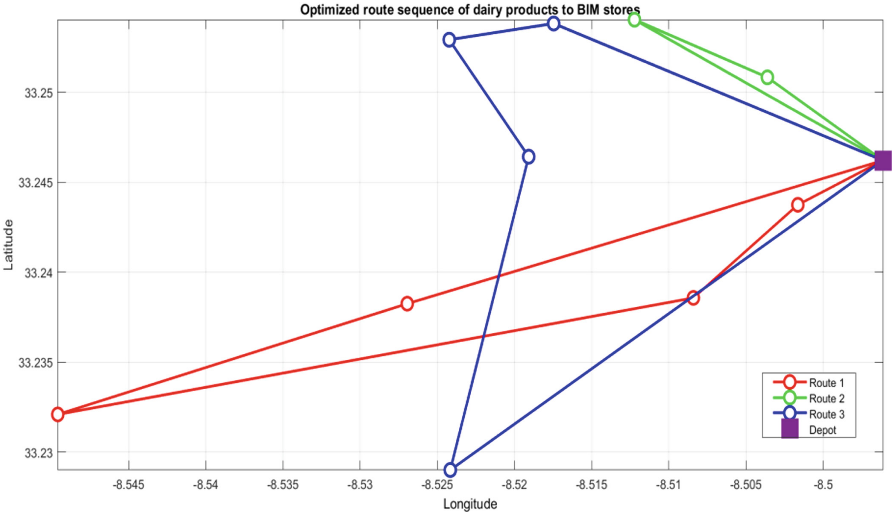


Fig. 4. Optimized route sequence.

In the figure above, optimized route sequence is presented using 100 iteration:

- **Route 1:** El Wahda -- > Touria Chaoui -- > Mwilha -- > Al Moukawama, $\boldsymbol{\mu} = 4.84$,$\boldsymbol{\sigma} = 0.66$, $\boldsymbol{C_{pk}} = 1.09$**:** in this route we can conclude that the process is capable but with tight control.
- **Route 2:** Alaouine -- > Sidi Bouzid -- > Khalil jabrane -- > Ibnou Badis, $\boldsymbol{\mu} = 4.76$, $\boldsymbol{\sigma} = 0.5$, $\boldsymbol{C_{pk}} = 1.49$**:** in this route, we conclude that the process is capable.
- **Route 3:** Mohamed Errafii -- > Lala Zahra, $\boldsymbol{\mu} = 4.35$, $\boldsymbol{\sigma} = 0.58$, $\boldsymbol{C_{pk}} = 1.35$: **in this route, the process is** capable.

C_{pk} is obtained by considering the mean μ and σ of each route, it was computed basing on the Eq. (2).

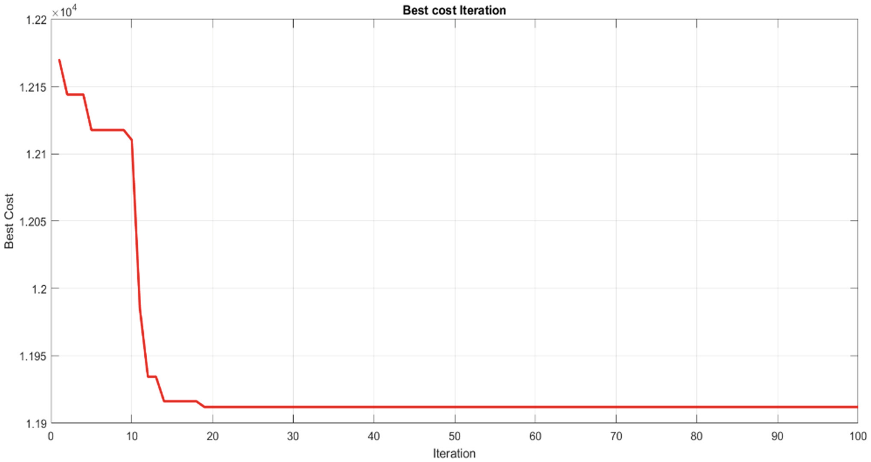


Fig. 5. Best cost Iteration.

The optimization of temperature control in the cold chain distribution system is of paramount importance to maintain the quality of dairy products during transportation. In addition to the optimized route sequence, the temperature values are also considered as an important parameter to be optimized. By using the developed algorithm, the temperature values are kept within the optimal interval, which lies between the specified limits of the milk and the dairy products, where USL = 7 °C and LSL = 2 °C. This ensures that the quality of the products is maintained and that they are following the specified limits. Then the optimized temperature is presented as follows:

For route 1:

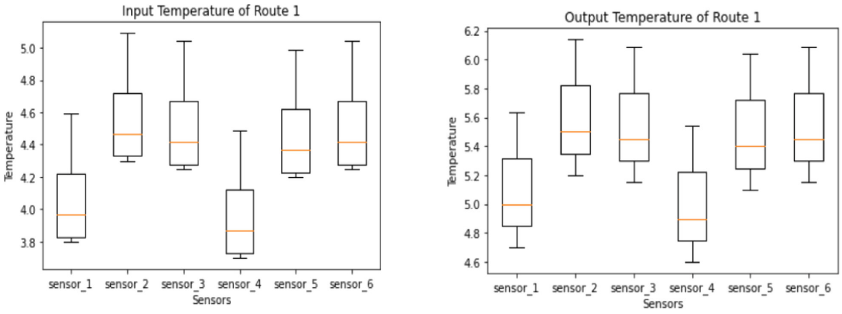


Fig. 6. Boxplot of Input and Output temperature of route 1.

For route 2:

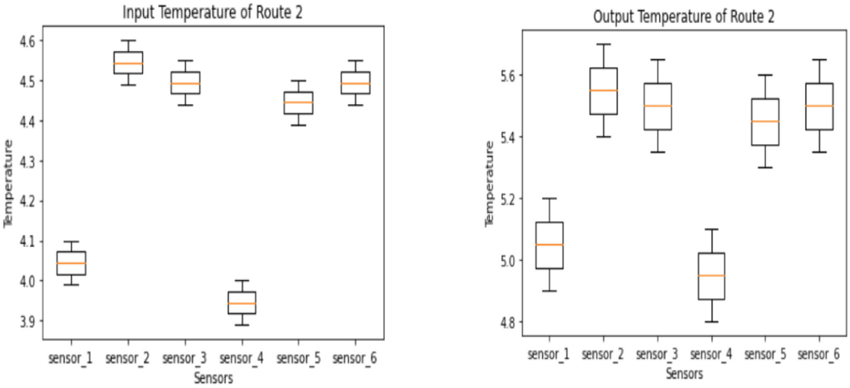


Fig. 7. Boxplot of Input and Output temperature of route 2.

For route 3:

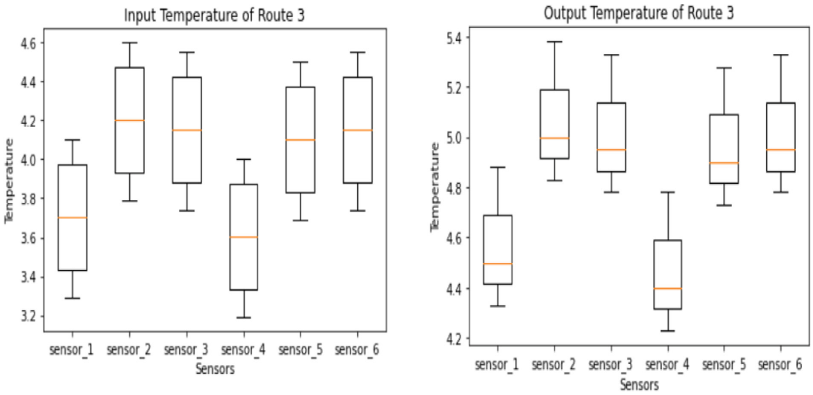


Fig. 8. Boxplot of Input and Output temperature of route 3.

Figures 6, 7, and 8 presents the boxplot of temperature values that were kept within the optimal range during cold chain distribution. The input temperature refers to the temperature of the product when the distributor arrives at the customer's location, while the output temperature represents the temperature of the product when the distributor leaves the customer's location, considering factors such as door opening frequency and discharging time. In addition, results of the developed model indicate that the product characteristics, especially product temperature inside the vehicle is related to travel time and its variation along the delivery route. Hence, specified surveillance and high attention have to be given to the product quality along short distance delivery route, and during hot weather conditions as well.

6 Conclusion and Perspectives

In this work, we propose a novel approach to the Vehicle Routing Problem (VRP) by incorporating temperature constraints as part of the objective function. The goal is to optimize the delivery route to minimize cost while simultaneously maintaining the quality of dairy products by ensuring they are kept at the appropriate temperature. The algorithm developed was applied to a simple case study, but the approach can be extended to more complex delivery processes with a larger number of customers. One of the challenges in incorporating temperature as a constraint is that it introduces an element of uncertainty, as temperature can fluctuate based on external factors such as weather conditions and door opening frequency. Temperature is considered in this work as a source of uncertainty. Hence, in future work, we can apply the simheuristic methods considering stochastic parameters such as travel time, loading/unloading time, etc. Simheuristic algorithms are very closed to a real-life distribution and transportation problems where several variables and parameters are modeled as random.

References

1. Wang, S.Y., Tao, F.M., Shi, Y.H.: Optimization of location-routing problem for cold chain logistics considering carbon footprint. Int. J. Environ. Res. Public Health **15**(1), 86–103 (2018)
2. Li, P., He, J., Zheng, D., Huang, Y., Fan, C.: Vehicle routing problem with soft time windows based on improved genetic algorithm for fruits and vegetables distribution. Discrete Dyn. Nat. Soc. (2015)
3. Behdani, B., Fan, Y., Bloemhof, J.M.: Cool chain and temperature-controlled transport: an overview of concepts, challenges, and technologies. Sustain. Food Supply Chains (2019)
4. Novaes, A.G.N., Lima, O.F., Jr., de Carvalho, C.C., Bez, E.T.: Thermal performance of refrigerated vehicles in the distribution of perishable food. Pesquisa Operacional **35**(2), 251–284 (2015). https://doi.org/10.1590/0101-7438.2015.035.02.0251
5. James, S.J., James, C., Evans, J.A.: Modelling of food transportation systems – a review. Int. J. Refrig. **29**, 947–957 (2006)
6. Evaluation of temperatures in a refrigerated container for chilled and frozen food transport (in Portuguese). Ciˆencia e Tecnologia de Alimentos, Campinas **30**(1), 158–165 (2010)
7. Zwietering, M.H., Wit, J.C.D., Notermans, S.: Application of predictive microbiology to estimate the number of Bacillus cereus in pasteurised milk at the point of consumption. Int. J. Food Microbiol. **30**(1–2), 55–70 (1996). https://doi.org/10.1016/0168-1605(96)00991-9. PMID: 8856374
8. Tarantilis, C.D., Kiranoudis, C.T.: Distribution of fresh meat. J. Food Eng. **51**, 85–91 (2002). https://doi.org/10.1016/S0260-8774(01)00040-1
9. Redi, A.A., et al.: Simulated annealing algorithm for solving the capacitated vehicle routing problem: a case study of pharmaceutical distribution. Jurnal Sistem dan Manajemen Industri **4**, 41–49 (2020)
10. Bogataj, M., Bogataj, L., Vodopivec, R.: Stability of perishable goods in cold logistic chains. Int. J. Prod. Econ. **93**, 345–356 (2005)
11. Han, J.W., Ruiz-Garcia, L., Qian, J.W., Yang, X.T.: Food Packaging: a comprehensive review and future trends. Compr. Rev. Food Sci. Food Saf. **17**, 860–877 (2018). https://doi.org/10.1016/j.asoc.2019.105733
12. Zhang, G.: Improving the structure of deep frozen and chilled food chain with Tabu search procedure. J. Food Eng. **6**, 67–79 (2003)

13. Wu, J., Chen, J., Zhu, F., Meng, F.: Developing technology of cold chain of aquatic products. Emporium Modernization J. Press **26**(4), 33–37 (2007)
14. Hsu, C.-I., Hung, S.-F., Li, H.-C.: Vehicle routing problem with time-windows for perishable food delivery. J. Food Eng. **80**, 465–475 (2007)
15. Osvald, A., Stim, L.Z.: A vehicle routing algorithm for the distribution of fresh vegetables and similar perishable food. J. Food Eng. (S0260–8774) **85**(2), 285- 295 (2008)
16. Lei, H., Laporte, G., Guo, B.: The capacitated vehicle routing problem with stochastic demands and time windows. Comput. Oper. Res. **38**(12), 1775–1783 (2011)
17. Zhang, J., Lam, W.H.K., Chen, B.Y.: On-time delivery probabilistic models for the vehicle routing problem with stochastic demands and time windows. Eur. J. Oper. Res. **249**(1), 144–154 (2016)
18. Estrada-Flores, S., Eddy, A.: Thermal performance indicators for refrigerated road vehicles. Int. J. Refrig. **29**, 889–898 (2006)
19. Chang, Y.S., Bai, D.S.: Control charts for positively-skewed populations with weighted standard deviations. Qual. Reliab. Eng. Int. **17**, 397–406 (2001)
20. Gonçalez, P.U., NC., Werner, L.: Comparison of process capability indices for non-normal distributions (in Portuguese). Gestao & Produçao **161**, 121–132 (2009)
21. Ait Mamoun, K., Hammadi, L., Novaes, A.G.N., Ballouti, A.E., Cursi, E.S.D.: An optimisation solution of capacitated vehicle routing problem (CVRP). In: 2022 11th International Symposium on Signal, Image, Video and Communications (ISIVC), El Jadida, Morocco, pp. 1-5 (2020) https://doi.org/10.1109/ISIVC54825.2022.9800726
22. Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P.: Optimization by simulated annealing. Science **220**(4598), 671–680 (1983)
23. Naderi, B., Roshanaei, V.: No-idle time scheduling of openshops: modeling and metaheuristic solution methods. Int. J. Supply Oper. Manag. **1**(1), 54–68 (2014)

Hilbert Basis Activation Function for Neural Network

J. E. Souza de Cursi[1], A. El Mouatasim[2](✉), T. Berroug[2], and R. Ellaia[3]

[1] Laboratoire de Mécanique de Normandie, Normandie Université,
INSA Rouen Normandie, 685, Avenue de l'Université Saint-Etienne du Rouvray,
Rouen, France
souza@insa-rouen.fr

[2] Department of Mathematics and Management, AMCS Team,
Faculty of Polydisciplinary Ouarzazate (FPO) - Ibnou Zohr University,
Agadir, Morocco
a.elmouatasim@uiz.ac.ma

[3] LERMA, Mohammadia School of Engineers - Mohammed V University in Rabat,
Avenue Ibn Sina BP765 - Agdal, Rabat, Morocco

Abstract. Artificial neural networks (NNs) have shown remarkable success in a wide range of machine learning tasks. The activation function is a crucial component of NNs, as it introduces non-linearity and enables the network to learn complex representations. In this paper, we propose a novel activation function based on Hilbert basis, a mathematical concept from algebraic geometry. We formulate the Hilbert basis activation function and investigate its properties. We also compare its performance with popular activation functions such as ReLU and sigmoid through experiments on MNIST dataset under LeNet architecture. Our results show that the Hilbert basis activation function can improve the performance of NNs, achieving competitive accuracy and robustness via probability analysis.

1 Introduction

Neural networks (NNs) have gained significant attention in recent years due to their remarkable performance in various machine learning tasks, such as image classification [1,2], speech recognition, and natural language processing [4,5]. NNs consist of interconnected nodes or neurons organized in layers, where each node applies an activation function to its input to introduce non-linearity and enable the network to learn complex representations [6]. The choice of activation function has a significant impact on the performance and behavior of the NN.

In this paper, we propose a novel activation function based on Hilbert basis, a mathematical concept from algebraic geometry. Hilbert basis is a set of monomials that generate the polynomial ideals in a polynomial ring [3,7]. We formulate the Hilbert basis activation function as follows:

$$f(x) = \sum_{i=1}^{n} h_i(x) + b, \tag{1}$$

J. E. S. De Cursi (Ed.): UQSM 2023, LNME, pp. 259–269, 2024.
https://doi.org/10.1007/978-3-031-47036-3_22

where x is the input to the activation function, $h_i(x)$ is the i-th monomial in the Hilbert basis, b is a bias term, and n is the number of monomials in the Hilbert basis.

The Hilbert basis activation function introduces a geometric interpretation to the activation process in NNs, and we hypothesize that it can enhance the performance of NNs by promoting geometric structures in the learned models. In this paper, we investigate the properties of the Hilbert basis activation function and compare its performance with popular activation functions such as ReLU and sigmoid through experiments on MNIST datasets.

The rest of the paper is organized as follows. In Sect. 2, we provide a brief overview of related work. In Sect. 3, we present the formulation of the Hilbert basis activation function and discuss its properties. In Sect. 4, we present the experimental results and analyze the performance of the Hilbert basis activation function compared to other activation functions. In Sect. 5, we conclude the paper and discuss future directions for research.

2 Related Work

The choice of activation function in neural networks has been an active area of research, and various activation functions have been proposed in the literature. Here, we review some of the related work on activation functions and their properties.

2.1 ReLU Activation Function

Rectified Linear Unit (ReLU) is a popular activation function that has been widely used in neural networks [8,9]. The ReLU activation function is defined as:

$$f(x) = \max(0, x), \tag{2}$$

where x is the input to the activation function.

ReLU has been shown to alleviate the vanishing gradient problem, which can occur in deep neural networks with traditional activation functions such as sigmoid and tanh. ReLU is computationally efficient and promotes sparsity in the network, as it sets negative values to zero. However, ReLU has some limitations, such as the "dying ReLU" problem where some neurons become inactive during training and never recover, and the unbounded output range which can lead to numerical instability.

2.2 Sigmoid Activation Function

The sigmoid activation function is another commonly used activation function, defined as:

$$f(x) = \frac{1}{1 + e^{-x}}, \tag{3}$$

where x is the input to the activation function.

Sigmoid function has a bounded output range between 0 and 1, which can be useful in certain applications, such as binary classification. It has been widely used in early neural network models, but it has some limitations, such as the vanishing gradient problem when the input is too large or too small, and the computational cost of exponentiation.

2.3 Other Activation Functions

There are also many other activation functions proposed in the literature, such as hyperbolic tangent (tanh), softmax, exponential linear unit (ELU), and parametric ReLU (PReLU), among others [10–12]. These activation functions have their own strengths and weaknesses, and their performance depends on the specific task and network architecture.

2.4 Geometric Interpretation in Activation Functions

Recently, there has been growing interest in exploring the geometric interpretation of activation functions in neural networks. Some researchers have proposed activation functions based on geometric concepts, such as radial basis functions (RBFs) [13] and splines [14]. These activation functions are designed to capture local geometric structures in the input space, which can improve the performance and interpretability of the neural network models.

In this paper, we propose a novel activation function based on Hilbert basis, a mathematical concept from algebraic geometry. Hilbert basis has been widely used in feature selection and dimensionality reduction methods [7,15], but its application in activation functions of neural networks has not been explored before. We hypothesize that the Hilbert basis activation function can promote geometric structures in the learned models and improve the performance of neural networks.

3 Methodology

In this section, we present the formulation of the Hilbert basis activation function for neural networks. We also discuss its properties and potential advantages compared to other activation functions.

3.1 Formulation of Hilbert Basis Activation Function

The Hilbert basis activation function is formulated as follows:

$$f(x) = \sum_{i=1}^{k} h_i(x) + b, \tag{4}$$

where x is the input to the activation function, $h_i(x)$ are the Hilbert basis functions, k is the number of Hilbert basis functions, and b is a bias term.

In this paper, the Hilbert basis functions are defined as:

$$h_i(x) = \alpha_i \cdot \max(0, x - \beta_i), \tag{5}$$

where α_i and β_i are learnable parameters for each Hilbert basis function.

The Hilbert basis activation function is designed to capture local geometric structures in the input space by using a weighted combination of max functions with different thresholds. The learnable parameters α_i and β_i allow the activation function to adaptively adjust the weights and thresholds to fit the data during training.

3.2 Properties of Hilbert Basis Activation Function

The Hilbert basis activation function has several interesting properties that make it unique compared to other activation functions:

Local Geometric Structures: The Hilbert basis activation function is designed to capture local geometric structures in the input space. The max functions with different thresholds allow the activation function to respond differently to different regions of the input space, which can help the neural network to capture complex geometric patterns in the data.

Adaptive and Learnable: The Hilbert basis activation function has learnable parameters α_i and β_i, which can be updated during training to adaptively adjust the weights and thresholds based on the data. This makes the activation function flexible and capable of adapting to the characteristics of the data, potentially leading to improved performance.

Sparse and Efficient: Similar to ReLU, the Hilbert basis activation function has a sparse output, as it sets negative values to zero. This can help reduce the computational cost and memory requirements of the neural network, making it more efficient in terms of computation and storage.

3.3 Advantages of Hilbert Basis Activation Function

The Hilbert basis activation function has several potential advantages compared to other activation functions:

Improved Geometric Interpretability: The Hilbert basis activation function is based on the Hilbert basis, a mathematical concept from algebraic geometry that has been widely used in feature selection and dimensionality reduction methods. This can potentially lead to improved interpretability of the learned models, as the activation function is designed to capture local geometric structures in the input space.

Enhanced Performance: The adaptive and learnable nature of the Hilbert basis activation function allows it to adapt to the characteristics of the data during training. This can potentially lead to improved performance, as the activation function can better model the underlying data distribution and capture complex patterns in the data.

Reduced Computational Cost: The sparse and efficient nature of the Hilbert basis activation function, similar to ReLU, can help reduce the computational cost and memory requirements of the neural network, making it more computationally efficient compared to other activation functions.

4 Computational Experiment

All of the tests were run on a personal PC with an HP i7 CPU processor running at 2.80 GHz, 16 GB of RAM, and Python 3.8 for Linux Ubuntu installed.

In this paper, we implement ANN using PyTorch3, an open source Python library for deep learning classification.

4.1 Hilbert Basis Neural Network

The Hilbert Basis Neural Network (HBNN) is a type of neural network that uses Hilbert basis functions as activation functions. It is composed of an input layer, a hidden layer, and an output layer.

Let $X \in \mathbb{R}^{n \times m}$ be the input data matrix, where n is the number of samples and m is the number of features. Let $W_1 \in \mathbb{R}^{m \times k}$ be the weight matrix that connects the input layer to the hidden layer. The hidden layer of the HBNN is defined as follows:

$$H = \max(0, XW_1 + b_1)\Phi, \tag{6}$$

where $\max(0, \cdot)$ is the rectified linear unit (ReLU) activation function, $b_1 \in \mathbb{R}^k$ is the bias term, and Φ is a matrix of k Hilbert basis functions defined as:

$$\Phi_{ij} = \frac{1}{\sqrt{j + 1/2}} \sin\left(\frac{(i + 1/2)j\pi}{k}\right), \tag{7}$$

where $i = 0, 1, \ldots, k - 1$ and $j = 0, 1, \ldots$. The output layer of the HBNN is defined as:

$$Y = HW_2 + b_2, \tag{8}$$

where $W_2 \in \mathbb{R}^{k \times p}$ is the weight matrix that connects the hidden layer to the output layer, $b_2 \in \mathbb{R}^p$ is the bias term, and p is the number of output classes Fig. 1.

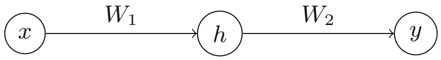


Fig. 1. HBNN model.

The loss function used to train the HBNN is the cross-entropy loss:

$$\mathcal{L} = -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{p} y_{ij}\log(\hat{y}_{ij}), \tag{9}$$

where y_{ij} is the ground truth label for sample i and class j, and $\hat{y}_{ij}$ is the predicted probability of class j for sample i.

We implemented a neural network model using Hilbert Basis Function (HBF) to classify the MNIST dataset. The HBF is a class of basis functions defined on the Hilbert space, which has been shown to be effective in approximating a wide range of functions with few parameters.

Our model consists of a single hidden layer with a nonlinear activation function based on the HBF. The input layer has 784 neurons corresponding to the 28×28 pixel images, and the hidden layer has 10 basis functions. We used the rectified linear unit (ReLU) activation function to introduce nonlinearity to the model.

We trained the model using stochastic gradient descent (SGD) or Adam with a learning rate of 0.01 and a batch size of 64. We also used weight decay with a regularization parameter of 0.01 to prevent overfitting. The model was trained for 20 epochs, and we used the cross-entropy loss function to optimize the weights.

The HBNN model can be achieved an accuracy of 96.6% on the test set for large epochs, which is comparable to the performance of other state-of-the-art models on the MNIST dataset. The use of HBF in our model allowed us to achieve good performance with a small number of parameters, making it a promising approach for neural network models with limited computational resources.

4.2 Probability Analysis

To investigate the effects of variability in the training set on the weights and loss function values of the neural network using Hilbert basis activation, we trained the model using 10 different random subsets of the training data. Each subset contained an equal number of samples, with the subsets covering the entire training set.

For each training subset, we recorded the final weights of the neural network and the corresponding value of the loss function after training for 20 epochs. We then computed the mean and standard deviation of these values across the 10 different subsets.

The results of this analysis are shown in Table 1. We observe that there is some variability in both the weights and loss function values across different training subsets. However, the standard deviations are relatively small compared to the mean values, indicating that the variability is not excessively large.

Overall, these results suggest that while there is some variability in the neural network weights and loss function values due to the selection of the training set, the effect is relatively small and should not have a major impact on the performance of the model.

Table 1. Results of probability analysis for Hilbert basis activation on MNIST dataset.

Statistic	Weights	Train Loss Function	Test Loss Function
Mean	−0.2253	0.2646	0.4520
Standard Deviation	1.3718	0.3444	0.0393

4.3 MNIST Dataset and LeNet 5

MNIST dataset, consisting 70,000 images 28 × 28 grayscale of handwritten digits in the range of 0 to 9, for a total of 10 classes, which includes 60,000 training and validation, and 10,000 test.

We present digits from the 54,000 MNIST training set to the network LeNet5 to train it, 6000 images for validation set and 10000 for test set. A 100 mini-batch size was used.

Deep learning models might have a lot of hyperparameters that need to be adjusted. The number of layers to add, the number of filters to apply to each layer, whether to subsample, kernel size, stride, padding, dropout, learning rate, momentum, batch size, and so on are all options. Because the number of possible choices for these variables is unlimited, using cross-validation to estimate any of these hyperparameters without specialist GPU technology to expedite the process is extremely challenging.

As a result, we suggest a model the LeNet-5 model. The LeNet architecture pad the input image with to make it 32 × 32 pixels, then convolution and subsampling with Relu activation in two layers. The next two layers are completely connected linear layers, followed by a layer of Gaussian connections, which are fully connected nodes that use mean squared-error as the loss function.

4.4 Comparing Results

We used optimizer Adam with fixed learning rate $lr = 1e - 3$ and $K_{\max} = 20$ epochs, the best epoch when we have training accuracy and valid accuracy. Table 2 give the results of the algorithms, and test results of the best epoch model.

Table 2. LeNet-5 model: Comparison between activation functions in terms of accuracy and loss.

Algorithm	best epoch	Training loss	Training accuracy %	Validation loss	Validation accuracy %	Test accuracy %
Relu	17	0.0083	99.86	0.0094	98.03	97.69
HP k = 1	18	0.0080	99.91	0.0091	98.35	98.06
HP k = 2	16	0.0078	99.76	0.0088	98.12	97.79
HP k = 3	17	0.0101	99.85	0.0105	98.00	97.76

Figures 2–3 exhibit the results of comparing approaches: training loss, training accuracy, validation loss, and validation accuracy, respectively.

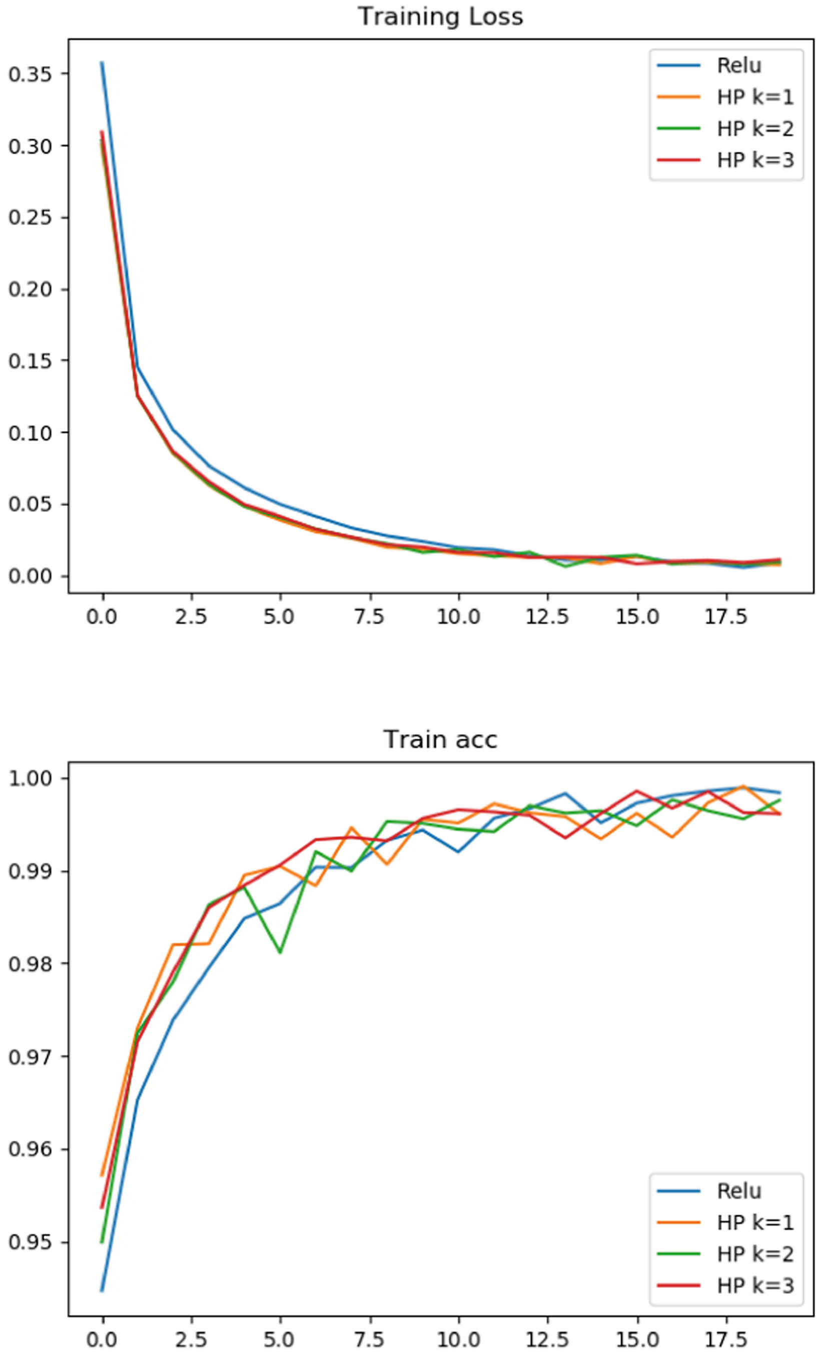


Fig. 2. LeNet-5 model Comparing training results.

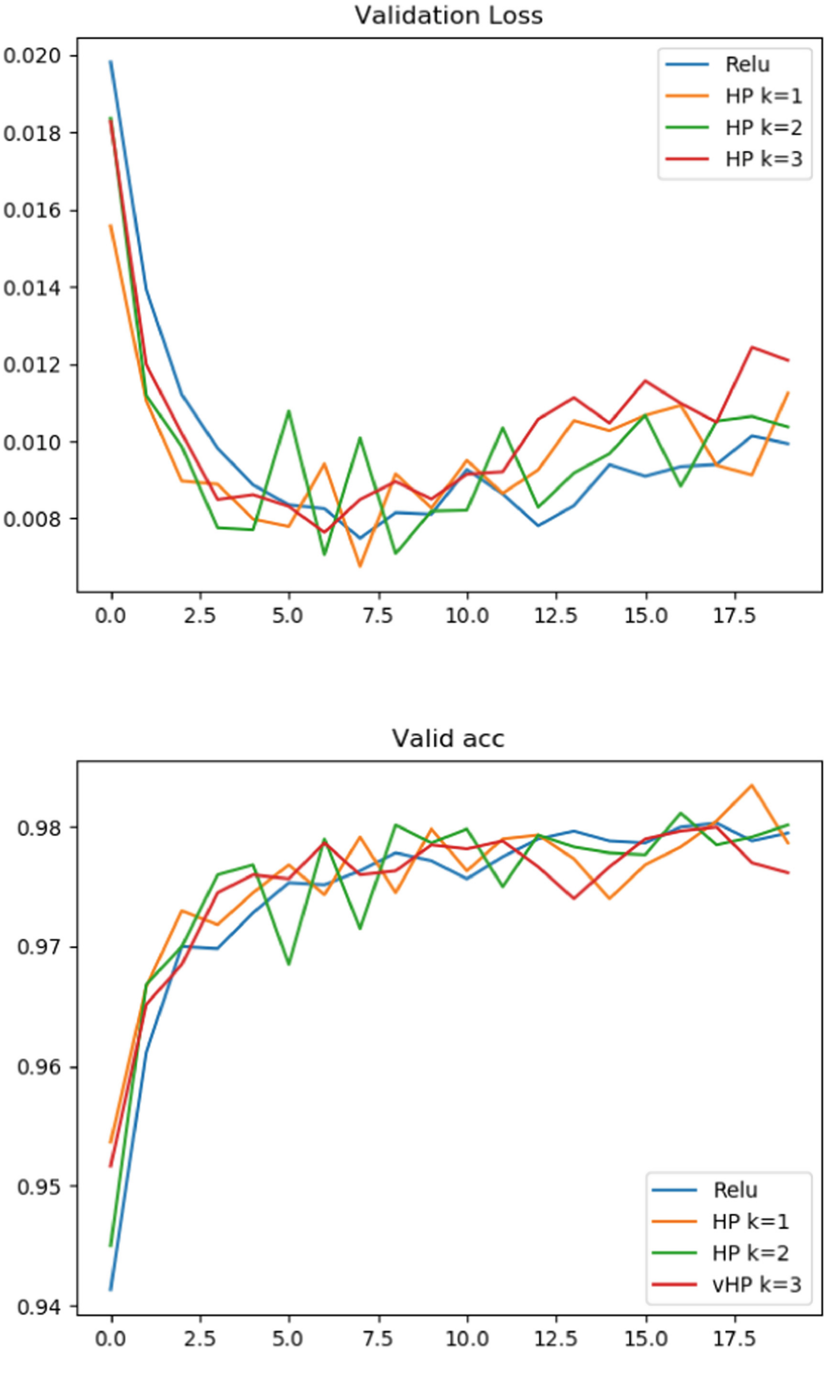


Fig. 3. LeNet-5 model Comparing evaluation results.

5 Conclusion

In this paper, we proposed a novel activation function based on Hilbert basis for neural networks. The Hilbert basis activation function is capable of capturing local geometric structures in the input space and has adaptive and learnable parameters, allowing it to adapt to the characteristics of the data during training. The experimental results demonstrate that the Hilbert basis activation function achieves competitive performance compared to other activation functions and exhibits improved geometric interpretability.

Future research directions could include further investigation of the properties and capabilities of the Hilbert basis activation function, such as its robustness to different types of data and its applicability to various neural network architectures. Additionally, exploring potential applications of the Hilbert basis activation function in other machine learning tasks, such as reinforcement learning or generative models, could be an interesting direction for future research.

In conclusion, the proposed Hilbert basis activation function offers a promising approach for enhancing the interpretability and performance of neural networks. By leveraging the local geometric structures in the input space, and incorporating adaptive and learnable parameters, the Hilbert basis activation function presents a unique and effective activation function for neural networks. Further research and experimentation can shed more light on the potential of the Hilbert basis activation function and its applications in various machine learning tasks.

References

1. El Mouatasim, A., de Cursi, J.S., Ellaia, R.: Stochastic perturbation of subgradient algorithm for nonconvex deep neural networks. Comput. Appl. Math. **42**, 167 (2023)
2. El Mouatasim, A.: Fast gradient descent algorithm for image classification with neural networks. J. Sign. Image Video Process. (SIViP) **14**, 1565–1572 (2020)
3. Khalij, L., de Cursi, E.S.: Uncertainty quantification in data fitting neural and Hilbert networks. Uncertainties **2020**(036), v2 (2020)
4. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**(7553), 436–444 (2015)
5. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. MIT Press, Cambridge (2016)
6. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer, Cham (2006)
7. Ng, A.Y., Jordan, M.I.: On spectral clustering: analysis and an algorithm. In: Advances in Neural Information Processing Systems, pp. 849-856 (2004)
8. Glorot, X., Bengio, Y.: Deep sparse rectifier neural networks. In: Proceedings of the 14th International Conference on Artificial Intelligence and Statistics, pp. 315-323 (2011)
9. Nair, V., Hinton, G.E.: Rectified linear units improve restricted boltzmann machines. In: Proceedings of the 27th International Conference on Machine Learning, pp. 807-814 (2010)
10. Agostinelli, F., Hoffman, M., Sadowski, P., Baldi, P.: Learning activation functions to improve deep neural networks. In: Proceedings of the 31st International Conference on Machine Learning, pp. 478-486 (2014)

11. Clevert, D.A., Unterthiner, T., Hochreiter, S.: Fast and accurate deep network learning by exponential linear units (ELUs). arXiv preprint: arXiv:1511.07289 (2015)
12. Maas, A.L., Hannun, A.Y., Ng, A.Y.: Rectifier nonlinearities improve neural network acoustic models. In: Proceedings of the 30th International Conference on Machine Learning, pp. 3-11 (2013)
13. Park, H.S., Sandberg, I.W.: Universal approximation using radial-basis-function networks. Neural Comput. **3**(2), 246–257 (1991)
14. Luo, C., Liu, Q., Wang, H., Lin, Z.: Spline-based activation functions for deep neural networks. In: Proceedings of the European Conference on Computer Vision, pp. 99-115 (2018)
15. Wang, H., Zhou, Z.H., Zhang, Y.: Dimensionality reduction using Hilbert-schmidt independence criterion. In: Proceedings of the 12th International Conference on Artificial Intelligence and Statistics, pp. 656-663 (2009)

Author Index

J. E. S. De Cursi (Ed.): UQSM 2023, LNME, pp. 271–272, 2024.
https://doi.org/10.1007/978-3-031-47036-3