



**HAL**  
open science

# Modèles statistiques de la carbonatation des structures en béton armé exposées à la carbonatation naturelle et introduction de covariables dans les modèles à réduction géométrique du taux de défaillance

Franck Kouassi

## ► To cite this version:

Franck Kouassi. Modèles statistiques de la carbonatation des structures en béton armé exposées à la carbonatation naturelle et introduction de covariables dans les modèles à réduction géométrique du taux de défaillance. Génie civil. INSA de Toulouse, 2023. Français. NNT : 2023ISAT0048 . tel-04548213

**HAL Id: tel-04548213**

**<https://theses.hal.science/tel-04548213>**

Submitted on 16 Apr 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# THÈSE

En vue de l'obtention du  
**DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE**  
Délivré par l'Institut National des Sciences Appliquées de  
Toulouse

---

Présentée et soutenue par  
**Franck KOUASSI**

Le 12 décembre 2023

**Modèles statistiques de la carbonatation des structures en béton armé exposées à la carbonatation naturelle et introduction de covariables dans les modèles à réduction géométrique du taux de défaillance**

---

Ecole doctorale : **EDMITT - Ecole Doctorale Mathématiques, Informatique et Télécommunications de Toulouse**

Spécialité : **Mathématiques et Applications**

Unité de recherche :  
**IMT : Institut de Mathématiques de Toulouse**

Thèse dirigée par  
**Jean-Yves DAUXOIS et Frederic DUPRAT**

Jury

M. Franck SCHOEFS, Rapporteur  
Mme Mitra FOULADIRAD, Rapporteur  
M. Akim ADEKPEDJOU, Examineur  
Mme Anne BARROS, Examinatrice  
M. Jean-Yves DAUXOIS, Directeur de thèse  
M. Frédéric DUPRAT, Co-encadrant de thèse



---

## Résumé

---

La prédiction de la durée de vie d'une structure revêt une grande importance à diverses étapes de son existence. Dans le cas des ouvrages en béton armé, la corrosion des aciers d'armature, induite par la carbonatation ou la pénétration d'ions chlorures, demeure l'une des causes majeures de défaillance. La carbonatation, processus chimique découlant de la diffusion du  $\text{CO}_2$  gazeux au sein de la matrice cimentaire, est inévitable pour les structures en béton armé non immergées. Ainsi, il devient impératif de comprendre, modéliser et prédire la carbonatation, car cela permet de programmer des opérations de maintenance correctives en vue d'améliorer la fiabilité et la sécurité de ces structures.

De nombreux modèles de carbonatation ont été élaborés pour anticiper la dégradation structurale de manière réaliste. Cependant, la mise à disposition de ces modèles représente un défi majeur, car ils requièrent la saisie de nombreux paramètres d'entrée, dont certains peuvent être difficiles à identifier, rendant ainsi leur utilisation complexe pour les ingénieurs en Génie Civil. Dans cette thèse, une approche statistique est adoptée pour élaborer des modèles de carbonatation simples et performants intégrant les principaux facteurs influençant ce phénomène. L'analyse est abordée sous deux angles. Le premier se focalise sur la profondeur de carbonatation, avec la présentation de deux modèles statistiques pour la prédire : le premier basé sur la régression linéaire généralisée et le second sur les processus de dégradation Gamma.

La seconde approche se concentre sur le temps de carbonatation, défini comme le temps requis pour que la carbonatation atteigne les armatures d'acier. Le temps de carbonatation est modélisé à l'aide d'un modèle de survie, plus particulièrement d'un modèle à risque proportionnel qui permet de prendre en compte des covariates (caractéristiques du béton, caractéristiques environnementales et les conditions de cure) sur la modélisation du temps de carbonatation. Ce modèle à l'avantage d'intégrer la dimension temporelle des facteurs environnementaux, ce qui permet une prédiction plus précise de leur impact sur la carbonatation et donc sur la durée de vie de la structure, même face à divers scénarios de réchauffement climatique.

D'autre part, une fois que des actions de maintenance sont entreprises pour améliorer la fiabilité d'un système ou d'une structure dans le contexte du génie civil, se pose la question cruciale de l'intégration de ces informations dans les prévisions de défaillance future. De nombreux modèles de maintenance imparfaite ont été proposés dans la littérature pour résoudre ce problème, mais peu d'entre eux intègrent l'effet des caractéristiques internes ou externes du système, également appelées ci-dessus covariables. Ainsi, cette thèse présente un modèle de maintenance imparfaite avec covariables. Plus spécifiquement, un modèle de réduction géométrique du taux de défaillance intégrant les covariables est proposé. Les estimateurs des paramètres de ce modèle sont élaborés, et les propriétés asymptotiques sont établies. En outre, une étude numérique menée sur des données simulées confirme ces propriétés asymptotiques.

---

# Abstract

---

The lifetime prediction of structures is of great importance at various stages of its existence. In the case of reinforced concrete structures, steel reinforcement corrosion induced by carbonation or chloride ion penetration remains one of the major causes of failure. Carbonation, a chemical process resulting from the diffusion of gaseous  $\text{CO}_2$  within the cement matrix, is inevitable for non-immersed reinforced concrete structures. Thus, it is imperative to understand, model, and predict carbonation, as it allows for the scheduling of corrective maintenance operations to improve the reliability and safety of these structures.

Numerous carbonation models have been developed to anticipate structural degradation realistically. However, making these models available poses a significant challenge as they require the input of numerous parameters, some of which may be difficult to identify, thus making their use complex for Civil Engineering professionals. In this thesis, a statistical approach is adopted to develop simple and efficient carbonation models that integrate the key factors influencing this phenomenon. The analysis is approached from two perspectives. The first focuses on the depth of carbonation, presenting two statistical models for prediction : the first based on generalized linear regression and the second on Gamma degradation processes.

The second approach concentrates on the time-to-carbonation, defined as the time required for carbonation to reach the steel reinforcement. The time-to-carbonation is modeled using a survival model, specifically a proportional hazard model that accounts for main covariates (concrete characteristics, environmental conditions, and curing conditions) in the modeling. This model has the advantage of integrating the temporal dimension of environmental factors, allowing for a more accurate prediction of their impact on carbonation and thus on the structure's lifetime, even in the face of various climate warming scenarios.

Furthermore, once maintenance actions are undertaken to enhance the reliability of a system or structure in the civil engineering context, the crucial question arises regarding the integration of this information into future failure forecasts. Many imperfect maintenance models have been proposed in the literature to address this issue, but few of them incorporate the effects of the system's internal or external characteristics, also referred to as covariates above. Thus, this thesis presents an imperfect maintenance model with covariates. Specifically, a geometric failure rate reduction model with covariates is proposed. Estimators of the parameters of this model are performed, and asymptotic properties are established. Additionally, a numerical study conducted on simulated data confirms these asymptotic properties.

---

## Un énorme Merci !

---

*Je tiens à exprimer ma profonde gratitude envers toutes les personnes qui ont contribué d'une manière ou d'une autre à l'achèvement de cette thèse. Votre soutien et votre implication ont été des éléments clés de cette grande aventure.*

*En premier lieu, je souhaite adresser mes remerciements chaleureux à l'équipe d'encadrement de ma thèse. Je suis particulièrement reconnaissant envers mon directeur de thèse, Jean-Yves Dauxois, pour ses conseils perspicaces, qu'ils soient d'ordre professionnel ou personnel. Sa disponibilité et sa bienveillance ont été des éléments déterminants, que ce soit à Toulouse ou à Casablanca.*

*Mes remerciements vont également à Frédéric Duprat, dont les conseils et le soutien ont été essentiels, notamment lors de la "pêche" aux données de carbonatation.*

*Thomas DeLarrard, je te suis reconnaissant pour ta disponibilité et les explications approfondies que tu m'as fournies sur la chimie des matériaux. Ton ouvrage en "Béton" a été ma bible.*

*Fabrice Deby, merci pour tes précieuses remarques.*

*Bien que l'acte d'écrire puisse être solitaire, ce manuscrit de thèse n'aurait pas atteint ce niveau sans les remarques pertinentes et les nombreuses relectures, pour lesquelles je suis particulièrement reconnaissant envers Jean-Yves et Frédéric. J'espère que cette collaboration fructueuse constituera le prélude à de futures réalisations communes.*

*J'adresse également mes sincères remerciements à Mme Mitra Fouladirad et M. Franck Schoefs pour l'honneur qu'ils ont accepté de me faire en rapportant cette thèse malgré le délais relativement juste. Leur remarques, les questions pertinentes soulevées et la qualité de leurs rapports sont très enrichissants. Mes remerciements s'étendent également à Mme Anne Barros et M. Akim Adekpedjou, qui ont accepté de faire partie de mon jury.*

*Un grand merci à tous les membres de l'IMT et du département Génie Mathématiques et Modélisation de l'INSA Toulouse pour leur accueil chaleureux et les moments partagés. Malgré les changements de bureau, les moments partagés avec mes différents collègues de bureau, tels que Mahamoud, Moustapha, Clément, Alban, Evgenia et Ousman, restent des souvenirs particulièrement appréciés. Je souhaite également exprimer ma gratitude envers l'équipe de la direction du GMM, en particulier à Sandrine et Cathy, pour leur grande gentillesse et leur disponibilité. Mes remerciements vont également aux équipes pédagogiques avec lesquelles j'ai collaboré au fil des années, que ce soit l'équipe pédagogique Math des 1A (de Math 0 à Math 2), les équipes pédagogiques des 4 et 5 ModIA, ou l'équipe pédagogique des 4GMM. J'ai beaucoup appris en travaillant avec vous.*

*Je tiens à exprimer ma gratitude envers l'INSA Toulouse pour avoir rendu ce projet possible en le finançant.*

*Mes remerciements s'étendent également à toutes les personnes rencontrées au cours de mon cursus universitaire, qui, de près ou de loin, m'ont apporté leur précieuse aide. Un merci particulier au Dr Pkata pour sa motivation constante à aller toujours plus loin. Mes remerciements vont également au Dr*

---

*Assane pour son soutien dans ma candidature au master Paris Saclay, ainsi qu'à Madame Guilloux pour avoir accepté ma candidature au Master 2 Data Science à l'université d'Evry. Mes amis du master, Julie, Judith et Herman, sont aujourd'hui pour moi une deuxième famille, et je leur en suis reconnaissant. Un Big Up spécial à la DiappyTeam. Merci pour vos remarques sur mes slides et votre soutien.*

*Je tiens à exprimer mes sincères remerciements à tous mes amis et proches. Un merci particulier à mes "LOUPS", mes frères d'autres mères, pour leur présence et leur soutien constant pendant et après ces années de "Kosso". Sans ces entraides et encouragements mutuels, aucun de nous ne serait probablement là où il est aujourd'hui, alors merci pour tout. Nicaise, merci pour ton soutien, frérot !*

*À Cynthia, ma Loulouta, merci pour ta présence et de m'avoir remonté le moral pendant ces moments à la cité universitaire et même après. Sarah, Issa, Angeo (le CD), Trevor (le Senoufou), Stéphane (le gros) et Serge (le sergent), votre soutien a été précieux.*

*Mes remerciements vont également à toutes les personnes qui m'ont soutenu à un moment donné de ma vie. Même si les circonstances ont évolué, je n'oublie pas les actions positives que vous avez entreprises pour moi.*

*Parce sans vous, rien de tout cela n'aurait été possible, je vous suis très reconnaissant chers membres de ma famille. À ma Reine Mère, aucune quantité de mots ne pourrait suffire à exprimer ma gratitude, je me contenterai donc de dire un simple "MERCI!". Merci pour tes prières et tes sacrifices.*

*À mon père, merci pour tes conseils avisés et pour m'avoir appris la patience et le sens des priorités. Ma reconnaissance s'étend également à mon parrain pour ses conseils, sa présence et son aide inestimable. Merci de m'avoir toujours encouragé à croire en moi.*

*À "ma promise", un sincère merci pour ton soutien constant, en particulier sur le plan moral, tout au long de cette aventure.*

*Je souhaite également exprimer ma gratitude envers mes tantes (tata Isa, tata Ceci, tata Lydie, tata Angie, tata Cristelle, tata Constance, Tata Anna) et mes oncles (Bruno, Oumar, Erico et Elisé) pour leurs prières et encouragements. Mes remerciements vont également à mon oncle Koffi (paix à son âme) pour avoir initié ce cheminement. Enfin, merci à ma sœur Solo pour ses prières et encouragements constants, ainsi qu'à mes petits frères (Devys, Nolann et Ethan) pour leur présence et leur soutien.*

---

# Table des matières

---

<b>Table des matières</b>	<b>v</b>
<b>Introduction Générale</b>	<b>1</b>
<b>1 État de l'art</b>	<b>4</b>
1.1 Introduction	4
1.2 Le matériau béton	4
1.2.1 Composition minéralogique du béton	4
1.2.2 Microstructure de la pâte de ciment	5
1.2.3 Propriétés de transfert dans le béton	5
1.3 La corrosion des armatures du béton armé	6
1.3.1 Généralité sur le mécanisme de corrosion	6
1.3.2 Corrosion induite par la carbonatation	6
1.3.2.1 Le mécanisme	6
1.3.2.2 Les facteurs de la carbonatation	7
1.3.3 Corrosion induite par la pénétration des ions chlorures	8
1.3.3.1 Le mécanisme	8
1.3.3.2 Les facteurs influençant l'attaque par les ions chlorures	9
1.4 Durabilité des bétons armés face à la corrosion des armatures	10
1.4.1 Modélisation de la carbonatation	10
1.4.1.1 Les modèles numériques	10
1.4.1.2 Les modèles empiriques	10
1.4.1.3 Les modèles analytiques	11
1.4.2 Modélisation de la pénétration des ions chlorure	12
1.5 Méthodes statistiques pour la modélisation de la durabilité d'un matériau	13
1.5.1 Les Modèles Linéaires Généralisés	13
1.5.1.1 Présentation générale	13
1.5.1.2 Identification de la distribution et de la fonction de lien	15
1.5.1.3 Estimation dans un GLM	17
1.5.1.4 Sélection de modèles linéaires généralisés	19
1.5.2 Les modèles de dégradation	19
1.5.2.1 Présentation générale	19
1.5.2.2 Estimation dans un modèle de dégradation Gamma	21
1.5.2.3 Ajustement d'un modèle de dégradation Gamma	22
1.5.3 Les modèles de durée de vie	22
1.5.3.1 Présentation générale	22
1.5.3.2 Modèles à risques proportionnels en présence de current status data	24



1.5.3.3	Modèles à côtes proportionnelles en présence de current status data . . . . .	28
1.6	Conclusion . . . . .	30
<b>2</b>	<b>Présentation des bases de données de modélisation</b>	<b>31</b>
2.1	Introduction . . . . .	32
2.2	Présentation de la base de données initiale et des caractéristiques environnementales . . . . .	32
2.2.1	Sources de données et description des variables . . . . .	32
2.2.2	Description des caractéristiques environnementales . . . . .	33
2.3	La base de données de la modélisation de la profondeur de carbonatation . . . . .	35
2.3.1	Description générale . . . . .	35
2.3.2	Présentation de la variable cible : la profondeur de carbonatation . . . . .	35
2.3.3	Présentation des variables explicatives . . . . .	36
2.3.3.1	L'âge $t$ du matériau . . . . .	36
2.3.3.2	La résistance moyenne à la compression $f_{cm}$ . . . . .	36
2.3.3.3	Le paramètre de cure $k_c$ . . . . .	36
2.3.3.4	Le paramètre de cure $k_{cfmt}$ . . . . .	37
2.3.3.5	Le paramètre environnemental $k_e$ . . . . .	38
2.3.3.6	Le paramètre environnemental $W_t$ . . . . .	39
2.3.3.7	Le paramètre environnemental $C_s$ . . . . .	40
2.4	La base de données de la modélisation du temps de carbonatation . . . . .	40
2.4.1	Présentation générale . . . . .	40
2.4.2	La nouvelle variable cible . . . . .	40
2.4.3	Prise en compte des variables explicatives dans les modèles de durée de vie . . . . .	42
2.5	Conclusion . . . . .	43
<b>3</b>	<b>Engineering statistical models for carbonation depth</b>	<b>45</b>
3.1	Introduction . . . . .	46
3.2	Review on carbonation models . . . . .	48
3.3	Scope of the present study . . . . .	49
3.3.1	The data set . . . . .	49
3.3.2	Variables selected for the modelling . . . . .	49
3.4	Statistical models for carbonation depth prediction . . . . .	52
3.4.1	Carbonation Model based on Generalized Linear Regression Methods . . . . .	52
3.4.1.1	A brief presentation of GLM . . . . .	52
3.4.1.2	GLM fitting of carbonation depth . . . . .	53
3.4.2	Model based on Gamma degradation models . . . . .	57
3.4.2.1	Overview of Gamma degradation process . . . . .	57
3.4.2.2	Gamma degradation model fitting of carbonation depth . . . . .	58
3.5	Application of proposed models to real sheltered structures and comparison with the <i>fib</i> model . . . . .	60
3.6	Conclusion . . . . .	61
<b>4</b>	<b>Survival analysis of the time-to-carbonation</b>	<b>63</b>
4.1	Introduction . . . . .	64
4.2	A brief presentation of survival analysis . . . . .	65
4.2.1	Some basics quantities of the survival analysis . . . . .	66
4.2.2	Some survival regression models . . . . .	66
4.2.2.1	Proportional hazard models . . . . .	67
4.2.2.2	Proportional odds models . . . . .	67

4.3	Survival model of carbonation . . . . .	68
4.3.1	The time-to-carbonation dataset . . . . .	68
4.3.2	Time-to-carbonation model : model fitting and model selection . . . . .	69
4.3.2.1	Proportional hazard model of the time-to-carbonation . . . . .	70
4.3.2.2	Proportional Odds Model of the Time-to-carbonation . . . . .	73
4.3.2.3	Selection of the best time-to-carbonation model . . . . .	75
4.3.3	Some applications of the proposed model . . . . .	75
4.3.3.1	Effect of carbonation factors on the reliability of the reinforced concrete under real climates. . . . . .	75
4.3.3.2	Analysis of the carbonation of a reinforced concrete under climate change scenarios . . . . .	76
4.4	Conclusion . . . . .	78
<b>5</b>	<b>An Extended Geometric Failure Rate Reduction model with covariates</b>	<b>79</b>
5.1	Introduction . . . . .	80
5.2	Geometric Failure Rate Reduction models without covariates . . . . .	81
5.3	Extended Geometric Failure Rate Reduction with covariates . . . . .	82
5.4	Estimation in the EGFRR model with covariates . . . . .	83
5.4.1	Estimation of the regression parameter $\beta$ . . . . .	84
5.4.2	Estimation of the parameters $\gamma$ and $\theta$ . . . . .	85
5.5	Large sample behaviour of estimators . . . . .	86
5.5.1	Asymptotic behaviour of $\hat{\Lambda}_{X_k^s}(\cdot)$ . . . . .	87
5.5.2	Asymptotic behaviour of $\hat{\beta}$ . . . . .	87
5.5.3	Asymptotic behaviour of $\hat{\zeta} = (\hat{\gamma}, \hat{\theta})$ . . . . .	89
5.6	Simulation study . . . . .	94
5.6.1	Simulation design . . . . .	94
5.6.2	MC simulation results for $\hat{\beta}$ . . . . .	96
5.6.3	MC simulation results for the estimators $(\hat{\gamma}, \hat{\theta})$ . . . . .	96
5.7	Conclusion and Perspectives . . . . .	98
	<b>Conclusion Générale</b>	<b>99</b>
	<b>Bibliographie</b>	<b>101</b>

---

# Introduction Générale

---

Une fois qu'un système est opérationnel, il est inévitable qu'il subisse une détérioration progressive au fil du temps, qu'il soit réparable ou non. Cette dégradation peut être influencée par des facteurs tels que l'usure et d'autres éléments comme les contraintes environnementales, dont les impacts sont de plus en plus accentués par le réchauffement climatique. Face à ces défis, les propriétaires et les gestionnaires de ces systèmes se trouvent confrontés à la nécessité d'optimiser la durée de vie de leurs équipements en fonction de leur fiabilité et leur efficacité opérationnelle. C'est dans ce cadre que la compréhension approfondie du processus de détérioration devient une priorité majeure. En identifiant avec précision les facteurs qui influencent la détérioration du système, il devient possible de modéliser et de prédire plus efficacement sa durée de vie. Cette approche, en prenant en compte les considérations environnementales, permet de déterminer la fréquence et la nature des interventions de maintenance nécessaires pour maintenir le système en bon état de fonctionnement et de sécurité.

Dans le domaine spécifique du Génie Civil où la durée attendue de service des structures est souvent plus longue que celle attendue, cette problématique est particulièrement présente surtout pour les structures en béton armé. Au cours de leur cycle de vie, ces structures subissent une détérioration. L'un des principaux contributeurs à cette détérioration est la corrosion des armatures, un phénomène électrochimique en deux phases (initiation et propagation) qui peut être déclenché par la carbonatation et/ou par la pénétration d'ions chlorures. La carbonatation, qui nous intéresse particulièrement ici, correspond à la diffusion du dioxyde de carbone ( $\text{CO}_2$ ) dans les pores du béton, déclenchant ainsi des réactions de dissolution/précipitation avec la portlandite, qui agit comme un tampon de pH. Lorsque la portlandite est épuisée, la solution interstitielle perd de sa basicité, conduisant à l'amorçage de la phase d'initiation de la corrosion.

La carbonatation, un phénomène inévitable pour les structures non immergées, est influencée par des paramètres intrinsèques au matériau, des conditions de traitement et de cure ainsi que des variables environnementales. Les approches de modélisation de la carbonatation se déclinent en trois catégories principales : numériques, empiriques et analytiques. Les ingénieurs ont généralement recours aux modèles analytiques pour évaluer le niveau de carbonatation, adoptant différentes approches de modélisation. Néanmoins, la majorité de ces modèles souffrent de limitations. Certains sont conçus spécifiquement pour un type particulier de béton, d'autres sont calibrés à partir de données issues de tests de carbonatation accélérée en laboratoire, et certains requièrent un grand nombre de paramètres d'entrée. Toutes ces caractéristiques entravent leur applicabilité et leur pertinence pour évaluer la fiabilité des structures en béton armé exposées à la carbonatation naturelle. En conséquence, de nombreux chercheurs se sont engagés dans l'élaboration de modèles simples et efficaces capables de surmonter ces limitations. La première partie de ce projet de recherche, développée en collaboration entre l'Institut de Mathématiques de Toulouse (IMT) et le Laboratoire Matériaux et Durabilité des

Constructions (LMDC) de Toulouse, s'inscrit dans cette optique. L'objectif principal est de développer un modèle statistique répondant aux critères de simplicité et de performance. Ce modèle vise à améliorer la compréhension et la gestion de la durabilité des structures en béton armé exposées à la carbonatation naturelle, permettant ainsi d'évaluer leur fiabilité dans le contexte actuel de réchauffement climatique. Pour ce faire, deux approches seront adoptées : la première se concentrera sur la modélisation de la profondeur de carbonatation, tandis que la seconde se basera sur la modélisation du temps de carbonatation (c'est-à-dire le temps nécessaire au processus de carbonatation pour atteindre les armatures). L'approche statistique devient de plus en plus utilisée dans l'étude de la carbonatation. Elle offre des outils simples et performants qui permettent d'analyser et comprendre le processus de carbonatation tout en identifiant les facteurs qui sont les plus influents.

D'autre part, une fois qu'une action de maintenance est réalisée, se pose une question cruciale : comment intégrer l'impact de cette maintenance dans la modélisation de la fiabilité du système, en vue de planifier de manière optimale les prochaines maintenances ? En effet, toutes les maintenances ne se valent pas. Lorsqu'un système rencontre une panne, différentes approches peuvent être adoptées. D'un côté, il est possible de simplement le rétablir dans l'état dans lequel il était avant la défaillance, c'est ce qu'on appelle une réparation « As Bad As Old » (ABAO). À l'opposé, il est également envisageable de le restaurer de manière à ce qu'il retrouve un état proche du neuf, voire même de le remplacer par un système neuf identique, ce qui est qualifié de réparation « As Good As New » (AGAN). Dans la réalité, la plupart des scénarios de maintenance se situent entre ces deux extrêmes, engendrant ainsi un effet intermédiaire. Cette situation est désignée sous le terme de « maintenance imparfaite ».

La modélisation de la maintenance imparfaite a attiré l'attention de nombreux chercheurs. Parmi les modèles les plus réputés de la littérature, on peut citer le modèle BP introduit par BORDES et MERCIER 1983, les modèles d'âge virtuel présentés par KIJIMA 1989, les modèles de réduction arithmétique (réduction arithmétique de l'intensité de défaillance (ARI) et réduction arithmétique de l'âge virtuel (ARA)) développés par DOYEN et GAUDOIN 2004, ainsi que le modèle de Processus Géométrique (GP) conçu par LAM 1998. Plus récemment, Dauxois et ses collaborateurs ont proposé (voir DAUXOIS et al. 2018) une généralisation du modèle GFRR (Geometric Failure Rate Reduction) de Finkelstein FINKELSTEIN 2008. Ce modèle étendu a été nommé « EGFRR » pour « Extended Geometric Failure Rate Reduction ».

Afin d'accroître encore davantage les performances des modèles de maintenance imparfaite, les chercheurs se sont attachés à intégrer l'impact des covariables dans ces modèles. Ces covariables fournissent des informations complémentaires précieuses pour l'analyse et la modélisation de la fiabilité des systèmes réparables. C'est dans cette optique que des travaux supplémentaires ont été entrepris au cours de cette thèse. Plus précisément, l'objectif de ces travaux était de prendre en compte l'effet des covariables dans le modèle EGFRR.

Le présent mémoire, structuré en cinq chapitres, aborde de manière approfondie les deux problématiques majeures exposées.

Le premier chapitre offre une revue bibliographique exhaustive des connaissances essentielles relatives aux bétons armés, au processus de corrosion induit par la carbonatation et aux méthodes statistiques.

Le deuxième chapitre, plus concis, se concentre sur les bases de données employées pour modéliser la profondeur de carbonatation et le temps de carbonatation, en se basant sur une source de données variée et riche.

Les trois derniers chapitres sont constitués d'articles soit déjà publiés, soit en cours de soumission à des revues scientifiques. Le chapitre 3 détaille les modèles de carbonatation élaborés à travers l'utilisation des GLMs (Generalized Linear Models) ainsi que des processus

de dégradation Gamma. Une analyse comparative de ces modèles avec le modèle de la *fib-Model Code 2010* CEB-FIP 2010 est également exposée.

Le chapitre 4 se focalise sur la modélisation du temps de carbonatation par le biais de modèles de régressions en survie. La procédure de modélisation est explicitée, aboutissant à la sélection du modèle le plus approprié et performant. De plus, ce modèle est employé pour évaluer la fiabilité de matériaux en béton armé sous diverses conditions climatiques.

Le chapitre 5 introduit le modèle EGFR (Extended Geometric Failure Rate Reduction) avec covariables, présentant également les estimateurs associés à ses paramètres. En complément, les résultats de l'étude asymptotique et numérique de ces estimateurs sont fournis dans ce même chapitre.

Enfin, ce mémoire se termine par une conclusion générale et une bibliographie complète.

# État de l'art

---

## 1.1 Introduction

Le béton présente une excellente résistance aux contraintes de compression, mais cette résistance diminue considérablement lorsqu'il est soumis à des forces de traction. D'après la norme EN 206-1, la résistance à la traction du béton est d'environ un dixième de sa résistance à la compression. Pour créer un matériau capable de supporter à la fois des forces de compression et de traction, on intègre des barres d'acier (armatures) dans le béton, formant ainsi le béton armé. Cependant, la pénétration d'agents agressifs tels que le dioxyde de carbone gazeux et/ou les ions chlorures dans le béton peut entraîner la corrosion des armatures, ce qui réduit la sécurité et la durabilité de la structure.

L'étude de la durabilité des structures en béton armé requiert une compréhension approfondie du phénomène de corrosion, ainsi que la modélisation des phénomènes déclencheurs tels que la carbonatation et la pénétration d'ions chlorure.

Ainsi, ce premier chapitre commencera par une brève description du matériau béton, suivi de l'explication du phénomène de corrosion. Ensuite, différentes approches existantes pour modéliser la durabilité du béton armé seront présentées, en mettant l'accent sur la carbonatation. Enfin, ce chapitre conclura en exposant les méthodes statistiques utilisées pour étudier la durabilité des matériaux en béton.

## 1.2 Le matériau béton

### 1.2.1 Composition minéralogique du béton

Le béton est le matériau le plus utilisé dans la construction. C'est un solide poreux résultant du mélange de trois éléments : du sable, des gravillons et du ciment. Le ciment, composant principal du béton, réagit avec l'eau de gâchage pour former la pâte de ciment qui durcit suite à une succession de réactions chimiques. Les propriétés du béton (mécaniques, chimiques, esthétiques et de durabilité) dépendent du type de ciment utilisé, des techniques de mise en place et de son environnement d'exposition.

Il existe plusieurs types de ciments, dont le plus couramment utilisé est le ciment Portland. Ce dernier est constitué à 95% de clinker et 5% de sulfate de calcium. Le clinker, constitué de silicates de calcium et d'aluminates, est issu de la cuisson à très haute température (1450 °C) d'un mélange de calcaire et d'argile.

Lorsque le ciment est mélangé à l'eau, un ensemble de réactions de dissolution-précipitation se produit, constituant l'hydratation du ciment (MAI-NHU 2013). Les principaux produits de l'hydratation, appelés hydrates, sont :

- les silicates de calcium hydratés (**C-S-H**), qui sont les principaux acteurs du développement des résistances mécaniques de la matrice cimentaire ;
- la portlandite (**CH**), responsable de la basicité de la solution interstitielle du béton ;
- l'étringite (**AFt**) et le monosulfoaluminate (**AFm**).

### 1.2.2 Microstructure de la pâte de ciment

En contact avec l'eau, les différents constituants du ciment se dissolvent d'abord, puis se précipitent pour former de petits grains (les hydrates) qui se superposent et s'agglomèrent. Les espaces entre ces grains, qui étaient initialement occupés par l'eau, constituent la porosité capillaire du béton. L'ensemble de ces pores forme le réseau capillaire, qui est très connecté aux premiers stades de la formation du béton (OLLIVIER et VICHOT 2008a). Au fil du temps, ce réseau devient de moins en moins connecté en raison de l'augmentation du volume des hydrates formés. La porosité capillaire, caractérisée par son volume et sa connectivité, joue un rôle crucial dans la résistance du béton. Le volume des pores augmente avec le rapport eau sur ciment (E/C). En effet, pour une quantité de ciment donnée, plus la quantité d'eau utilisée pour le mélange est importante, plus petit est le volume des hydrates formés, ce qui entraîne un espacement plus important entre eux.

### 1.2.3 Propriétés de transfert dans le béton

La saturation en eau des pores du béton ainsi que leur connectivité jouent un rôle important dans les propriétés de transfert du béton. En effet, l'eau présente dans les pores favorise la pénétration des espèces ioniques tandis qu'elle entrave la pénétration du gaz (OLLIVIER et VICHOT 2008b). Une fois coulé, le béton est exposé à l'intrusion d'agents agressifs au sein de la matrice cimentaire, sous forme de liquide ou de gaz.

La capacité d'un matériau poreux à être traversé par un fluide sous un gradient de pression est appelée perméabilité. La perméabilité d'un matériau, notée  $k$ , est définie par la relation de Darcy, qui exprime le débit  $Q$  d'un fluide de viscosité  $\mu$  traversant une épaisseur  $dx$  de matériau de section  $A$  sous une différence de pression  $dp$  :

$$Q = \frac{k}{\mu} A \frac{dp}{dx}. \quad (1.1)$$

La perméabilité d'un béton est relativement faible. Les travaux de KART et THOMPSON 1986 ainsi que ceux de DULLIEN 1979 montrent que la perméabilité du béton dépend de la structure de son réseau capillaire. Pour réduire la perméabilité du béton, il convient donc de diminuer la valeur du rapport E/C (eau sur ciment) et de soigner le processus de cure.

Le transfert des agents agressifs dans le béton se produit également par diffusion, que ce soit sous forme liquide ou gazeuse. En général, le processus de diffusion est caractérisé par le coefficient de diffusion  $D$ , exprimé en mètres carrés par seconde ( $m^2/s$ ), qui est défini par la première loi de Fick :

$$J_x = -D \frac{\partial c}{\partial x}, \quad (1.2)$$

où  $J_x$  est le flux du composant dans la direction  $x$  et  $\frac{\partial c}{\partial x}$  est le gradient de concentration dans cette direction.

Dans les matériaux poreux, les parois perturbent le transport des substances dans la direction du gradient de concentration (OLLIVIER et VICHOT 2008b). Ainsi, la diffusion dans un milieu

poreux est décrite en utilisant le flux effectif  $J_{e,x}$ , qui représente la quantité de composant traversant le matériau par unité de temps et de surface dans la direction  $x$  :

$$J_{e,x} = -D_e \frac{\partial c}{\partial x}. \quad (1.3)$$

Dans ce cas, le coefficient de diffusion effectif  $D_e$  dépend non seulement de l'espèce diffusante, mais aussi de la structure poreuse du matériau.

En résumé, le béton est un matériau poreux qui peut être pénétré par des agents agressifs. La pénétration de ces agents se fait par diffusion conformément à la loi de diffusion de Fick ou par infiltration conformément à la loi de Darcy. La capacité du béton à permettre la pénétration des agents agressifs dépend de ses propriétés de transfert (principalement sa diffusivité, sa perméabilité et sa porosité), qui sont liées à la valeur du rapport E/C, au type de ciment utilisé et aux conditions de cure.

## 1.3 La corrosion des armatures du béton armé

### 1.3.1 Généralité sur le mécanisme de corrosion

La corrosion est un processus électrochimique qui se produit entre le métal et son environnement. Selon TUUTTI 1992a, la corrosion se déroule en deux phases : la phase d'amorçage ou d'initiation, et la phase de propagation.

Au cours de la première phase de corrosion, également appelée phase dormante, la stabilité du système formé par le métal immergé dans la matrice cimentaire se détériore, créant ainsi des conditions favorables à la corrosion. En effet, le métal est en contact avec la solution interstitielle du béton, dont le pH est compris entre 12 et 13. Une fine couche protectrice d'oxydes, appelée film passif, se forme alors à la surface de l'acier, protégeant ainsi les armatures (OLLIVIER et VICHOT 2008c). Cependant, la pénétration progressive des agents agressifs conduira à la dépassivation des armatures, favorisant ainsi la corrosion.

Lorsque les agents agressifs atteignent les armatures en quantité suffisante, la phase de propagation (deuxième phase) débute en présence d'oxygène et d'humidité adéquats. À ce stade, des produits de corrosion se forment au niveau des armatures, donnant lieu à la formation d'oxydes et d'hydroxydes grâce à une série de réactions électrochimiques. Ces produits de corrosion (oxydes et hydroxydes), dont les volumes sont bien supérieurs à celui de l'acier sain, génèrent des contraintes dans l'enrobage, créant ainsi des fissures à terme. Ces fissures favorisent la pénétration continue des agents agressifs, augmentant ainsi la cinétique de la corrosion (OLLIVIER et VICHOT 2008c).

La cinétique de la corrosion dépend donc non seulement des propriétés du matériau (rapport E/C, type de ciment, résistivité électrique, etc.), mais également des caractéristiques de l'environnement (humidité relative, température, pluviométrie, etc.). La pénétration de dioxyde de carbone ( $\text{CO}_2$ ) et d'ions chlorures sont les deux principaux phénomènes inducteurs de la corrosion.

### 1.3.2 Corrosion induite par la carbonatation

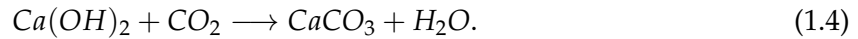
#### 1.3.2.1 Le mécanisme

La carbonatation est un phénomène chimique causé par l'intrusion du dioxyde de carbone dans la matrice cimentaire. Le  $\text{CO}_2$  présent dans l'air sous forme gazeuse se dissout dans les pores du béton, puis réagit avec les hydrates présents dans la solution interstitielle, en particulier



avec la portlandite qui est responsable de la basicité de la solution. La consommation de portlandite entraîne une diminution du pH de la solution interstitielle, ce qui, en combinaison avec la présence d'oxygène, active la phase d'initiation de la corrosion.

Le mécanisme de la carbonatation, dont les étapes sont résumées dans la Figure ??, est principalement gouverné par la carbonatation de la portlandite, décrite dans la littérature par l'équation (1.4) suivante :



La carbonatation a un impact positif sur la microstructure du béton. En effet, la formation de calcite par la carbonatation des hydrates contribue à augmenter la phase solide de la matrice cimentaire, réduisant ainsi la porosité du béton. De plus, l'eau libérée par la carbonatation de la portlandite accroît la quantité d'eau présente dans les pores, ce qui à long terme ralentit la cinétique de la carbonatation.

En fin de compte, la principale conséquence néfaste de la carbonatation est la diminution du pH de la solution interstitielle due à la réduction de la quantité de portlandite, ce qui déclenche la corrosion des armatures.

### 1.3.2.2 Les facteurs de la carbonatation

La cinétique de la carbonatation est influencée par de nombreux facteurs, qui peuvent être regroupés en trois catégories : les facteurs liés au matériau, les facteurs de mise en œuvre et de cure, ainsi que les facteurs environnementaux.

#### a) Les facteurs du matériau

Comme mentionné à la section 1.2.2, la porosité du béton, qui dépend du rapport E/C, influe sur le transfert des agents agressifs dans le matériau. En effet, un béton présentant un rapport E/C élevé possède un réseau capillaire plus développé, ce qui favorise une diffusion rapide du CO<sub>2</sub> à travers les pores.

De plus, le dosage en ciment est un facteur influent sur la carbonatation du béton. En effet, une augmentation du dosage en ciment réduit la porosité du matériau et, par conséquent, la diffusion du CO<sub>2</sub> est ralentie (VENUAT et ALEXANDRE 1968 ; ZHANG et F. GLASSER 2005).

Par ailleurs, le rapport E/C et le dosage en ciment sont deux caractéristiques directement liées à la résistance mécanique du béton durci, ce qui est une information plus accessible pour les ingénieurs en génie civil. De nombreuses études indiquent que la profondeur de carbonatation est moindre dans les bétons à résistance à la compression élevée (OLLIVIER et VICHOT 2008c).

#### b) Facteurs liés à la mise en œuvre et à la cure

La plupart des bétons sont constitués d'un mélange de sable, de gravier et de pâte de ciment. Une fois placé dans le moule, ce mélange est soumis à des vibrations pour assurer un meilleur agencement de ces grains de tailles différentes. Ces vibrations contribuent à obtenir un béton avec un réseau poreux moins ouvert, réduisant ainsi sa perméabilité (MAI-NHU 2013).

La cure fait référence à un ensemble de mesures mises en œuvre pour protéger le béton coulé dans des conditions climatiques optimales, de manière à ce que les espaces entre les grains

de la pâte fraîche soient autant que possible remplis par les produits d'hydratation (HYVERT 2009). Elle contribue à réduire la porosité du béton, ce qui le rend moins perméable à la carbonatation.

### c) Facteurs liés à l'environnement d'exposition

Les caractéristiques du lieu où la construction est exposée jouent un rôle considérable dans la cinétique de la carbonatation. Parmi ces caractéristiques, on peut citer l'humidité relative, la pluviométrie, la température et la pression atmosphérique en  $\text{CO}_2$ .

L'humidité relative, qui détermine la teneur en eau du béton, exerce une influence significative sur la carbonatation. Le  $\text{CO}_2$  peut pénétrer dans la matrice cimentaire par l'intermédiaire de la phase liquide ou de la phase gazeuse. Avec une augmentation de l'humidité relative, la proportion de pores saturés augmente. Par conséquent, la pénétration du  $\text{CO}_2$  se produit principalement à travers la phase liquide, dont le coefficient de diffusion est 10 000 fois plus faible que celui de la phase gazeuse ( $10^{-12}$  contre  $10^{-10}$ ) (MAI-NHU 2013). En outre, une faible humidité relative ralentit la carbonatation en raison du manque d'eau nécessaire à l'hydratation des hydrates.

La pluviométrie constitue un autre facteur étroitement lié à l'humidité relative. En l'absence de protection contre la pluie, les pores du béton sont complètement saturés pendant les périodes de précipitation. Par conséquent, lorsque la porosité présente un taux de saturation supérieure à 70%, la pénétration du dioxyde de carbone ralentit.

La concentration en  $\text{CO}_2$  dans l'environnement d'exposition est un facteur crucial de la carbonatation. Logiquement, plus cette concentration est élevée, plus le niveau de carbonatation du matériau est important. Les conclusions des études menées par NISCHER 2013 et HYVERT 2009 viennent confirmer cette hypothèse.

## 1.3.3 Corrosion induite par la pénétration des ions chlorures

### 1.3.3.1 Le mécanisme

Les ions chlorures pénètrent dans la matrice cimentaire du béton et contribuent à la dépassivation, ce qui engendre la corrosion des armatures. La manière dont les ions chlorures pénètrent dans le béton dépend de son exposition à l'environnement marin. La Figure 1.1, extraite de FABRICE 2008, énumère les divers types d'expositions possibles.

Les ions chlorures accèdent au cœur du béton par diffusion sous gradient de concentration dans le cas des parties immergées. Pour d'autres types d'exposition, les ions chlorures pénètrent par absorption capillaire, puis migrent avec le liquide par convection et diffusion (TAHLAITI 2010).

Une fois dans les pores du béton, une partie des ions chlorures est absorbée par les C-S-H ou réagit avec les aluminates et les aluminoferrites pour former des sels de Friedel (OLLIVIER et VICHOT 2008c). Ces chlorures sont appelés chlorures fixes. La partie restante des ions chlorures, présente sous forme ionique dans la solution interstitielle, constitue les chlorures dits "libres". Ce sont ces derniers qui jouent un rôle actif dans la corrosion des armatures. En effet, lorsque la concentration de la solution interstitielle en ions chlorures atteint une certaine valeur critique, notée  $C_{crit}$ , la couche passive autour des armatures se dissout. Des ions chlorures migrent à travers cette couche, modifiant sa morphologie en formant des ions chlorure de fer. Ces ions chlorure de fer consomment les ions hydroxydes et entraînent ainsi une baisse du pH de la solution autour des armatures. Ces différentes réactions chimiques sont résumées à la figure 1.2.

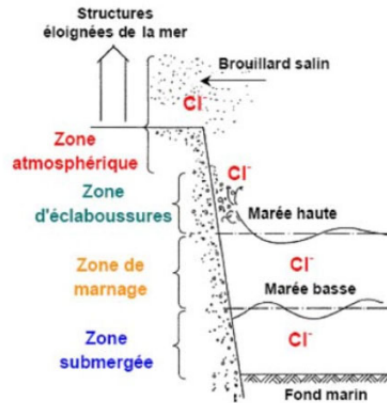


Figure 1.1 – Les différents types d'expositions d'un béton à l'environnement marin (FABRICE 2008)

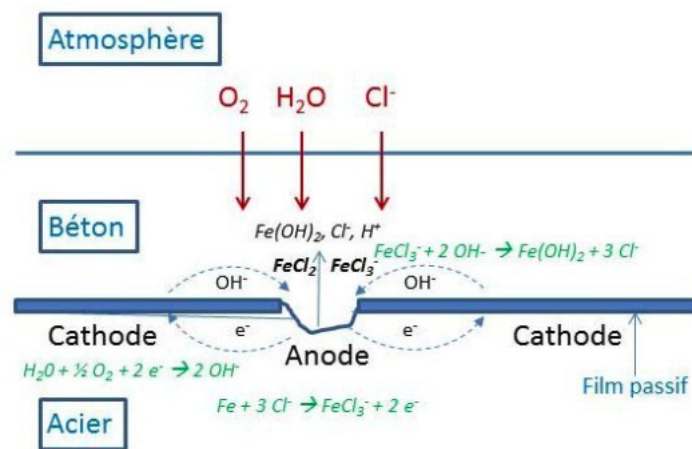


Figure 1.2 – Mécanisme de corrosion dans le cas d'une attaque par les ions chlorure (MAI-NHU 2013)

### 1.3.3.2 Les facteurs influençant l'attaque par les ions chlorures

Comme pour la carbonatation, la pénétration d'ions chlorures dans la matrice cimentaire est influencée par les caractéristiques du matériau, de l'environnement d'exposition et des facteurs technologiques tels que la mise en œuvre et la cure.

La migration des ions dépend des propriétés de transfert du béton, notamment de sa porosité. Ainsi, le rapport E/C, lié à la porosité, joue un rôle important. Un béton avec une grande valeur du rapport E/C possède un réseau poreux très connecté, favorisant une migration plus rapide des ions chlorures. D'autre part, le type de ciment utilisé influence la capacité du béton à fixer les ions chlorures. Les bétons à base de ciments riches en aluminates présentent une grande capacité de fixation des ions chlorure, réduisant ainsi la quantité de chlorures libres et ralentissant le processus de corrosion (HYVERT 2009).

Parmi les paramètres environnementaux, la concentration en chlorures à la surface du béton, la température et l'humidité relative jouent un rôle dans la pénétration des ions chlorures (HYVERT 2009).

Les propriétés de transfert du béton, telles que la perméabilité et la diffusivité, sont également liées à sa mise en œuvre et à sa cure. Comme mentionné précédemment, la cure et la mise en œuvre du béton modifient sa porosité et la quantité d'hydrates, impactant respectivement la pénétration des chlorures et la capacité du béton à les fixer.

## 1.4 Durabilité des bétons armés face à la corrosion des armatures

### 1.4.1 Modélisation de la carbonatation

La carbonatation représente la principale source de corrosion dans les structures en béton non immergées. Pour les gestionnaires d'infrastructures, avoir une connaissance précise du niveau de carbonatation tout au long de la durée d'exploitation de la structure est crucial. Cela permet de mettre en place des stratégies de maintenance préventive et corrective de manière efficace. La littérature propose divers modèles de carbonatation, qui se distinguent par leurs méthodes de modélisation et le type ainsi que le nombre de paramètres pris en compte. Tous ces modèles trouvent leur origine dans la première loi de Fick (Equation (1.3)). En fonction de leur approche de modélisation, on peut classer les modèles de carbonatation en trois grandes catégories : les modèles numériques, les modèles empiriques et les modèles analytiques.

#### 1.4.1.1 Les modèles numériques

Les modèles numériques visent à fournir une description précise du processus de carbonatation en résolvant, par des méthodes telles que les éléments finis ou les volumes finis, des systèmes d'équations qui représentent les transformations physico-chimiques se produisant lors de la carbonatation.

L'un des premiers modèles numériques a été développé par SAETTA et al. 1993. Ce modèle 1D décrit la cinétique globale de la carbonatation de la portlandite, en tenant compte des transferts hydriques et thermiques à l'intérieur du béton. Cependant, ce modèle présente des limites car il ne décompose pas suffisamment les mécanismes de la carbonatation, ce qui l'empêche de prédire l'état réel du matériau. Plus tard, STEFFENS et al. 2002 ont repris ce modèle en introduisant un paramètre reliant le degré de saturation des pores à l'humidité relative.

Le modèle de BARRY et SELIER 1990 est largement mentionné dans la littérature. Il repose sur les équations de conservation de la masse de l'eau, du CO<sub>2</sub> et du calcium. Outre les phénomènes chimiques, il intègre également les aspects physiques tels que la variation de la porosité due à la carbonatation. Néanmoins, ce modèle est limité par sa calibration sur des données expérimentales spécifiques et certains paramètres d'entrée qui peuvent être difficiles à obtenir.

D'autres modèles vont au-delà en incorporant des phénomènes tels que la présence de microfissures dans le béton. C'est notamment le cas du modèle proposé par ISGOR et RAZAQPUR 2004. La présence de fissures facilite la diffusion du CO<sub>2</sub> dans le béton, et en tenant compte de ce paramètre, ces modèles offrent une meilleure estimation du niveau de carbonatation.

#### 1.4.1.2 Les modèles empiriques

En supposant que le CO<sub>2</sub> qui pénètre dans les pores du béton réagit immédiatement avec les hydrates présents, la profondeur de carbonatation  $X_c$  à l'âge  $t$  peut être exprimée selon la relation suivante (OLLIVIER et VICHOT 2008c) :

$$X_c = At^B, \quad (1.5)$$

où  $X_c$  est exprimée en millimètres [mm] et  $t$  en années. Les paramètres réels  $A$  et  $B$  sont des coefficients d'ajustement. Les auteurs de modèles empiriques utilisent des données provenant de tests de carbonatation accélérée ou naturelle pour estimer ces paramètres  $A$  et  $B$ .

Cependant, ces modèles empiriques ont leurs limites. Ils sont principalement calibrés en utilisant des données de carbonatation et donc leur utilité pour prédire le niveau de carbonatation

est limitée. De plus, même lorsqu'ils sont calibrés avec des données de carbonatation naturelle, ils ne prennent généralement pas en compte les cycles d'humidification/séchage, qui ont un impact significatif sur la profondeur de carbonatation. Parmi les modèles empiriques, on peut citer les travaux tels que ALEXANDER, MACKECHNIE et AM 2007; BAMFORTH 2004; BARY et SEILLIER 2003; GREVE-DIERFELD et GEHLEN 2014; MORINAGA 1990; PARROT 2009; SILVA et al. 2014; THIERY 2006 ou encore RIZVI 2017.

### 1.4.1.3 Les modèles analytiques

En analysant les profondeurs de carbonatation de bétons exposés à la fois à la carbonatation accélérée et à la carbonatation naturelle, les chercheurs ont observé une tendance selon laquelle la profondeur de carbonatation est proportionnelle à la racine carrée du temps. En conséquence, les modèles analytiques ont été développés en utilisant la solution analytique de la première loi de Fick, qui est formulée comme suit :

$$X_c = K\sqrt{t}, \tag{1.6}$$

où  $X_c$  est exprimée en millimètres et  $t$  en années. Le paramètre  $K$ , appelé coefficient de carbonatation (exprimé en  $\text{mm}/\text{année}^{0.5}$ ), est la principale variable d'intérêt dans les modèles analytiques. Les chercheurs ont utilisé des méthodes mathématiques pour déterminer une expression pour ce coefficient de carbonatation qui prend en compte les effets des facteurs de carbonatation.

La littérature scientifique propose de nombreux modèles analytiques qui se distinguent les uns des autres par les méthodes mathématiques qu'ils utilisent et par la forme de l'expression de  $K$ . On peut trouver des modèles statistiques, des modèles probabilistes et d'autres basés sur des approches plus modernes comme les arbres de régression et les réseaux de neurones. La Table 1.1 présente une liste non exhaustive de ces deux dernières catégories de modèles analytiques.

Références	Facteurs pris en compte			Conditions de carbonatation
	Matériau	Technologie	Environnement	
Schiessl P. 1976	Propriétés chimiques		$P_{CO_2}$	Accélérée
Tuuti TUUTTI 1982	Propriétés chimiques		$P_{CO_2}$	Naturelle
Papadakis et al. V. G. PΑΡΑΔΑΚΙΣ et al. 1991	Propriétés chimiques		$P_{CO_2}$	Naturelle et accélérée
Jiang et al. J. L. et al. 2000	Propriétés chimiques		$P_{CO_2}$	Accélérée
Bouquet G.C. 2004	Propriétés chimiques		$P_{CO_2}$ , Pluviométrie	Naturelle et accélérée
Sisomphon et al. K. et F. L. 2007	Propriétés chimiques			Accélérée
Hyvert HYVERT 2009	Propriétés chimiques et mécaniques	Temps de cure	$P_{CO_2}$ , RH, Pluviométrie	Accélérée
fib CEB-FIP 2010	Propriétés chimiques		$P_{CO_2}$ , RH, Pluviométrie	Naturelle et accélérée
Salvoldi et al. B.G. et al. 2015	Propriétés chimiques		$P_{CO_2}$ , RH, Pluviométrie	Accélérée
Van-Loc Ta et al. V.L TA et al. 2016	Propriétés chimiques et mécaniques	Temps de cure	$P_{CO_2}$ , RH, Pluviométrie, Température	Accélérée et naturelle
Ekolu EKOLU 2020	Propriétés chimiques et mécaniques		$P_{CO_2}$ , RH, Pluviométrie	Naturelle
Liu et al. LIU et al. 2008	Propriétés chimiques et mécaniques		$P_{CO_2}$ , RH, Pluviométrie	Naturelle
Taffese et al. TAFSESE, Esko SISTONEN et PUTTONEN 2015	Propriétés chimiques et mécaniques		$P_{CO_2}$ , RH, Pluviométrie	Naturelle

Table 1.1 – Synthèse des modèles analytiques de carbonatation SCHMITT 2019

En ce qui concerne les modèles basés sur l'approche statistique, plusieurs travaux importants peuvent être mentionnés.

Le modèle de ATIŞ 2004 est l'un des premiers modèles statistiques de carbonatation. En utilisant la régression linéaire sur des données de carbonatation accélérée, il propose une expression pour  $K$  impliquant deux facteurs liés au matériau : la porosité et la résistance mécanique du béton à 28 jours.

Monteiro et ses co-auteurs ont également utilisé la régression linéaire (dans MONTEIRO et al. 2012) pour étudier l'effet de facteurs tels que la résistance à la compression, la classe d'exposition et la présence d'une couche de protection à la surface du béton (peinture) sur la valeur du coefficient de carbonatation pour des structures exposées à la carbonatation naturelle. Ils proposent une valeur moyenne de  $K = 3.76\text{mm}/\text{année}^{0.5}$  pour prédire la profondeur de carbonatation pour un matériau non peint exposé dans un environnement avec une humidité modérée. Cependant, ils n'ont pas fourni une formule précise pour le calcul de  $K$ , ce qui rend ce modèle empirique.

Le modèle présenté par P.HILL et al. 2015 est l'un des modèles statistiques de carbonatation les plus cités dans la littérature. Les auteurs ont appliqué la régression linéaire multiple à une base de données comprenant des mesures de carbonatation accélérée et naturelle. Ils ont proposé deux expressions pour  $\log(K)$  : la première tient compte des effets des facteurs du matériau, des conditions d'exposition et de l'âge du béton ; la seconde prend en compte les effets des facteurs du matériau, notamment la résistance mécanique, et des conditions d'exposition. Cependant, les auteurs n'ont pas précisé clairement dans quelles conditions utiliser chaque expression. Ils ont également noté que l'hypothèse de normalité pour leur variable cible (c'est-à-dire la transformation logarithmique de  $K$ ) n'était pas vérifiée par leurs données.

En plus de ces modèles, d'autres peuvent être mentionnés, comme ceux de KIM et LEE 1999 ainsi que celui de LEEMANN et al. 2015, qui sont également basés sur une approche statistique. Cependant, ces analyses ont été principalement menées sur des données provenant de la carbonatation accélérée.

#### 1.4.2 Modélisation de la pénétration des ions chlorure

Les modèles de pénétration d'ions chlorures se basent sur la résolution des lois de Fick. En supposant que la concentration en ions chlorures ainsi que le coefficient de diffusion restent constants au fil du temps, la solution analytique pour la concentration en ions chlorures libres à une distance  $x$  en mètre [m] et un temps  $t$  en seconde [s] dans un milieu semi-infini est exprimée par l'équation suivante :

$$C(x, t) = C_0 \left( 1 - \operatorname{erf} \left( \frac{x}{2\sqrt{D_a t}} \right) \right), \quad (1.7)$$

où :

- $C(x, t)$  représente la concentration en ions chlorures à la distance  $x$  et au temps  $t$  [mol/m<sup>3</sup>].
- $C_0$  est la concentration en ions chlorures à la surface du béton [mol/m<sup>3</sup>].
- $D_a$  correspond au coefficient de diffusion apparent des ions chlorures [m<sup>2</sup>/s].
- $\operatorname{erf}(\cdot)$  désigne la fonction d'erreur de Gauss, définie par  $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-s^2} ds$ .

Cette équation décrit comment la concentration en ions chlorures évolue en fonction de la distance à partir de la surface et du temps écoulé. Notez que dans des situations réelles, les valeurs de  $C_0$  et  $D_a$  peuvent varier en fonction de facteurs spécifiques tels que les conditions environnementales et les propriétés du béton.

Cependant, le modèle tel que défini à l'équation (1.7) ne tient pas compte de certains phénomènes cruciaux, tels que l'interaction entre les ions chlorures et les autres espèces ioniques présentes dans la solution interstitielle, la porosité du béton ainsi que sa structure interne de pores, et les conditions extérieures comme l'humidité et la température. Ces facteurs jouent en effet un rôle significatif dans la pénétration des ions chlorures dans le béton. En conséquence,

de nombreuses recherches ont été entreprises pour enrichir le modèle (1.7) en intégrant les effets de ces éléments importants.

Les modèles numériques de pénétration de chlorures sont les plus nombreux. Bien que complexes, ces modèles sont basés sur la résolution numérique des équations de transfert des agents agressifs dans le béton, avec l'utilisation d'expressions empiriques pour les paramètres physiques tels que les concentrations molaires des espèces ioniques dans la solution interstitielle, les coefficients de diffusion, etc. Dans cette catégorie, on peut citer le modèle ClinConc de TANG et NILSSON 1996, ainsi que les travaux de KONG et al. 2002, de JOHANNESSON 2003 et de MEIJERS et al. 2005.

En s'inspirant de l'expression du modèle (1.7), certains auteurs adoptent une approche probabiliste pour développer des modèles qui intègrent l'effet des facteurs environnementaux et des caractéristiques du matériau dans l'expression du paramètre  $D_a$ . Les lois de distribution des paramètres d'entrée de ces modèles sont définies à la suite de nombreuses études expérimentales. Des modèles tels que le modèle Duracrete (voir par exemple CEB-FIP 2010) ou encore le modèle de PETRE-LAZAR 2000 utilisent cette approche. Cependant, ces modèles supposent généralement que le coefficient de diffusion apparent et/ou la concentration de chlorure à la surface restent constants dans le temps. Dans certains modèles, comme le modèle de LAY 2003, le coefficient de diffusion apparent  $D_a$  est exprimé en fonction du temps.

## 1.5 Méthodes statistiques pour la modélisation de la durabilité d'un matériau

Lorsque l'objectif d'une analyse est de proposer une formule analytique mettant en relation le niveau de dégradation d'un composant ou d'un système et les facteurs de dégradation (internes ou externes), certains outils de la modélisation statistique peuvent s'avérer très utiles. Selon la problématique considérée et la nature des facteurs, on peut citer les outils tels que : les modèles linéaires généralisés, les modèles de dégradation et les modèles de durée de vie. Cette partie sera consacrée à la description de ces outils.

### 1.5.1 Les Modèles Linéaires Généralisés

#### 1.5.1.1 Présentation générale

Le modèle de régression linéaire (simple ou multiple) est un outil simple et efficace de la modélisation statistique qui permet de décrire le lien entre une variable quantitative d'intérêt et une ou plusieurs variables explicatives (quantitatives et/ou qualitatives).

Les domaines d'application sont divers. Il peut s'agir, par exemple, d'expliquer : le prix d'un bien immobilier en fonction de sa superficie et de sa localisation ; le salaire d'un employé en fonction de certaines caractéristiques socio-démographiques ; ou encore, dans notre cas ici, le niveau de carbonatation d'un béton en fonction des facteurs de carbonatation.

Désignons par  $Y$  la variable aléatoire réelle que l'on veut expliquer (on l'appelle aussi variable à expliquer ou variable réponse) et  $\mathbf{X} = (X^1, \dots, X^p)$  le vecteur des  $p$  variables explicatives (aussi appelées prédicteurs ou covariables). Le modèle linéaire suppose que  $Y$  est la somme d'une combinaison linéaire des variables explicatives et d'un terme d'erreur. En d'autres termes :

$$Y = \beta_0 + \sum_{j=1}^p \beta_j X^j + e, \quad (1.8)$$

où :

- les paramètres inconnus  $\beta_j, j = 0, \dots, p$  sont appelés coefficients de régression et sont à estimer.
- $e$  est une v.a.r représentant le terme d'erreur; c'est une variable non observée telle que

$$\mathbb{E}(e) = 0 \text{ et } \text{Var}(e) = \sigma^2 > 0, \text{ avec } \sigma^2 \text{ inconnu et à estimer.}$$

Lorsqu'on suppose que le terme d'erreur  $e$  est une variable aléatoire gaussienne (c'est-à-dire  $e \rightsquigarrow \mathcal{N}(0, \sigma^2)$ ), le modèle obtenu est appelé le modèle linéaire gaussien. C'est sous cette forme que les modèles linéaires sont les plus connus.

Si on désigne par  $\mathbf{X} = (1, X^1, \dots, X^p)$  le vecteur des variables explicatives (avec  $X^0 = 1$ , la variable qui prend la valeur 1 pour toutes les observations), l'équation (1.8) s'écrit :

$$Y = \sum_{j=0}^p \beta_j X^j + e, \quad (1.9)$$

Dans ce cas, les conditions imposées au terme d'erreur impliquent que :

$$\mathbb{E}(Y | \mathbf{X}) = \sum_{j=0}^p \beta_j X^j \text{ et } \text{Var}(Y | \mathbf{X}) = \sigma^2.$$

Autrement dit, en moyenne  $Y$  s'écrit comme combinaison linéaire des  $X^j, j = 0, \dots, p$ . Ainsi, le lien entre la variable à expliquer et les variables explicatives est de nature linéaire. Cependant, dans la réalité, le lien entre la variable d'intérêt et les prédicteurs ne respecte pas toujours l'hypothèse de linéarité. C'est le cas, par exemple, du lien entre la profondeur de carbonatation  $X_c$  et le temps  $t$  donné par l'équation (1.6).

Ainsi, le modèle linéaire généralisé, également connu sous le nom de Generalized Linear Model (GLM) en anglais, a été introduit par J. A. NELDER et WEDDERBURN 1972 dans le but de surmonter les hypothèses contraignantes du modèle linéaire gaussien. Ces hypothèses incluent la linéarité de la relation entre  $Y$  et les  $X^j, j = 1, \dots, p$ , ainsi que la distribution normale du terme d'erreur. Le modèle GLM suppose que la variable  $Y$  suit une distribution de probabilité  $f_Y$ , et que la transformation de l'espérance conditionnelle de  $Y$  en fonction des covariables est une combinaison linéaire de ces dernières. En d'autres termes, le modèle GLM repose sur les hypothèses suivantes :

$$\begin{cases} Y \sim f_Y \\ g(\mu) = \sum_{j=1}^p \beta_j X^j \Leftrightarrow \mu = g^{-1} \left( \sum_{j=1}^p \beta_j X^j \right), \end{cases} \quad (1.10)$$

où

- $\mu = \mathbb{E}[Y | \mathbf{X}]$  l'espérance de  $Y$  conditionnellement à  $X$ ;
- $f_Y$  est la fonction de densité de  $Y$ ; elle est supposée connue et appartenir à une famille exponentielle des distributions de probabilité.
- $g$  est une fonction réelle différentiable et strictement monotone appelée *fonction de lien*; elle est supposée connue.
- $\beta = (\beta_1, \dots, \beta_p)$  est le vecteur des coefficients de régression; c'est un vecteur inconnu qui est à estimer.

Dans la formulation ci-dessus, on remarque que si  $f_Y$  est la densité d'une loi normale et que  $g$  est la fonction identité, on retrouve le modèle linéaire classique.

De plus, il convient de rappeler que la distribution d'une variable  $Y$  est qualifiée de "famille exponentielle" lorsque sa fonction de densité prend la forme suivante :



$$f_Y(y, \theta, \varphi) = \exp \left( \frac{y\theta - b(\theta)}{a(\varphi)} + c(y, \varphi) \right), \quad (1.11)$$

où  $a(\cdot)$ ,  $b(\cdot)$  et  $c(\cdot)$  sont des fonctions spécifiques connues (déterministes). Les paramètres  $\varphi$  (paramètre de dispersion) et  $\theta$  (paramètre naturel) sont inconnus. Si l'on prend l'exemple de la densité de la loi normale :

$$f_Y(y, \theta, \varphi) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left( -\frac{(y - \mu)^2}{2\sigma^2} \right) = \exp \left( \frac{y\mu - \mu^2/2}{\sigma^2} - \frac{1}{2} \left( \frac{y^2}{\sigma^2} + \log(2\pi\sigma^2) \right) \right).$$

Dans ce cas,  $\theta = \mu$ ,  $\varphi = \sigma^2$ , et nous avons :

- $a(\varphi) = \sigma^2$ ;
- $b(\theta) = \mu^2/2$ ;
- $c(y, \varphi) = -\frac{1}{2}(y^2/\sigma^2 + \log(2\pi\sigma^2))$ .

L'ajustement d'un modèle linéaire généralisé à des données s'effectue en trois grandes étapes : 1) identification la loi de distribution la plus appropriée; 2) estimer les paramètres du modèle; 3) faire une sélection de modèle. Dans la suite, nous présenterons les outils permettant d'effectuer ces différentes étapes.

### 1.5.1.2 Identification de la distribution et de la fonction de lien

La première étape de l'ajustement d'un modèle GLM consiste à identifier la loi de distribution la plus adaptée aux données. De plus, il faudra identifier la fonction de lien  $g(\cdot)$  qui permet un meilleur ajustement aux données.

L'identification de la loi de distribution se fait par le biais d'analyses graphiques et/ou de tests d'adéquation de lois. Supposons que l'on dispose d'un échantillon  $Y_1, \dots, Y_n$  de taille  $n$  formé de répliques indépendantes et identiquement distribuées (i.i.d) de  $Y$ . On estime la fonction de répartition empirique  $\hat{F}_Y(\cdot)$  qui est donnée par :

$$\hat{F}_Y(y) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}(Y_i \leq y), \quad (1.12)$$

où  $\mathbb{I}(\cdot)$  est la fonction indicatrice classique.

L'analyse graphique consiste à identifier visuellement la loi de la famille exponentielle dont la courbe de la fonction de répartition est la plus proche de celle de  $\hat{F}_Y(\cdot)$ .

Le test d'adéquation de Kolmogorov-Smirnov (voir par exemple BAGDONAVIČIUS et al. 2011) est le test le plus couramment utilisé pour identifier la loi dont la fonction de répartition est proche de  $\hat{F}_Y(\cdot)$ . Si  $F_0$  est la fonction de répartition d'une loi connue, l'idée du test de Kolmogorov-Smirnov est de comparer les valeurs  $\hat{F}_Y(y_{(i)})$  et  $F_0(y_{(i)})$ , où  $y_{(i)}$  est la statistique d'ordre de l'échantillon observé  $(y_1, \dots, y_n)$ . L'hypothèse nulle de ce test est :

$$\mathcal{H}_0 : \text{la loi de l'échantillon a pour fonction de répartition } F_0(\cdot)$$

L'adéquation de la fonction de répartition empirique à la fonction  $F_0(\cdot)$  est mesurée par la distance de Kolmogorov-Smirnov, qui correspond à la distance de la norme uniforme entre les fonctions  $F_0(\cdot)$  et  $\hat{F}_Y(\cdot)$ . En d'autres termes, la statistique de test de ce test est donnée par :

$$\begin{aligned}
 D_{KS}(F_0, \hat{F}_Y) &= \sup_{y \in \mathbb{R}} \left| F_0(y) - \hat{F}_Y(y) \right| \\
 &= \max_{i=1, \dots, n} \left\{ \left| F_0(y_{(i)}) - \hat{F}_Y(y_{(i)}) \right| ; \left| F_0(y_{(i)}) - \hat{F}_Y(y_{(i-1)}) \right| \right\} \\
 &= \max_{i=1, \dots, n} \left\{ \left| F_0(y_{(i)}) - \frac{i}{n} \right| ; \left| F_0(y_{(i)}) - \frac{i-1}{n} \right| \right\}
 \end{aligned}$$

où  $y_{(0)} = -\infty$ .

Sous l'hypothèse  $H_0$ , la distribution de la statistique  $D_{KS}(F_0, \hat{F}_Y)$  ne dépend pas de  $F_0$  et a été tabulée. De plus, la proposition 5.3, issue de KOLMOGOROV 1933, démontre que  $D_{KS}(F_0, \hat{F}_Y)$  converge asymptotiquement vers une loi qui ne dépend pas de  $F_0$ .

**Proposition 1.1** : Sous l'hypothèse  $H_0$ , on a, pour tout  $t > 0$  :

$$\lim_{n \rightarrow \infty} \mathbb{P}_{H_0}[\sqrt{n}D_{KS}(F_0, \hat{F}_Y) \leq t] = 1 - 2 \sum_{k=1}^{+\infty} (-1)^{k-1} \exp(-2k^2 t^2).$$

Ainsi, nous rejetons l'hypothèse  $H_0$  au niveau  $\alpha \in ]0, 1[$  lorsque  $D_{KS}(F_0, \hat{F}_Y) \geq d_{n, 1-\alpha}$ , où  $d_{n, 1-\alpha}$  est choisi de manière à satisfaire l'inégalité

$$\mathbb{P}_{H_0} \left[ \sqrt{n}D_{KS}(F_0, \hat{F}_Y) \geq d_{n, 1-\alpha} \right] \leq \alpha.$$

Une fois la loi  $F_0$  identifiée, on recherche la forme la plus appropriée de la fonction de lien (c'est-à-dire la fonction  $g(\cdot)$ ). La qualité d'ajustement d'un modèle de régression peut être mesurée à travers son erreur de prédiction. Il existe plusieurs métriques qui quantifient l'erreur de prédiction d'un modèle. Les plus connues sont : l'erreur absolue (MAE, pour Mean Absolute Error), l'erreur quadratique moyenne (MSE, pour Mean Squared Error) et l'erreur quadratique moyenne enracinée (RMSE, pour Root Mean Squared Error). En désignant par  $(\hat{y}_i)_{1 \leq i \leq n}$  les différentes valeurs prédites  $(y_i)_{1 \leq i \leq n}$  par un modèle de régression, les formules pour le calcul de MAE, MSE et RMSE sont les suivantes :

$$MAE = \frac{1}{n} \sum_i^n |y_i - \hat{y}_i|, \tag{1.13}$$

$$MSE = \frac{1}{n} \sum_i^n (y_i - \hat{y}_i)^2, \tag{1.14}$$

$$RMSE = \sqrt{MSE}. \tag{1.15}$$

Dans ce cas, le meilleur modèle est celui qui a la plus petite valeur de l'erreur.

La qualité d'ajustement d'un modèle peut aussi se mesurer à travers le critère d'information d'Akaike ou AIC (Akaike Information Criterion) (AKAIKE 1998). L'AIC d'un modèle est donné par :

$$AIC = 2k - 2 \ln(L), \tag{1.16}$$

où  $L$  est le maximum de la fonction de vraisemblance du modèle et  $k$  est le nombre de paramètres à estimer du modèle. Parmi deux modèles, le meilleur est celui qui possède le plus petit AIC.

Ainsi, pour identifier la fonction  $g(\cdot)$  la plus adaptée, on ajuste les modèles avec plusieurs expressions de fonctions de lien, puis on retient celui qui nous fournit la plus petite valeur d'AIC ou de l'erreur de prédiction (MAE, MSE ou RMSE). Lorsque la loi  $F_0$  et la fonction  $g$  qui fournissent le meilleur ajustement sont identifiées, on estime les paramètres  $\theta$ ,  $\varphi$  et  $\beta$  du modèle par le maximum de vraisemblance, que nous décrivons dans la suite.

### 1.5.1.3 Estimation dans un GLM

Considérons l'échantillon i.i.d  $\{Y_1, \dots, Y_n\}$  de  $Y$ , et supposons que la loi de  $Y$  est issue de la famille exponentielle de fonction de densité  $f_Y(\cdot, \theta, \varphi)$ . Le logarithme de la fonction de vraisemblance des observations est alors donné par :

$$l_n(\theta, \varphi) = \sum_{i=1}^n \log(f_Y(y_i, \theta, \varphi)).$$

Il est important de rappeler que dans cette équation, le paramètre  $\theta$  est une fonction du vecteur des coefficients de régression  $\beta$ .

En remplaçant  $f_Y(\cdot, \theta, \varphi)$  par son expression donnée à l'équation (1.11), on obtient :

$$l_n(\theta, \varphi) = \sum_{i=1}^n \left( \frac{y_i \theta - b(\theta)}{a(\varphi)} + c(y_i, \varphi) \right). \quad (1.17)$$

Rappelons d'abord que l'estimateur du maximum de vraisemblance vérifie, sous certaines hypothèses, les relations classiques suivantes pour tout  $i = 1, \dots, n$  :

$$\mathbb{E} \left[ \frac{\partial}{\partial \theta} l_n(Y_i, \theta, \varphi) \mid \mathbf{X} \right] = 0, \quad (1.18)$$

$$\text{var} \left[ \frac{\partial}{\partial \theta} l_n(Y_i, \theta, \varphi) \mid \mathbf{X} \right] = -\mathbb{E} \left[ \frac{\partial^2}{\partial \theta^2} l_n(Y_i, \theta, \varphi) \mid \mathbf{X} \right]. \quad (1.19)$$

Avec la relation (1.17), on montre que  $\forall i = 1, \dots, n$ ,

$$\frac{\partial}{\partial \theta} l_n(y_i, \theta, \varphi) = \frac{y_i - b'(\theta)}{a(\varphi)},$$

et

$$\frac{\partial^2}{\partial \theta^2} l_n(y_i, \theta, \varphi) = -\frac{b''(\theta)}{a(\varphi)}.$$

Ainsi, en introduisant ces valeurs respectivement dans les équations (1.18) et (1.19), on déduit que :

$$\mathbb{E} [Y_i \mid \mathbf{X}_i] = \mu_i = b'(\theta) \quad (1.20)$$

et

$$\frac{\text{var} [Y_i \mid \mathbf{X}_i]}{a^2(\varphi)} = -\frac{b''(\theta)}{a(\varphi)} \Leftrightarrow \text{var} [Y_i \mid \mathbf{X}_i] = b''(\theta)a(\varphi) \quad (1.21)$$

Par ailleurs, rappelons que le paramètre de régression  $\beta$  et  $\mu_i = \mathbb{E}[Y_i | \mathbf{X}_i]$  sont reliés par la relation  $g(\mu_i) = \sum_{j=1}^p \beta_j X_i^j = \beta^T X_i$ . Aussi, de la relation (1.17), on a :

$$\begin{aligned}
 \frac{\partial}{\partial \beta} l_n(y_i, \theta, \varphi) &= \frac{1}{a(\varphi)} \sum_{i=1}^n \frac{\partial}{\partial \beta} (y_i \theta - b(\theta)) + \sum_{i=1}^n \frac{\partial}{\partial \beta} c(y_i, \varphi) \\
 &= \frac{1}{a(\varphi)} \sum_{i=1}^n y_i \frac{\partial \theta}{\partial \beta} - \frac{\partial b(\theta)}{\partial \beta} \quad (\text{car la quantité } c(y_i, \varphi) \text{ ne dépend pas de } \beta) \\
 &= \frac{1}{a(\varphi)} \sum_{i=1}^n y_i \frac{\partial \theta}{\partial \beta} - \frac{\partial \theta}{\partial \beta} \frac{\partial b(\theta)}{\partial \theta} \\
 &= \frac{1}{a(\varphi)} \sum_{i=1}^n \left( y_i - \frac{\partial b(\theta)}{\partial \theta} \right) \frac{\partial \theta}{\partial \beta} \\
 &= \frac{1}{a(\varphi)} \sum_{i=1}^n (y_i - \mu_i) \frac{\partial \theta}{\partial \beta} \quad (\text{en utilisant la relation (1.20)}) \\
 &= \frac{1}{a(\varphi)} \sum_{i=1}^n (y_i - \mu_i) \frac{\partial \theta}{\partial \mu_i} \frac{\partial \mu_i}{\partial \beta}.
 \end{aligned}$$

D'autres part, on a les relations suivantes :

$$\frac{\partial \theta}{\partial \mu_i} = \left( \frac{\partial \mu_i}{\partial \theta} \right)^{-1} = \left( \frac{\partial^2 b(\theta)}{\partial \theta^2} \right)^{-1} \Leftrightarrow \frac{\partial \theta}{\partial \mu_i} = \frac{1}{b''(\theta)}$$

et

$$\frac{\partial \mu_i}{\partial \beta} = \frac{\partial \mu_i}{\partial g(\mu_i)} \frac{\partial g(\mu_i)}{\partial \beta} = \left( \frac{\partial g(\mu_i)}{\partial \mu_i} \right)^{-1} \frac{\partial \beta^T X_i}{\partial \beta} = \left( \frac{\partial g(\mu_i)}{\partial \mu_i} \right)^{-1} X_i$$

En utilisant donc ces deux dernières relations, on déduit que :

$$\frac{\partial}{\partial \beta} l_n(y_i, \theta, \varphi) = \frac{1}{a(\varphi)} \sum_{i=1}^n \frac{y_i - \mu_i}{b''(\theta) g'(\mu_i)} X_i, \quad (1.22)$$

avec  $X_i = (x_i^1, \dots, x_i^p)$  et  $g'(\mu_i) = \frac{\partial g(\mu_i)}{\partial \mu_i}$ . Si on pose  $\mathbf{W} = \text{diag}(w_1, \dots, w_p)$ , avec  $w_i = [b''(\theta) g'^2(\mu_i)]^{-1}$  et  $\Delta = \text{diag}(g'(\mu_1), \dots, g'(\mu_n))$ ,  $\mu = (\mu_1, \dots, \mu_n)^T$  et  $\mathbf{Y} = (y_1, \dots, y_n)^T$ , la relation (1.22) peut s'écrire sous la forme :

$$\frac{\partial}{\partial \beta} l_n(\mathbf{Y}, \theta, \varphi) = \frac{1}{a(\varphi)} \mathbf{X}^T \mathbf{W} \Delta (\mathbf{Y} - \mu). \quad (1.23)$$

Ainsi, les coefficients de régressions sont obtenus en résolvant l'équation :

$$\mathbf{X}^T \mathbf{W} \Delta (\mathbf{Y} - \mu) = 0_{\mathbb{R}^p}. \quad (1.24)$$

Cette équation n'admet pas de solution explicite. Elle est résolue de façon itérative avec les méthodes classiques telles que la méthode de Newton-Raphson. Pour une description plus détaillée de cette méthode appliquée aux GLMs, l'on peut se référer à McCULLOCH et SEARLE 2001a.

### 1.5.1.4 Sélection de modèles linéaires généralisés

Une fois que les étapes 2 et 3 de la procédure d'ajustement sont effectuées, le modèle obtenu est appelé modèle saturé. Il s'agit du modèle qui contient toutes les variables explicatives (covariables). La sélection de modèle consiste à trouver parmi les covariables  $X^j$ ,  $j = 1, \dots, n$ , celles qui ont un effet significatif. C'est donc un sous-modèle du modèle saturé.

La significativité de la covariable  $X^j$ , pour tout  $j = 1, \dots, p$  est évaluée à travers le test de Wald, dont les hypothèses de test sont les suivantes :

$$\mathcal{H}_0 : \beta^j = 0 \text{ contre } \mathcal{H}_1 : \beta^j \neq 0.$$

Il repose sur la statistique de test  $\mathcal{W}$  définie par :

$$\mathcal{W} = [\text{var}(\hat{\beta}_j)]^{-1} \hat{\beta}_j^2,$$

où  $\hat{\beta}$  est l'estimateur de  $\beta$  obtenu en résolvant l'équation (1.24).

Sous  $\mathcal{H}_0$ , la variable aléatoire  $\mathcal{W}$  suit une loi de Khi-deux ( $\chi_1^2$ ) à un degré de liberté. Ainsi, on rejette  $\mathcal{H}_0$  au niveau  $\alpha \in ]0, 1[$  si

$$\mathcal{W} > q_{1-\alpha}^{\chi_1^2},$$

où  $q_{1-\alpha}^{\chi_1^2}$  est le quantile d'ordre  $1 - \alpha$  d'une  $\chi_1^2$ .

## 1.5.2 Les modèles de dégradation

### 1.5.2.1 Présentation générale

Les modèles de dégradation sont un ensemble de méthodes qui permettent de modéliser la détérioration d'un composant ou d'un système au fil du temps. On distingue généralement deux catégories de modèles de dégradation : les modèles à espace fini d'états (ou modèles multi-états) et les modèles de dégradation continue. Cette section se concentre sur la deuxième catégorie, en mettant particulièrement l'accent sur les modèles de dégradation Gamma. Pour une discussion sur les modèles de dégradation à espaces finis d'états, nous renvoyons le lecteur au livre de ANDERSEN et KEIDING 2002.

Mathématiquement, la dégradation est caractérisée par un processus stochastique  $(Y_t)_{t \geq 0}$  continu à droite avec limite à gauche (un processus càdlàg), dont on suppose que les trajectoires sont croissantes (ou croissantes en moyenne). Le processus  $(Y_t)_{t \geq 0}$  peut être multidimensionnel, mais nous nous restreignons au cas unidimensionnel. Ainsi, pour un instant  $t \geq 0$  donné, la variable aléatoire réelle (v.a.r)  $Y_t$  associe la valeur de la dégradation à l'instant  $t$ .

Soient  $A : \mathbb{R}^+ \mapsto \mathbb{R}^+$  une fonction positive et croissante telle que  $A(0) = 0$ , et  $\theta$  un nombre réel strictement positif. Le processus de dégradation  $(Y_t)_{t \geq 0}$  est appelé processus Gamma de fonction de forme  $A(\cdot)$  et de paramètre d'échelle  $\theta$  et on note  $(Y_t)_{t \geq 0} \sim \mathcal{G}(A(\cdot), \theta)$  si (voir KAHLE et al. 2016) :

- $Y_0 = 0$  presque sûrement :  $\mathbb{P}(Y_0 = 0) = 1$ ;
- les accroissements de  $(Y_t)_{t \geq 0}$  sont indépendants :  $\forall n \in \mathbb{N}^*$  et pour toute suite  $(t_1, \dots, t_{n+1})$  telle que  $0 \leq t_1 < t_2 < \dots < t_{n+1}$ , les accroissements  $Y_{t_{i+1}} - Y_{t_i} \forall i = 1, \dots, n$  sont indépendantes;
- $\forall 0 < s < t$ ,  $\Delta Y_t = Y_t - Y_s$  est une v.a.r qui suit une loi Gamma de paramètre de forme  $A(t) - A(s)$  et de paramètre d'échelle  $\theta$ .

On rappelle qu'une v.a.r continue suit une loi Gamma de paramètre  $(A(t) - A(s), \theta)$  si sa fonction de densité est donnée par :

$$f(x; A(t) - A(s), \theta) = \frac{x^{A(t)-A(s)-1} e^{-\frac{x}{\theta}}}{\Gamma(A(t) - A(s)) \theta^{A(t)-A(s)}}, \quad (1.25)$$

où  $\Gamma$  est la fonction Gamma standard définie sur  $\mathbb{R}_+^*$  par :  $\Gamma(z) = \int_0^{+\infty} t^{z-1} e^{-t} dt$ .

Par ailleurs, pour des réels  $s$  et  $t$  tels que  $0 < s < t$ , on montre facilement les égalités suivantes :

$$\begin{cases} \mathbb{E} [\Delta Y_t] = \theta (A(t) - A(s)) \\ \text{var} [\Delta Y_t] = \theta^2 (A(t) - A(s)). \end{cases} \quad (1.26)$$

Les modèles de dégradation sont utilisés pour modéliser les données de type longitudinales. C'est-à-dire des données obtenues en observant les réalisations du phénomène d'intérêt (la dégradation) à plusieurs instants. Autrement dit, pour un composant  $i$  donné, on observe  $N_i$  valeurs de  $Y_i$  aux temps  $t_j$  avec  $j = 1, \dots, N_i$ . Il peut arriver que pour chacun des matériaux, on ne dispose seulement que d'une seule valeur  $y_i$  de la gradation ( $N_i = 1$  pour tout  $i = 1, \dots, n$ ). C'est dans ce dernier cadre que nous nous situons.

Il est possible d'intégrer des covariables, dans le modèle de dégradation Gamma. Désignons par  $Z = (Z^1, Z^2, \dots, Z^p)$ , le vecteurs des  $p$  covariables du composant. Dans la modélisation, les covariables sont souvent appelées *stress*. Les effets de ces covariables peuvent être pris en compte à travers le paramètre fonctionnel  $A(\cdot)$  comme suggéré par BAGDONAVICIUS et NIKULIN 2000a ou à travers le paramètre d'échelle  $\theta$  comme proposé par LAWLESS et M.CROWDER 2004a.

PARK et PADGETT 2005 propose d'incorporer l'effet des stress dans la fonction de forme du processus en posant :

$$A(t) = g(Z) t^\lambda, \quad (1.27)$$

avec  $g$  une fonction définie de  $\mathbb{R}^p$  à valeur dans  $\mathbb{R}^+$  appelée fonction de lien. Ces auteurs proposent les fonctions de lien suivantes (le nom du modèle résultant est donné entre parenthèse) :

- fonction affine (modèle inverse-linéaire) :  $\forall Z = (Z_1, Z_2, \dots, Z_p)$ ,  $g(\mathbf{z}) = \alpha + \sum_{j=1}^p \beta_j Z_j$  avec  $\alpha \in \mathbb{R}$  et  $\beta \in \mathbb{R}^p$  tels que  $g(Z) \geq 0$ ;
- fonction exponentielle (modèle exponentiel) :  $\forall Z = (Z_1, Z_2, \dots, Z_p)$   $g(Z) = \alpha \exp(\sum_{j=1}^p \beta_j Z_j)$ ,  $\alpha > 0$ ;
- fonctions de la forme  $g(Z) = \alpha \exp(\sum_{j=1}^p \beta_j h(Z_j))$  où :
  - $h(Z_j) = \frac{1}{Z_j} \forall Z_j \neq 0_{\mathbb{R}^n}$  (modèle Arrhenius);
  - $h(Z_j) = \log(Z_j)$ ,  $\forall Z_j > 0$  (modèle puissance).

La deuxième manière de prendre en compte l'effet des stress est de considérer que le paramètre d'échelle de la distribution est une fonction des stress. En d'autres termes, on suppose que :

$$\theta = \exp \left( \sum_{j=1}^p \beta_j Z_j \right), \quad (1.28)$$

avec  $\beta_j \in \mathbb{R} \forall j = 1, \dots, p$ , des paramètres inconnus à estimer.

La forme de la fonction  $A(\cdot)$  peut être supposée totalement inconnue (modèle semi-paramétrique) ou dépendant d'un ensemble de paramètres (modèle paramétriques).

### 1.5.2.2 Estimation dans un modèle de dégradation Gamma

Dans les deux types de modèles, les paramètres du modèle sont estimés par maximum de vraisemblance. Supposons que nous disposons de  $K$  observations aléatoires du processus de dégradation Gamma  $(Y_t)_{t \geq 0}$  avec les paramètres  $(A(\cdot), \theta)$  à des instants  $0 = T_0 < T_1 < \dots < T_K$ . De plus, supposons que le vecteur aléatoire  $(K, T_1, \dots, T_K)$  est indépendant du processus  $(Y_t)_{t \geq 0}$ . Pour chaque  $k = 1, \dots, K$ , définissons  $Y_k = Y(T_k)$ . Nous disposons alors de  $n$  réalisations i.i.d  $(K_i, T_{i,1}, \dots, T_{i,K}, Y_{i,1}, \dots, Y_{i,K})_{1 \leq i \leq n}$  de  $(K, T_1, \dots, T_K, Y_1, \dots, Y_K)$ . Les estimateurs  $\hat{A}_n(\cdot)$  et  $\hat{\theta}_n$  des paramètres  $A(\cdot)$  et  $\theta$  respectivement sont obtenus en maximisant la fonction pseudo log de vraisemblance :

$$l_n(A(\cdot), \theta) = \sum_{i=1}^n \sum_{k=1}^{K_i} \left[ A(T_{K_i,k}) \log Y_{K_i,k} - \frac{Y_{K_i,k}}{\theta} - \log \Gamma(A(T_{K_i,k}) - A(T_{K_i,k}) \log(\theta)) \right]. \quad (1.29)$$

La procédure d'estimation et les propriétés des estimateurs font l'objet de nombreux livres et articles dont KAHLE et al. 2016 et WANG 2009.

Comme mentionné précédemment, l'étude que nous présenterons au Chapitre 3 se focalisera sur le scénario où nous disposons d'une seule observation de la dégradation pour chaque composant.

Supposons donc que, pour chaque composant  $i$ , nous observons une seule valeur (c'est-à-dire  $K_i = 1$ ) de la dégradation au temps  $T_i$ . Les observations obtenues se présentent sous la forme de paires  $(T_i, Y_i)$ , avec  $i = 1, \dots, n$ . Dans ce cas, la fonction de pseudo-vraisemblance est équivalente à la fonction de vraisemblance classique :

$$l_n(A(\cdot), \theta) = \sum_{i=1}^n \left[ A(T_i) \log Y_i - \frac{Y_i}{\theta} - \log \Gamma(A(T_i) - A(T_i) \log(\theta)) \right]. \quad (1.30)$$

Les paramètres du modèle sont estimés en résolvant les équations de vraisemblance, en fonction de la manière dont l'effet des covariables est pris en compte dans le modèle.

Par ailleurs, dans ce cas particulier où chaque  $K_i = 1$ , on montre sans difficulté que le modèle de dégradation Gamma peut être assimilé à un modèle linéaire généralisé Gamma. En effet, supposons que les effets des covariables soient pris en compte à travers l'expression de la fonction de forme  $A(\cdot)$ . Dans ce contexte, le processus de dégradation étant un processus Gamma, on obtient les relations suivantes :

$$\begin{cases} Y_i = Y(T_i) \sim \mathcal{G}(A(T_i), \theta) \\ \mu_i = \mathbb{E}[Y_i | \mathbf{Z}_i] = \theta A(T_i) = g(\mathbf{Z}_i) T_i^\lambda \end{cases} \quad (1.31)$$

Ainsi, pour une fonction de lien bien choisie (en fonction de la forme de  $A$ ), on remarque que le modèle devient un GLM Gamma.

Supposons par exemple que pour tout  $t > 0$ , fonction  $A(\cdot)$  est donnée par :

$$A(t) = \exp\left(\sum_{j=1}^p \beta_j Z_j\right) t^\lambda.$$

Alors de (1.31), on déduit que :

$$\begin{cases} Y_i = Y(T_i) \rightsquigarrow \mathcal{G}(A(T_i), \theta) \\ \mu_i = \theta A(T_i) = \theta \exp(\sum_{j=1}^p \beta_j Z_j) T_i^\lambda \Rightarrow \log(\mu_i) = \sum_{j=1}^p \beta_j Z_i^j + \lambda \log(T_i) + \log(\theta) \end{cases} .$$

En posant  $\tilde{Z} = (1, Z_i^1, \dots, Z_i^p, \log(T_i))$  et  $\tilde{\beta} = (\log(\theta), \beta_1, \dots, \beta_p, \lambda)$ , on obtient

$$\begin{cases} Y_i = Y(T_i) \sim \mathcal{G}(A(T_i), \theta) \\ \log(\mu_i) = \sum_{j=1}^p \tilde{\beta}_j \tilde{Z}_i^j \end{cases} .$$

Ce qui n'est rien d'autre qu'un GLM Gamma avec la fonction de lien logarithme classique. Ainsi, la procédure d'estimation présentée à la Section 1.5.1.3 précédente est appliquée pour estimer les paramètres  $\beta$ ,  $\lambda$ , et  $\theta$ .

### 1.5.2.3 Ajustement d'un modèle de dégradation Gamma

Pour ajuster un modèle de dégradation à des données, il faut d'abord identifier la fonction  $A(\cdot)$  la mieux adaptée parmi les différentes expressions possibles. Ensuite, il faut estimer les paramètres du modèle. Enfin, il est nécessaire d'identifier quelles covariables ont un effet significatif sur la dégradation.

L'identification de la fonction  $A(\cdot)$  se déroule de la même manière que l'identification de la fonction de lien dans un GLM (voir la Section 1.5.1.2). En d'autres termes, on ajuste les modèles avec différentes expressions de  $A(\cdot)$  et on choisit celle qui a la plus petite valeur d'AIC ou d'erreur de prédiction.

Comme nous l'avons dit, l'estimation se fait par la maximisation de la fonction pseudo log de vraisemblance. La sélection de modèles se fait, elle, par la biais des tests de Wald, comme décrit dans la Section 1.5.1.4.

## 1.5.3 Les modèles de durée de vie

Les GLMs et les modèles de dégradation Gamma sont des outils statistiques couramment utilisés pour étudier la dégradation et les facteurs qui peuvent l'influencer. Cependant, comme tout outil, ils présentent certaines limites. Les GLMs, par exemple, ne prennent en général pas en compte les covariables dépendant du temps. Même si certaines extensions permettent de modéliser les données longitudinales, elles deviennent difficiles à appliquer lorsque l'on dispose de plusieurs moments de mesure des covariables.

Comme mentionné précédemment, les modèles de dégradation Gamma sont bien adaptés pour la modélisation des données longitudinales et permettent de prendre en compte l'effet des variables dépendant ou non du temps. Cependant, ils ne sont pas adaptés lorsque l'on dispose d'une seule valeur de dégradation pour chaque composant, comme c'est le cas dans notre étude.

De plus, ces outils (GLMs et modèles de dégradation Gamma) ne tiennent généralement pas compte des phénomènes de censure souvent présents dans les applications et que nous rencontrons dans notre étude.

Ainsi, dans cette partie, nous introduirons les modèles de durée de vie, qui non seulement permettent la prise en compte des phénomènes de censure mais aussi l'effet des covariables dépendant du temps.

### 1.5.3.1 Présentation générale

L'analyse de survie est un ensemble de méthodes statistiques permettant d'étudier le temps avant l'occurrence d'un événement d'intérêt, ainsi que les différents facteurs qui peuvent



influencer cet événement. Ces méthodes sont applicables à de nombreux domaines. En effet, l'événement d'intérêt peut prendre diverses formes, telles que la signature du premier contrat à durée indéterminée pour un nouveau diplômé, la résiliation d'un contrat ou d'un abonnement, la guérison d'un patient suite à un traitement, la défaillance d'un composant ou d'un système, etc. Lorsque l'événement d'intérêt est la panne d'un composant, le temps avant l'occurrence de cette défaillance, que nous désignons ici par la variable aléatoire positive  $T$ , est appelé la durée de vie (Time-To-Failure en anglais) du composant. La connaissance de la durée de vie des composants est un élément clé pour garantir la fiabilité du système dans lequel ils sont intégrés, car elle permet de planifier des maintenances préventives.

La durée de vie  $T$  est caractérisée par sa fonction de survie, également appelée fonction de fiabilité (reliability function en anglais), ainsi que par sa fonction de risque instantané (hazard function).

La fonction de fiabilité, notée  $R(\cdot)$  est définie sur  $\mathbb{R}^+$  et à valeur dans l'intervalle  $[0, 1]$ , pour tout  $t > 0$ , par

$$R(t) = \mathbb{P}(T > t). \quad (1.32)$$

La quantité  $R(t)$  désigne donc la probabilité de fonctionnement sans défaillance du composant sur l'intervalle de temps  $[0, t]$ .

Le risque instantané ou taux de défaillance (hazard rate) à l'instant  $t > 0$  est la probabilité infinitésimale que le composant tombe en panne juste après l'instant  $t$ , sachant qu'il était en fonctionnement juste avant  $t$ . Il es donné par :

$$\lambda(t) = \lim_{h \rightarrow 0^+} \frac{\mathbb{P}(t \leq T < t+h \mid T \geq t)}{h} = -\frac{dR(t)}{R(t)}. \quad (1.33)$$

Ainsi, la fonction de fiabilité et la fonction de risque instantané sont liées par la relation :

$$R(t) = \exp\left(-\int_0^t \lambda(s)ds\right) = \exp(-\Lambda(t)), \quad (1.34)$$

où la fonction définie par  $\Lambda(t) = \int_0^t \lambda(s)ds$  est appelée fonction de risque cumulé. Elle est aussi utilisée pour caractériser la v.a.r  $T$  car elle permet d'avoir la fonction de fiabilité grâce à la relation (1.34).

Un indice de fiabilité d'un composant est donné par sa durée moyenne de fonctionnement avant la défaillance ou *MTTF* (Mean Time To Failure). Elle est donnée par :

$$MTTF = \mathbb{E}(T) = \int_0^{+\infty} tf(t)dt = \int_0^{+\infty} R(t)dt, \quad (1.35)$$

où la fonction  $f$  désigne la fonction de densité de la v.a.r  $T$ .

Lorsqu'un composant n'a pas encore subi de défaillance après une durée de fonctionnement  $x$ , il est souvent utile de modéliser la durée supplémentaire pendant laquelle il peut encore fonctionner avant de tomber en panne. La durée de vie résiduelle à l'âge  $x$  désignée par la v.a.r  $\tau_x$ , fourni cette information. Ainsi, la probabilité que cette durée résiduelle soit supérieur à  $t \geq 0$ , notée  $R(t \mid x)$  est donnée par :

$$R(t \mid x) = \mathbb{P}(\tau_x > t) = \mathbb{P}(T - x > t \mid T > x) = \frac{R(x+t)}{R(x)}. \quad (1.36)$$

Par ailleurs, l'un des points forts des modèles de durée de vie est leur capacité à prendre en compte les informations censurées. En effet, dans certaines expériences en vue de l'étude de fiabilité, il arrive que l'on ne soit pas capable d'observer pour certains composants, le moment exact de la défaillance; dans ce cas l'information obtenue est dite censurée. Les types de données censurées les plus couramment rencontrés sont (KLEIN et MOESCHBERGER 2003) :

- les données censurées à droite (right censored data) : si on observe l'état du composant à l'instant  $t_r$  et qu'à cette date il n'a pas encore eu de défaillance, la donnée obtenue ( $t_r$ ) est une donnée censurée à droite et la seule information disponible sur le temps de défaillance  $t$  du composant est :  $t > t_r$ ;
- les données censurées à gauche (left censored data) : si on n'observe l'état du composant à  $t_l$  et qu'on constate que le composant a déjà été défaillant, le temps  $t_l$  sera une donnée censurée à gauche et dans ce cas :  $t < t_l$ ;
- les données censurées par intervalle (interval censored data) : si le composant est survenu à un instant inconnu entre les dates  $t_l$  et  $t_r$ , il s'agit d'une observation censurée par intervalle, et on sait seulement que :  $t_l < t < t_r$ .

Dans le cas particulier où tous les composants de l'étude sont observés qu'une seule fois et que la donnée obtenue pour chacun d'eux est soit censurée à droite, soit censurée à gauche, les données obtenues sont appelées données à statut d'état connues sous le terme anglo-saxon *current status data*. Elles sont aussi appelées données censurées par intervalle de type I.

Comme indiqué précédemment, l'analyse de survie permet également d'identifier et d'étudier les facteurs qui ont une influence sur la durée de vie du composant ou du système. Les modèles de régression en survie sont les outils développés à cet effet. Ils permettent de modéliser et quantifier l'effet de ces facteurs ou covariables sur la durée de vie du composant. Les modèles de régression les plus couramment rencontrés dans la bibliographie sont :

- les modèles à risques proportionnels qui sont également connus sous les noms de modèles de Cox ou *Proportional Hazards models* en anglais;
- les modèles à côtes proportionnelles (*proportional odds models*);
- les modèles de durée de vie accélérée (*Accelerated Failure Time models*).

En présence de données complètes et/ou censurées à droite (éventuellement censurées à gauche et/ou par intervalle), il existe une vaste bibliographie d'articles et de livres décrivant les modèles de régression en survie ainsi que les méthodes d'estimation de paramètres associés à ces modèles. On peut énumérer par exemple le livre de KLEIN et MOESCHBERGER 2003 qui traite de façon détaillée les modèles de régression en présence de données censurées à droite et/ou à gauche, et celui de SUN 1978 qui se focalise sur les données censurées par intervalles dont les *current status data*.

Dans cette thèse, nous nous focalisons sur les techniques qui traitent des *current status data*. Plus précisément nous présenterons les modèles à risques proportionnels et les modèles à côtes proportionnelles en présence de *current status data*.

### 1.5.3.2 Modèles à risques proportionnels en présence de *current status data*

#### a) Présentation

Plaçons nous dans le cadre des données à statut d'état et désignons toujours par  $T$  la variable aléatoire désignant la durée de vie, par  $X$  la v.a.r associant les instants d'observations des composants, et par  $\delta = I(T < X)$  l'indicateur de la censure ( $\delta$  vaut 1 si l'observation est censurée à gauche et 0 sinon). Désignons aussi par  $Z(t) = (Z^1(t), \dots, Z^p(t))^T$ , le vecteur

des  $p$  covariables observées a un certain instant  $t \geq 0$  donné. Soulignons que pour tout vecteur  $u$ , le vecteur  $u^T$  désigne le vecteur transposé de  $u$ . De plus, pour tout  $j = 1, \dots, p$ , la covariable  $Z^j$  peut être indépendant du temps et dans ce cas on le notera  $Z^j$ . Dans ce cas l'information contenue dans les données est résumée à travers la variable  $(X, \delta, Z)$ . Désignons aussi par  $Z^*(t) = \{Z(u), 0 \leq u < t\}$ , l'histoire du processus des covariables jusqu'au temps  $t$  et supposons que  $X$  est indépendante de  $T$  et  $Z(\cdot)$ .

Le modèle à risques proportionnels proposé dans Cox 1972 suppose que le risque instantané, à tout instant  $t$ , conditionnellement à l'histoire  $Z^*(t)$ , des covariables est donné par la relation :

$$\lambda(t | Z^*(t)) = \lambda_0(t) \exp(\beta^T Z(t)), \quad (1.37)$$

avec :

- $\beta = (\beta_1, \dots, \beta_p) \in \mathbb{R}^p$  le vecteur des coefficients de régression ; il est inconnu et est à estimer.
- $\lambda_0(\cdot)$  la fonction de *risque instantané de base* (baseline). Elle est inconnue et correspond au risque instantané du composant lorsque les valeurs des covariables à l'instant  $t$  sont toutes nulles.

En fonction de l'hypothèse faite sur la fonction de risque de base, on distingue deux types de modèles à risques proportionnels :

- les modèles semi-paramétriques dans lesquels aucune hypothèse n'est faite sur la forme de la fonction de risque de base  $\lambda_0(\cdot)$ . Dans ce cas elle est entièrement estimée.
- les modèles paramétriques dans lesquelles on suppose que  $\lambda_0(\cdot)$  a la forme de la fonction de risque d'une famille de distribution connue. Le Tableau 1.2 extrait de KLEIN et MOESCHBERGER 2003 donne les expressions des MTTF et des fonctions de risque instantané, de fiabilité et de densité des familles de distribution les plus couramment utilisées.

Distribution	Hazard Rate $b(x)$	Survival Function $S(x)$	Probability Density Function $f(x)$	Mean $E(X)$
Exponential $\lambda > 0, x \geq 0$	$\lambda$	$\exp[-\lambda x]$	$\lambda \exp(-\lambda x)$	$\frac{1}{\lambda}$
Weibull $\alpha, \lambda > 0, x \geq 0$	$\alpha \lambda x^{\alpha-1}$	$\exp[-\lambda x^\alpha]$	$\alpha \lambda x^{\alpha-1} \exp(-\lambda x^\alpha)$	$\frac{\Gamma(1+1/\alpha)}{\lambda^{1/\alpha}}$
Gamma $\beta, \lambda > 0, x \geq 0$	$\frac{f(x)}{S(x)}$	$1 - I(\lambda x, \beta)^*$	$\frac{\lambda^\beta x^{\beta-1} \exp(-\lambda x)}{\Gamma(\beta)}$	$\frac{\beta}{\lambda}$
Log normal $\sigma > 0, x \geq 0$	$\frac{f(x)}{S(x)}$	$1 - \Phi\left(\frac{\ln x - \mu}{\sigma}\right)$	$\frac{\exp\left(-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2\right)}{x(2\pi)^{1/2}\sigma}$	$\exp(\mu + 0.5\sigma^2)$
Log logistic $\alpha, \lambda > 0, x \geq 0$	$\frac{\alpha x^{\alpha-1} \lambda}{1 + \lambda x^\alpha}$	$\frac{1}{1 + \lambda x^\alpha}$	$\frac{\alpha x^{\alpha-1} \lambda}{(1 + \lambda x^\alpha)^2}$	
Normal $\sigma > 0, -\infty < x < \infty$	$\frac{f(x)}{S(x)}$	$1 - \Phi\left(\frac{x - \mu}{\sigma}\right)$	$\frac{\exp\left(-\frac{1}{2}\left(\frac{x - \mu}{\sigma}\right)^2\right)}{(2\pi)^{1/2}\sigma}$	$\mu$
Exponential power $\alpha, \lambda > 0, x \geq 0$	$\alpha \lambda^\alpha x^{\alpha-1} \exp(-[\lambda x]^\alpha)$	$\exp\{1 - \exp(-[\lambda x]^\alpha)\}$	$\alpha e \lambda^\alpha x^{\alpha-1} \exp(-[\lambda x]^\alpha) - \exp\{-\exp(-[\lambda x]^\alpha)\}$	$\int_0^\infty S(x) dx$
Gompertz $\theta, \alpha > 0, x \geq 0$	$\theta e^{\alpha x}$	$\exp\left(\frac{\theta}{\alpha}(1 - e^{\alpha x})\right)$	$\theta e^{\alpha x} \exp\left(\frac{\theta}{\alpha}(1 - e^{\alpha x})\right)$	$\int_0^\infty S(x) dx$
Inverse Gaussian $\lambda \geq 0, x \geq 0$	$\frac{f(x)}{S(x)}$	$\Phi\left(\left(\frac{\lambda}{2}\right)^{1/2}\left(1 - \frac{x}{\mu}\right)\right) - e^{2\lambda/\mu} \Phi\left[-\left(\frac{\lambda}{2}\right)^{1/2}\left(1 - \frac{x}{\mu}\right)\right]$	$\left(\frac{\lambda}{2\pi x^3}\right)^{1/2} \exp\left(-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right)$	$\mu$
Pareto $\theta > 0, \lambda > 0, x \geq \lambda$	$\frac{\theta}{x}$	$\frac{\lambda^\theta}{x^\theta}$	$\frac{\theta \lambda^\theta}{x^{\theta+1}}$	$\frac{\theta \lambda}{\theta - 1}$ if $\theta > 1$
Generalized gamma $\lambda > 0, \alpha > 0, \beta > 0, x \geq 0$	$\frac{f(x)}{S(x)}$	$1 - I(\lambda x^\alpha, \beta)$	$\frac{\alpha \lambda^\beta x^{\beta-1} \exp(-\lambda x^\alpha)}{\Gamma(\beta)}$	$\int_0^\infty S(x) dx$

\*  $I(t, \beta) = \int_0^t u^{\beta-1} \exp(-u) du / \Gamma(\beta)$  avec  $\Gamma(\cdot)$  la fonction Gamma usuelle.

**Table 1.2** – Les fonctions de fiabilité et de risque instantané de certaines lois usuelles extraites de KLEIN et MOESCHBERGER 2003.

La popularité des modèles de Cox est due en partie à la facilité d'interprétation des coefficients de régression. En effet, si pour deux composants  $i$  et  $i'$  donnés, on suppose que les valeurs des covariables  $Z^k$ ,  $k \neq j$  sont identiques à l'instant  $t$  sauf pour la covariable  $Z^j$ , alors on obtient les égalités suivantes :

$$\frac{\lambda(t | Z_i^*(t))}{\lambda(t | Z_{i'}^*(t))} = \exp(\beta_j(Z_i^j(t) - Z_{i'}^j(t))) \Rightarrow \lambda(t | Z_i^*(t)) = \exp(\beta_j(Z_i^j(t) - Z_{i'}^j(t))) \lambda(t | Z_{i'}^*(t)).$$

On en déduit que :

- si  $\beta_j > 0$ , le composant qui possède la plus grande valeur de  $Z^j$  des deux est le plus à risque à  $t$  (plus grande valeur du risque) et dans ce cas on dit que la covariable  $Z^j$  est défavorable à la survie ;
- si  $\beta_j < 0$ , le composant qui a la plus grande valeur de  $Z^j$  est le moins à risque à  $t$  (plus petite valeur du risque), et on dit que  $Z^j$  est favorable à la survie.

En particulier si  $Z_i^j = Z_{i'}^j + 1$  presque sûrement, alors

$$\frac{\lambda(t | Z_i^*(t))}{\lambda(t | Z_{i'}^*(t))} = \exp(\beta_j) \Rightarrow \lambda(t | Z_i^*(t)) = \exp(\beta_j) \lambda(t | Z_{i'}^*(t)).$$

Donc en supposant que les valeurs observées des covariables  $Z^k$ ,  $k \neq j$  sont identiques, l'évolution d'une unité de la covariable  $Z^j$  implique la multiplication du risque par  $\exp(\beta_j)$ .

### b) Estimation dans un modèle à risques proportionnels

Soit  $\{X_i, \delta_i, Z^*(X_i)\}_{1 \leq i \leq n}$  un échantillon de  $n$  observations i.i.d de  $(X, \delta, Z)$ . Le logarithme de la fonction de vraisemblance est alors donnée par :

$$l_n(\beta, \lambda_0) = \sum_{i=1}^n \delta_i \log[1 - R(X_i | Z_i^*(X_i))] + (1 - \delta_i) \log[R(X_i | Z_i^*(X_i))].$$

En utilisant le lien entre la fonction de fiabilité et la fonction de risque cumulé, on obtient :

$$l_n(\beta, \lambda_0) = \sum_{i=1}^n \delta_i \log[1 - \exp(-\Lambda(X_i | Z_i^*(X_i)))] - (1 - \delta_i) \Lambda(X_i | Z_i^*(X_i)), \quad (1.38)$$

où

$$\Lambda(X_i | Z_i^*(X_i)) = \int_0^{X_i} \lambda_0(s) e^{\beta^T Z_i(s)} ds.$$

Dans le cas des modèles sémi-paramétriques, les approches permettant d'estimer les paramètres  $\Lambda_0(\cdot)$  et  $\beta$  sont essentiellement des méthodes itératives.

Lorsque toutes covariables sont indépendantes du temps, de nombreuses méthodes sont proposées dans la littérature pour estimer  $\Lambda_0(\cdot)$  et  $\beta$ . Dans ce cas en effet, l'équation (1.38) devient :

$$l_n(\beta, \Lambda_0) = \sum_{i=1}^n \delta_i \log[1 - \exp(-\Lambda_0(X_i) e^{\beta^T Z_i})] - (1 - \delta_i) \Lambda_0(X_i) e^{\beta^T Z_i}. \quad (1.39)$$

Désignons par  $0 < s_1 < s_2 < \dots < s_m$ , les  $m$  valeurs distinctes et ordonnées des  $(X_i)_{1 \leq i \leq n}$ , alors, l'estimateur du maximum de vraisemblance  $\hat{\Lambda}_n(\cdot)$  de  $\Lambda_0(\cdot)$  est une fonction en escalier qui change de valeur aux instants  $s_l$ , pour tout  $l = 1, \dots, m$  et de valeurs  $0 = \hat{\Lambda}_n(0) \leq \hat{\Lambda}_n(s_1) \leq \dots \leq \hat{\Lambda}_n(s_m)$ . Ainsi, pour tout  $j = 1, \dots, m$ , en posant  $\Lambda_0(s_j) = \alpha_j$ , le logarithme de la fonction de vraisemblance dévient

$$l_n(\beta, \alpha) = \sum_{j=1}^m \sum_{i=1}^n I(X_i = s_j) \left( \delta_i \log[1 - \exp(-\alpha_j e^{\beta^T Z_i})] - (1 - \delta_i) \alpha_j e^{\beta^T Z_i} \right), \quad (1.40)$$

avec  $0 = \alpha_1 \leq \alpha_2 \leq \dots \leq \alpha_m$ . On peut remarquer avec cette dernière expression que la fonction est concave  $l_n(\beta, \alpha)$  en  $\alpha$  pour un  $\beta$  donné. Les estimateurs  $\hat{\alpha}_n$  et  $\hat{\beta}_n$  de  $\alpha$  et  $\beta$  respectivement sont obtenus à l'aide de méthodes itératives proposées dans des articles tels que JIAN 1996 et WELLNER et ZHAN 1997. Il s'agit de méthodes itératives qui consistent à :

1. Fixer une valeur arbitraire pour  $\beta$  puis trouver les  $\{\alpha_{l,\beta}, l = 1, \dots, m\}$  qui maximisent  $l_n(\beta, \alpha)$  sous la contrainte  $0 = \alpha_1 \leq \alpha_2 \leq \dots \leq \alpha_m$ ;
2. remplacer les  $\Lambda_0(X_i)$  par leur valeurs obtenues à l'étape 1) dans l'équation (1.40) puis estimer  $\hat{\beta}_n$ , l'estimateur de  $\beta$ ;
3. Répéter les étapes 1) et 2) jusqu'à ce qu'un critère de convergence soit vérifié.

Dans le cas où les covariables dépendent du temps, les travaux qui se focalisent sur l'estimation des  $\Lambda_0(\cdot)$  et de  $\beta$  du modèle à risques proportionnels en présence de current status data sont moins nombreuses. On peut cependant énumérer les travaux de Zeng et ses co-auteurs. Ces derniers, en supposant que la fonction de risque instantané est constante par morceaux, proposent un algorithme EM (Expectation-Maximization) pour estimer  $\Lambda_0(\cdot)$  et  $\beta$  en présence de données censurées par intervalle.

En outre, comme proposé par HUANG et WELLNER 1997, la procédure utilisée dans le cas des covariables indépendantes du temps peut être adaptée pour estimer les paramètres du modèle en présence de covariables dépendantes du temps et current status data. En effet, si on suppose que  $\lambda_0(\cdot)$  est une fonction qui prend des valeurs constantes notées  $\alpha_l$  (avec  $\alpha_l \geq 0$ ) sur les intervalles  $[s_{l-1}, s_l[$  pour tout  $l = 1, \dots, m$ , alors l'estimateur du risque cumulé de base est toujours une fonction en escalier dont le saut est  $\alpha_l$  à l'instant  $s_l$ . On obtient ainsi que

$$\Lambda(X_i | Z_i^*(t)) = \sum_{l=1}^m I(s_l \leq X_i) (s_l - s_{l-1}) \alpha_l e^{\beta^T Z(s_l)}.$$

En plus, pour chaque composant  $i$ , supposons que les valeurs des covariables soient mises à jour aux instants  $\tau_{i0}, \tau_{i1}, \dots, \tau_{iK_i}$  et restent constantes sur les intervalles  $[\tau_{ik}, \tau_{i(k+1)}[$  entre deux mises à jour successives. Notons par  $Z_{ik}$  la valeur du processus des covariables sur l'intervalle  $[\tau_{ik}, \tau_{i(k+1)}[$ , pour tout  $k = 0, 1, \dots, (K_i - 1)$ . Dans ce cas, l'expression du risque cumulé conditionnellement à l'histoire du processus du covariable devient :

$$\Lambda(X_i | Z_i^*(t)) = \sum_{l=1}^m \sum_{k=0}^{K_i-1} I(s_l \leq X_i) I(s_l \leq \tau_{i(k+1)}) (s_l - s_{l-1}) \alpha_l e^{\beta^T Z_{ik}}.$$

Ainsi, en remplaçant  $\Lambda(X_i | Z_i^*(t))$  par sa valeur dans l'équation (1.38), on détermine les estimateurs  $\hat{\alpha}_n$  de  $\alpha = (\alpha_1, \dots, \alpha_m)$  et  $\hat{\beta}_n$  de  $\beta$  en maximisant la fonction obtenue sous la contrainte  $\alpha_l \geq 0$ , pour tout  $l = 1, \dots, m$  en utilisant l'approche proposée par JIAN 1996 (présentée ci-dessus) pour les covariables indépendantes du temps.

Dans le cas des modèles paramétriques, c'est-à-dire lorsque la forme de  $\lambda_0(\cdot)$  est décrite à travers un paramètre  $\theta$  (dans ce cas  $\lambda_0(\cdot) = \lambda_\theta(\cdot)$ ), les auteurs supposent également que le vecteur des covariables prend des valeurs constantes sur les intervalles  $]\tau_{ik}, \tau_{i(k+1)}]$ , où les  $\tau_{ik}, k = 1, \dots, K_i$  sont les instants de mise à jour des covariables du composant  $i$ . Ainsi, les estimateurs  $\hat{\beta}_n$  et  $\hat{\theta}_n$  sont obtenus en maximisant le logarithme de la fonction de vraisemblance, qui est donnée par :

$$l_n(\beta, \theta) = \sum_{i=1}^n \delta_i \log \left[ 1 - \exp \left( - \sum_{k=0}^{K_i-1} I(X_i \leq \tau_{i(k+1)}) \left[ \Lambda_\theta(\tau_{i(k+1)}) - \Lambda_\theta(\tau_{ik}) \right] e^{\beta^T Z_{ik}} \right) \right] - (1 - \delta_i) \sum_{k=0}^{K_i-1} I(X_i \leq \tau_{i(k+1)}) \left[ \Lambda_\theta(\tau_{i(k+1)}) - \Lambda_\theta(\tau_{ik}) \right] e^{\beta^T Z_{ik}}. \quad (1.41)$$

Les estimateurs  $\hat{\beta}_n$  et  $\hat{\theta}_n$  n'ont pas d'expressions explicites. Ils sont obtenus grâce à des méthodes du type Newton-Raphson. SPARLING et al. 2006 étudie les propriétés asymptotiques de ces estimateurs.

### 1.5.3.3 Modèles à côtes proportionnelles en présence de current status data

#### a) Présentation

De façon générale, la cote d'un événement aléatoire à deux issues est définie comme le rapport entre la probabilité de succès et la probabilité d'échec. En partant du même principe, les modèles à cotes proportionnelles (ou Proportional Odds Ratio Models) sont définis en supposant que la cotes de l'occurrence de la défaillance conditionnellement à l'histoire des covariables, donnée par  $\frac{1-R(t|Z_i^*(t))}{R(t|Z_i^*(t))}$ , est proportionnelle à la cote de base qui est  $\frac{1-R_0(t)}{R_0(t)}$ . Pour tout instant  $t \geq 0$ ,

$$\frac{1 - R(t | Z_i^*(t))}{R(t | Z_i^*(t))} = \frac{1 - R_0(t)}{R_0(t)} \exp(\beta^T Z(t)), \quad (1.42)$$

où :

- $\beta = (\beta_1, \dots, \beta_p) \in \mathbb{R}^p$  est le vecteur des coefficients de régression à estimer ;
- $R_0(\cdot)$ , appelée fonction de fiabilité de base (baseline), correspond à la fonction de fiabilité en absence de covariables.

Comme pour les modèles de Cox, on distingue les modèles PO (référence au terme anglais Proportional Odds) paramétriques des modèles PO semi-paramétriques. Dans les modèles paramétriques, on suppose que  $R_0(\cdot)$  a la forme de la fonction de fiabilité d'une famille de distribution connue et décrite par un paramètre  $\theta$  tandis que dans les modèles semi-paramétriques, aucune hypothèse n'est faite sur  $R_0(\cdot)$ .

Aussi, les coefficients  $\beta_j$ , pour tout  $j = 1, \dots, p$ , peuvent être interprétés de la même manière que dans les modèles de Cox. En effet, pour deux composants  $i$  et  $i'$  donnés, si on suppose que les valeurs des covariables  $Z_k$  à l'instant  $t$  pour tout  $k \neq j$  sont identiques, alors on remarque que :

$$\frac{1 - R(t | Z_i^*(t))}{R(t | Z_i^*(t))} \bigg/ \frac{1 - R(t | Z_{i'}^*(t))}{R(t | Z_{i'}^*(t))} = \exp \left[ \beta_j (Z_i^j(t) - Z_{i'}^j(t)) \right].$$

Ainsi, si  $\beta_j < 0$ , le composant avec la plus grande valeur de  $Z^j$  aura la plus petite valeur de la cote et donc une plus grande valeur de la fiabilité à l'instant  $t$ . On dit dans ce cas que la covariable  $Z^j$  est un bon pronostique. Naturellement, si  $\beta_j > 0$ , on fait le constat opposé au cas  $\beta_j < 0$  et on dit dans ce cas que  $Z^j$  est un mauvais pronostique.

Par analogie aux modèles de Cox, si  $Z_i^j = Z_{i'}^j + 1$  presque sûrement, alors

$$\frac{1 - R(t | Z_i^*(t))}{R(t | Z_i^*(t))} \bigg/ \frac{1 - R(t | Z_{i'}^*(t))}{R(t | Z_{i'}^*(t))} = \exp(\beta_j).$$

En supposant donc que les covariables  $Z^k, k \neq j$  sont identiques, l'augmentation d'une unité de la covariable  $Z^j$  va entraîner la multiplication de la cote de  $i$  par  $\exp(\beta_j)$  par rapport à celle de  $i'$ .

### b) Estimation dans un modèle à côtes proportionnelles

Posons  $H(t) = -\text{logit}(R_0(t))$  où  $\text{logit}$  est la fonction définie pour tout  $p \in ]0, 1[$  par

$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right).$$

Dans ce cas, l'équation (1.42), peut s'écrire sous la forme :

$$-\text{logit}(R(t | Z^*(t))) = -\text{logit}(R_0(t)) \exp(\beta^T Z(t)). \quad (1.43)$$

Ainsi, lorsqu'on dispose d'un échantillon  $\{X_i, \delta_i, Z_i(\cdot)\}_{1 \leq i \leq n}$  formé de  $n$  observations i.i.d de  $(X, \delta, Z)$ , le logarithme de la fonction de vraisemblance sous l'hypothèse de côtes proportionnelles est donné par :

$$l_n(H, \beta) = \sum_{i=1}^n \delta_i \left[ H(X_i) + \beta^T Z(X_i) \right] - \log \left[ 1 + \exp \left( H(X_i) + \beta^T Z(X_i) \right) \right].$$

Les mêmes approches que celles présentées dans le cadre des modèles de Cox sont utilisées pour obtenir les estimateurs des paramètres  $H(\cdot)$  et  $\beta$ . Dans le cas des modèles paramétriques, on remplace les valeurs de  $H(t)$  par  $H_\theta(t)$  définie  $H_\theta(t) = -\text{logit}(R_\theta(t))$  où la fonction  $R_\theta(\cdot)$  est la fonction de fiabilité correspondante à une famille de distribution donnée (voir Table 1.2). Dans le cas semi-paramétrique, on suppose que la fonction  $H(\cdot)$  est constante par morceaux et prend les valeurs  $\alpha_l$  sur les intervalles  $]s_l, s_{l+1}]$  où les  $(s_l)_{1 \leq l \leq m}$  désignent les  $m$  valeurs distinctes ordonnées des instant d'observations  $(X_i)_{1 \leq i \leq n}$ . Dans ce cas, en posant  $\alpha = (\alpha_1, \dots, \alpha_m)$ , on obtient

$$l_n(\alpha, \beta) = \sum_{i=1}^n \delta_i \left[ \sum_{l=1}^{m-1} \alpha_l I_{X_i \in ]s_l, s_{l+1}] } + \beta^T Z(X_i) \right] - \log \left[ 1 + \exp \left( \sum_{l=1}^{m-1} \alpha_l I_{X_i \in ]s_l, s_{l+1}] } + \beta^T Z(X_i) \right) \right]$$

La procédure présentée à la section 1.5.3.2 est ainsi utilisée pour obtenir les estimateurs du modèles (voir la section 5.3 de SUN 1978).

## 1.6 Conclusion

Cette introduction met en évidence l'importance cruciale de la durabilité des structures en béton armé, en particulier compte tenu de la réduction significative de la résistance à la traction du béton par rapport à sa résistance à la compression. L'intégration d'armatures dans le béton est une solution courante pour renforcer sa capacité à résister à la fois à la compression et à la traction. Cependant, le défi majeur réside dans la prévention de la corrosion de ces armatures, ce qui peut compromettre la sécurité et la longévité de la structure. Pour aborder ces problèmes complexes, il est nécessaire de comprendre en profondeur le phénomène de corrosion, ainsi que les facteurs déclencheurs tels que la carbonatation et la pénétration d'ions chlorure. Ce premier chapitre a jeté les bases pour cette étude en présentant une vue d'ensemble du matériau béton, en exposant le phénomène de corrosion, et en soulignant l'importance de la modélisation de la durabilité du béton armé, en mettant particulièrement l'accent sur la carbonatation. En effet, il révèle l'existence de nombreux modèles de carbonatation qui diffèrent les uns des autres par les méthodes de modélisation et par le cadre spécifique à l'étude. Les modèles analytiques qui sont les plus couramment utilisés par les ingénieurs pour évaluer la vulnérabilité du matériau en béton armé vis à vis de la carbonatation sont divers mais ils contiennent trop de paramètres d'entrée et/ou sont développés pour des bétons spécifiques. Particulièrement, ceux basés sur l'approche statistiques présentent une facilité d'utilisation mais ils sont souvent calibrés sur des données issues de carbonatation accélérée ce qui limite leur application aux matériaux exposés à la carbonatation naturelle.

Ce constat met donc en avant la nécessité de disposer d'un modèle statistique de carbonatation fiable, intégrant les principaux facteurs de carbonatation et pouvant être utilisé pour prédire le niveau de carbonatation d'un matériau exposé à la carbonatation naturelle.

Ce chapitre décrit également certaines méthodes statistiques couramment utilisées pour répondre à ce type de problématiques. Il s'agit des modèles de régression généralisés, les processus de dégradation Gamma et les modèles de durée de vie. À mesure que nous progresserons à travers les prochains chapitres, nous utiliserons ces méthodes pour analyser la carbonatation d'une structure en béton armé.



## Chapitre 2

---

# Présentation des bases de données de modélisation

---

## 2.1 Introduction

Pour appliquer les différentes méthodes statistiques présentées dans le chapitre précédent, un élément essentiel qui constitue en quelque sorte le socle de l'analyse statistique est nécessaire : la base de données. L'objectif principal de cette thèse étant de modéliser et d'analyser la dégradation induite par la carbonatation d'un béton armé exposé à son environnement de fonctionnement naturel, nous avons constitué une base de données de carbonatation. Elle est formée de mesures de carbonatation prélevées sur des structures exposées à la carbonatation naturelle.

Ce deuxième chapitre se consacrera à la présentation des données utilisées pour les analyses. Plus précisément, les données proviennent essentiellement d'articles déjà publiés. Nous commencerons par discuter des sources de données utilisées et des informations initiales recueillies sur les matériaux. Ensuite, nous expliquerons les étapes de nettoyage et d'enrichissement de la base de données initiale afin de l'adapter aux modèles linéaires généralisés (GLMs) et aux modèles de dégradation Gamma. La deuxième section sera dédiée aux modifications apportées à la base de données et présentera les caractéristiques de la version finale de la base de modélisation. Enfin, dans la dernière partie de ce chapitre, nous décrirons les ajustements réalisés sur la base de données pour pouvoir appliquer des modèles de durée de vie dans l'analyse du temps de carbonatation.

## 2.2 Présentation de la base de données initiale et des caractéristiques environnementales

### 2.2.1 Sources de données et description des variables

La base de données initiale est constituée de 811 mesures de carbonatation, principalement extraites des bases de données présentes dans divers articles de la littérature.

Plus précisément, elle est constituée en fusionnant de bases de données provenant majoritairement de ALEXANDER, MACKECHNIE et YAM 2007 ; EKOLU et van EDE 2014 ; FIGUEIREDO et NEPOMUCENO 2005 ; GUIGLIA et TALIANO 2013 ; LAM 1998 ; LIANG et al. 2013 ; LIISMA et al. 2017 ; NEVES et al. 2013 ; RIZVI 2017 ; TERUZZI et al. 2003 ainsi que de REVIRON et URRUTIA 2016. Les autres données de carbonatation proviennent des murs d'une structure située à Paris (nous désignons par *Parisdata* cette dernière source de données).

Cette base de données initiale ainsi constituée comprend des mesures de carbonatation provenant de structures en béton de divers types, tels que des ponts, des viaducs, des murs et des piliers. Ces structures sont situées dans différents pays à travers le monde, notamment en Afrique du Sud, au Brésil, en Chine, en France, en Indonésie, en Inde, au Portugal, en Italie et à Taiwan. Au total, les données proviennent de 22 localités, comprenant des villes et des départements, réparties dans ces 9 pays. Le Tableau 2.1 présente une vue d'ensemble des localités incluses dans notre base de données. Il est important de noter que la moitié des structures recensées est localisée en France, et que la majorité se trouve en Europe.

Comme le montre le Tableau 2.2, les informations recueillies sur les structures diffèrent d'une source à une autre à l'exception des informations telles que :

- L'âge de la structure exprimé en année [année],
- La profondeur de carbonatation en millimètre [mm] du matériau au moment de l'observation,
- L'épaisseur de l'enrobage en millimètre [mm],
- La résistance mécanique moyenne du matériau en Méga-Pascal [MPa],
- La localisation de la structure,

## 2.2. Présentation de la base de données initiale et des caractéristiques environnementales

	Nombre de localités	Noms des localités
Afrique du Sud	1	Johanesbourg
Brasil	1	Brasilia
Chine	1	Pékin
France	11	Paris, Beaulieu-sur-loire(LL), Amberieu-en-BUGEY (GB), cattenom (MN), chooz (ZH), civaux (XV), Cruas (SR), Yvelines (RR), golfech (HC), Nogent-sur-seine (TN), Champagne-ardenne (TT).
Inde	1	Bhopal
Indonésie	1	Jakarta
Italie	1	Rome
Portugal	3	Lisbonne, Setubal, Cascais
Taiwan	1	Seoul

**Table 2.1** – Répartition des localités incluses dans la base de données selon les pays

— L'information selon laquelle le matériau ausculté est protégé ou non de la pluie.

	TERUZZI et al. 2003	LISMA et al. 2017	FIGUEIREDO et NEFOMUCENO 2005	LIANG et al. 2013	ÉKOLU et van EDE 2014	LAM 1998	GUGLIA et TALLANO 2013	RIZVI 2017	NÉVES et al. 2013	ALEXANDER, MACKECHIE et YAM 2007	REVIRON et URRUTIA 2016	Paris/Data
Lieu de fonctionnement	X	X	X	X	X	X	X	X	X	X	X	X
Âge de la structure	X	X	X	X	X	X	X	X	X	X	X	X
Profondeur de carbonatation	X	X	X	X	X	X	X	X	X	X	X	X
L'épaisseur de l'enrobage	X	X	X	X	X	X	X	X	X	X	X	X
Résistance à mécanique	X	X	X	X	X	X	X	X	X	X	X	X
Exposé ou pas à la pluie	X	X	X	X	X	X	X	X	X	X	X	X
Type de structure	X	X	X	X	X	X	X	X	X	X	X	X
La porosité										X	X	
Type de ciment		X							X			
Classe du béton		X										
Constituants de la pâte	X											

**Table 2.2** – Récapitulatif des informations fournies par chaque source de données de carbonatation

### 2.2.2 Description des caractéristiques environnementales

Les caractéristiques de l'environnement d'exposition ont une influence considérable sur la carbonatation. Cependant, ces caractéristiques ne sont pas fournies dans les sources de données. Par conséquent, pour chaque structure prise en compte, nous avons obtenu des informations sur les caractéristiques environnementales auprès de la station météorologique la plus proche. Plus précisément, nous avons recueilli, depuis la date de construction de chaque structure jusqu'à l'année 2021, les valeurs moyennes annuelles de la température, de l'humidité relative, ainsi que le nombre moyen annuel de jours de pluie. En ce qui concerne la pression atmosphérique de CO<sub>2</sub>, à notre connaissance, seules les valeurs moyennes annuelles de la période 2003-2021 sont disponibles pour les localités prises en compte. Par conséquent, ce sont ces valeurs que nous avons collectées.

Par ailleurs, toutes ces données environnementales proviennent du site *Climate data store* (<https://cds.climate.copernicus.eu/cdsapp#!/home>) du programme spatial de l'union européenne.

Les Figures 2.2, 2.1, 2.3 présentent les valeurs minimales, moyennes et maximales des caractéristiques pour chaque localité considérée dans notre base de données. Il est intéressant de noter que les valeurs de la pression atmosphérique de CO<sub>2</sub> sont relativement similaires d'une localité à l'autre, indiquant une certaine uniformité de cet aspect à l'échelle mondiale.

En ce qui concerne les conditions climatiques, nous observons des variations significatives entre les différentes localités. Par exemple, Bhopal en Inde se distingue comme la ville ayant

les températures les plus élevées parmi toutes les localités étudiées. Pékin en Chine, quant à elle, est la localité la plus sèche, tandis que Taipei à Taiwan est celle où il pleut le moins. À l’opposé, Jäktar (Estonia) se démarque en tant que ville la plus froide, la plus humide et celle où il pleut le plus abondamment.

Ces variations climatiques entre les localités reflètent la diversité des conditions environnementales dans lesquelles les structures en béton armé, soumises à la carbonatation, sont exposées. Cela permettra de mieux comprendre et de modéliser l’impact de ces caractéristiques climatiques sur la dégradation du béton par la carbonatation.

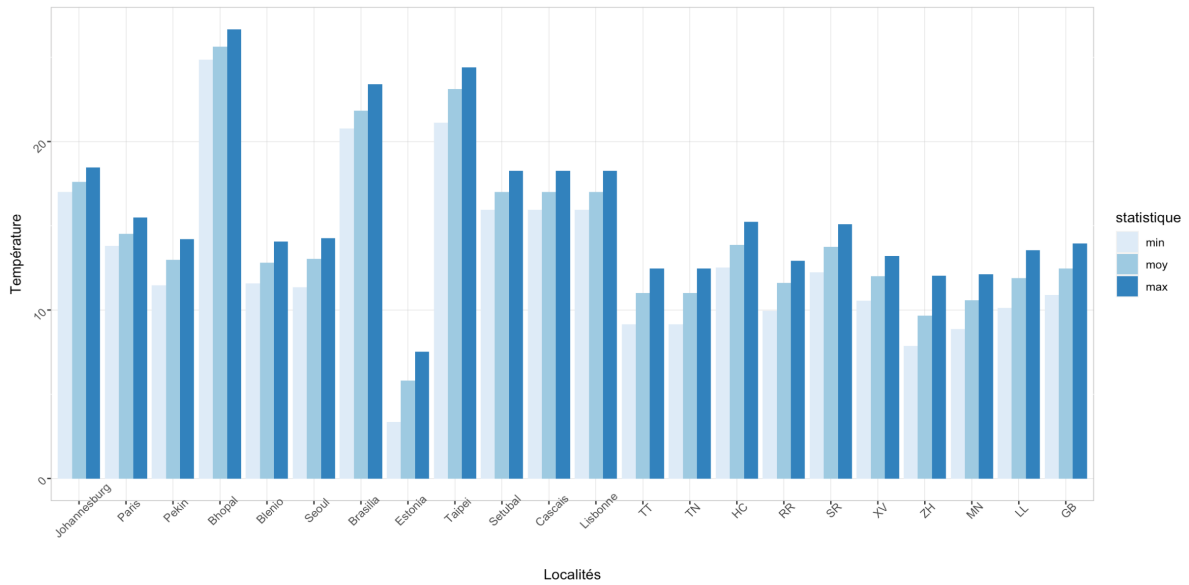


Figure 2.1 – Valeurs minimale, moyenne et maximale de la température de chaque localités

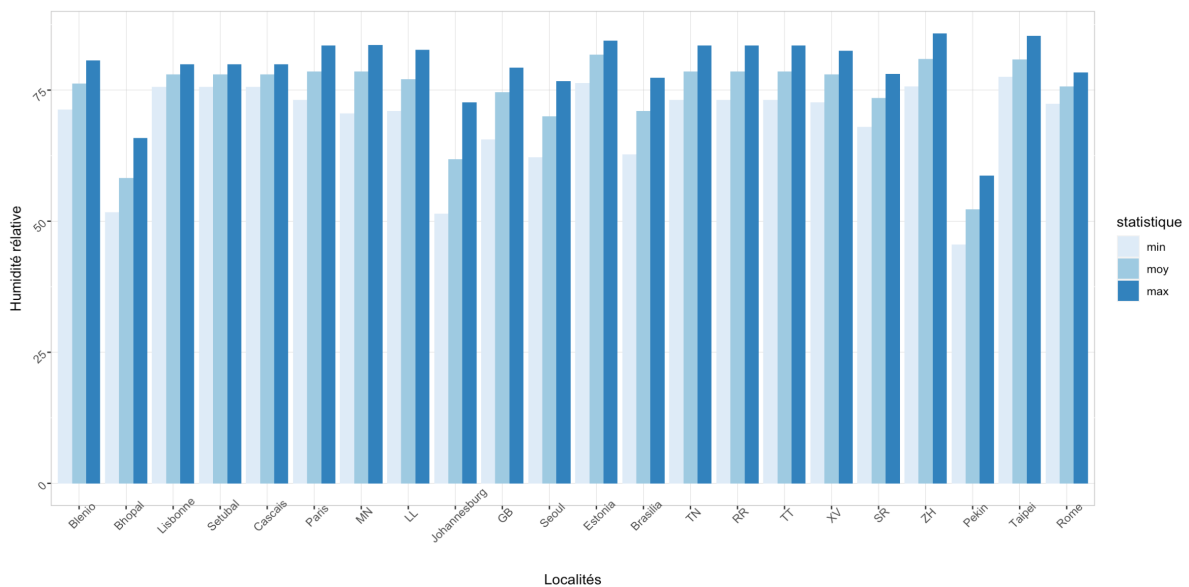
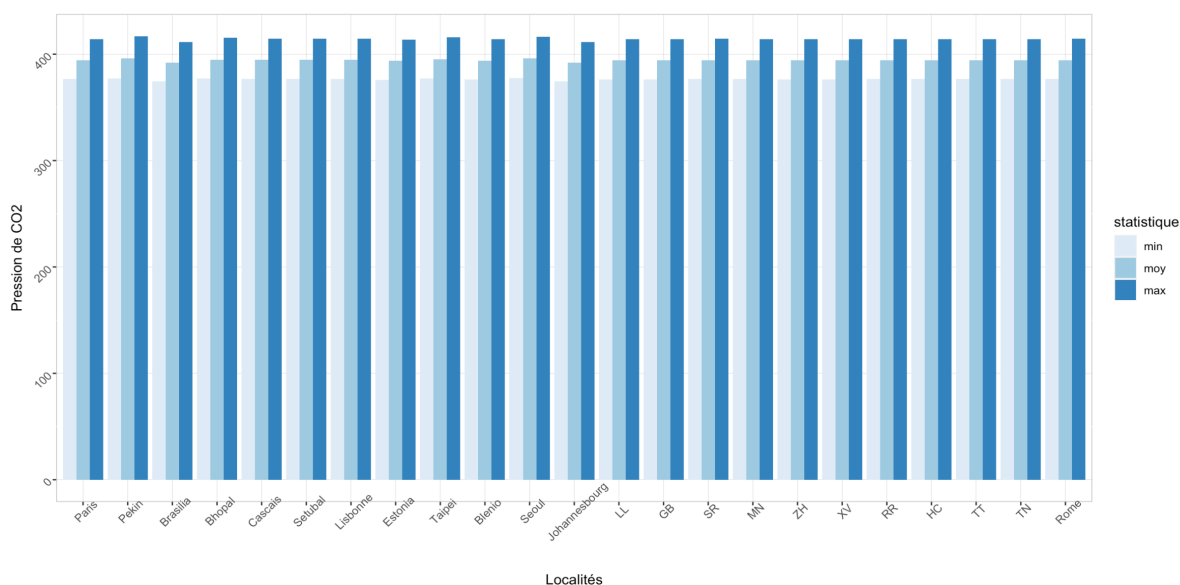


Figure 2.2 – Valeurs minimale, moyenne et maximale de l’humidité relative de chaque localités

Dans la suite de notre étude, nous avons constitué deux bases de données distinctes : une pour l’analyse de la profondeur de carbonatation et une autre pour l’analyse du temps de carbona-



**Figure 2.3** – Valeurs minimale, moyenne et maximale de la Pression atmosphérique de CO<sub>2</sub> de chaque localités

tation. Dans les sections suivantes, nous détaillerons les modifications et les enrichissements apportés à ces bases de données, ainsi que les caractéristiques des bases finales obtenues.

## 2.3 La base de données de la modélisation de la profondeur de carbonatation

### 2.3.1 Description générale

Après avoir mené une étude bibliographique sur les modèles de carbonatation, nous avons opté pour deux approches afin de modéliser la profondeur de carbonatation d'un matériau exposé à la carbonatation naturelle : les modèles linéaires généralisés (GLM) et les modèles de dégradation Gamma. Pour mettre en œuvre ces approches, nous avons créé une base de données à partir de la base initiale présentée ci-dessus, comprenant un total de 793 points de mesure. Nous avons exclu les structures dont l'âge était inférieur à un an. Cette base de données obtenue se compose de 8 variables : 7 variables explicatives dérivées des informations communes fournies par différentes sources de données, ainsi que la profondeur de carbonatation, qui constitue notre variable cible.

Les variables explicatives sont constituées de caractéristiques du matériau et d'informations sur le climat des lieux d'exposition. Les détails sur ces variables explicatives et leur signification seront présentés dans la section suivante.

### 2.3.2 Présentation de la variable cible : la profondeur de carbonatation

Comme indiqué précédemment, nous cherchons à modéliser la profondeur de carbonatation du matériau que nous désignons par  $Y$ , en fonction de ses autres caractéristiques. Les valeurs de profondeurs de carbonatation (en millimètre [mm]) de cette base se situent dans une plage allant de 1 mm à 60 mm. La valeur moyenne de la profondeur de carbonatation est de 15.94 mm, tandis que la valeur médiane est de 12.75 mm.

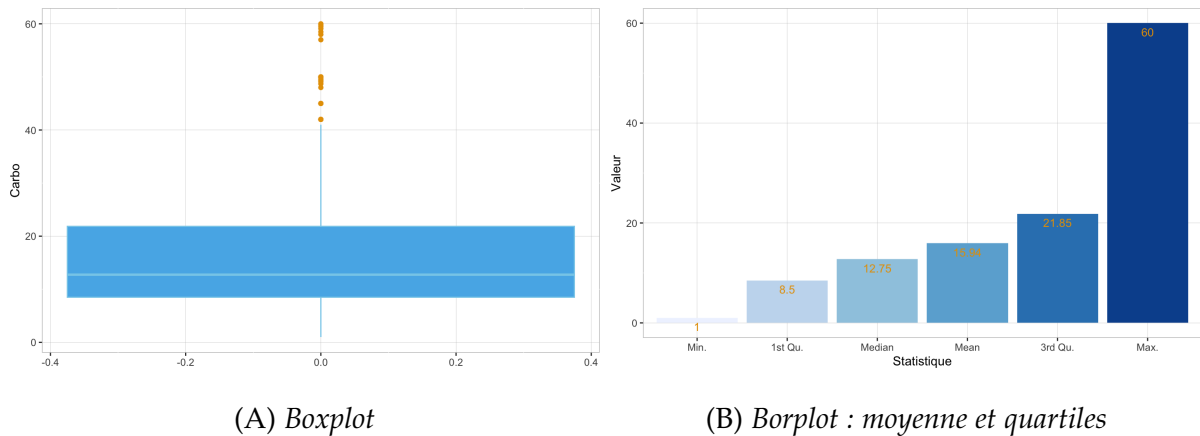


Figure 2.4 – Statistiques descriptives de la variable cible (profondeur de carbonatation)

### 2.3.3 Présentation des variables explicatives

Dans notre étude, nous avons retenu deux types de variables explicatives pour la modélisation de la profondeur de carbonatation : les variables du matériau, issues de la base de données originale, et les variables globales. En effet, cette étude ayant été inspirée par le modèle de la Fédération internationale du béton (*fib*), nous avons considéré les variables  $k_c$  (facteur de cure),  $W_t$  (facteur de la pluviométrie),  $k_e$  (facteur de l'humidité relative) et  $C_s$  (concentration atmosphérique de  $\text{CO}_2$ ) qui sont également utilisées dans ce modèle. Ces variables seront détaillées ultérieurement.

#### 2.3.3.1 L'âge $t$ du matériau

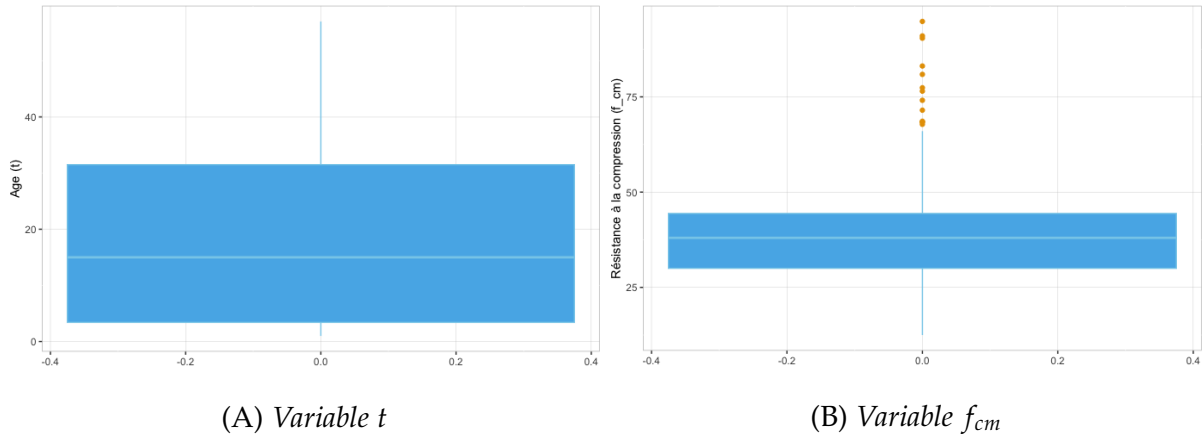
L'âge du matériau est une des caractéristiques du matériau les plus accessibles. Cette donnée, présente dans la base originale, est une caractéristique liée à la valeur de la profondeur de carbonatation. En effet, la solution analytique de la première loi de Fick montre que la profondeur de carbonatation d'un matériau est proportionnelle à la racine carrée de l'âge de ce dernier. Ainsi, l'âge du matériau est retenu pour la modélisation de la profondeur de carbonatation. Dans notre base de données, l'âge des matériaux varie entre 1 et 57 ans. La médiane de l'âge est de 15 ans et la moyenne est de 18,25 ans (voir la boîte à moustache (A) à gauche de la Figure 2.5).

#### 2.3.3.2 La résistance moyenne à la compression $f_{cm}$

Comme nous l'avons mentionné précédemment, la teneur en ciment du matériau influence la profondeur de carbonatation. Cependant, obtenir des informations précises sur les teneurs en ciment par des expérimentations peut être complexe et imprécis. En revanche, la résistance à la compression, qui est également liée à la teneur en ciment du béton (voir V. PAPADAKIS et DEMIS 2011), est une donnée systématiquement disponible dans les sources de données. Par conséquent, nous utilisons la résistance moyenne à la compression, représentée par la variable  $f_{cm}$ , comme une variable explicative pour modéliser la profondeur de carbonatation. Les valeurs de résistance mécanique des matériaux considérés se situent dans l'intervalle [12.50;98.80] avec une valeur médiane de 38.74 MPa (voir le graphique en boîte à moustaches (B) de la Figure 2.5).

#### 2.3.3.3 Le paramètre de cure $k_c$

Le processus de cure peut influencer la carbonatation du béton armé. Une cure insuffisante peut entraîner une détérioration de la couche de surface, favorisant ainsi la carbonatation



**Figure 2.5** – Boîtes à moustaches (boxplots) des facteurs internes au matériau : l'âge  $t$  et la résistance moyenne à la compression  $f_{cm}$ .

(BALAYSSAC et al. 1994). Le modèle *fib* intègre l'effet du temps de cure du matériau dans la modélisation de la carbonatation en utilisant le paramètre aléatoire  $k_c$  (sans unité [-]) défini par l'équation (2.1). Ce paramètre aléatoire est caractérisé par le paramètre  $b_c$ , dont les valeurs sont générées à partir d'une distribution normale avec une moyenne de  $m = -0.567$  et un écart type de  $sd = 0.024$  (CEB-FIP 2010).

$$k_c = \left( \frac{t_c}{7} \right)^{b_c}, \quad (2.1)$$

où  $t_c$  [jour] représente le temps de cure. Dans notre analyse, les valeurs de  $k_c$  sont calculées en utilisant  $b_c = m = -0.567$ . Ce paramètre est donc uniquement une fonction du temps de cure  $t_c$ .

Il faut noter que les valeurs des temps de cure ne sont pas systématiquement fournies par les sources de données. Par conséquent, la valeur du temps de cure de chaque matériau est déterminée en suivant les recommandations de la norme **NF EN 13670** (13670 2013). Cette norme recommande des valeurs minimales de temps de cure du béton en fonction de la température en surface et de la valeur de sa montée en résistance  $r$ . La montée en résistance  $r$  est le rapport de la résistance moyenne à la compression à 2 jours sur la résistance moyenne à la compression à 28 jours, déterminé par des essais préalables ou basé sur l'expérience d'un béton de composition similaire.

Dans notre analyse, nous avons supposé que la valeur de la montée en résistance  $r$  de chaque matériau se situe dans l'intervalle  $[0.3, 0.5]$ . En utilisant la Table 2.3 fourni par la norme **NF EN 13670**, nous avons déterminé la valeur du temps de cure  $t_c$  en fonction de la température au moment de la mise en place du béton. Ainsi, pour chaque matériau, la valeur du temps de cure a été déterminée en se référant aux recommandations de la norme et en tenant compte de la montée en résistance  $r$  estimée. La boîte à moustache des valeurs de  $k_c$  obtenues est représentée à la Figure 2.6 (graphique (A)). Ces valeurs se situent entre 1.28 et 2.03, avec une valeur médiane de 1.61.

#### 2.3.3.4 Le paramètre de cure $k_{cfmt}$

Le paramètre de cure  $k_c$  exprime simplement l'influence de la durée de la cure. Cependant, selon les travaux de Balayssac BALAYSSAC et al. 1994, l'effet de la cure dépend également de la composition du béton, telle que la résistance à la compression à 28 jours, le rapport E/C, etc. Par conséquent, nous avons également considéré le paramètre de cure proposé par KARI

Température de la surface du béton (t) en °C	Durée minimale de cure, jours		
	Evolution de la résistance du béton ( $f_{cm2}/f_{cm28} = r$ )		
	Rapide ( $r \geq 0,50$ )	Moyenne ( $0,50 > r \geq 0,30$ )	Lente ( $0,30 > r \geq 0,15$ )
$t \geq 25$	1,0	1,5	2,5
$25 > t \geq 15$	1,0	2,5	5
$15 > t \geq 10$	1,5	4	8
$10 > t \geq 5$	2,0	5	11

**Table 2.3** – Valeurs minimales de la durée de cure d'un béton en fonction de la température à la surface et de la valeur de sa montée en résistance extrait de 13670 2013.

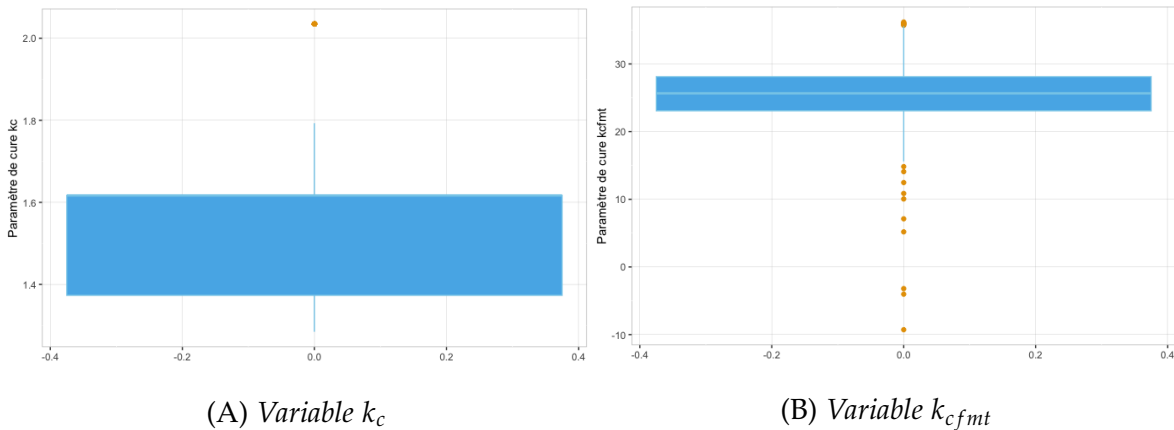
2011, que nous appelons  $k_{cfmt}$ . Ce dernier inclut, en plus de la durée de la cure, la valeur de la résistance à 28 jours  $f_{cm28}$  du béton, et il est défini par la formule suivante :

$$k_{cfmt} = \frac{1.9 \cdot 10^{-2}}{10^{-0.025 f_{cm28}}} + \left(1 - \frac{1.9 \cdot 10^{-2}}{10^{-0.025 f_{cm28}}}\right) \sqrt{\frac{28}{0.01 t_c}} \quad (2.2)$$

avec  $f_{cm28}$  donné en MPa et  $t_c$  en jour. La valeur de  $f_{cm28}$  est déterminée à partir de la valeur de la résistance à la compression  $f_{cm}$  à l'âge  $t$  à l'aide de la relation (2.3) proposée dans CEB-FIP 2010.

$$f_{cm28} = e^{-0.25 \left(1 - \sqrt{\frac{28}{t}}\right)} f_{cm} \quad (2.3)$$

où  $t$  est l'âge du béton exprimé en jour. Il est intéressant de remarquer que pour les valeurs élevées de la résistance à 28 jours, les valeurs du paramètre  $k_{cfmt}$  sont généralement très petites, voire négatives. En revanche, on observe l'opposé lorsque les valeurs de la résistance à 28 jours sont faibles. Cela est illustré dans le boxplot (b) de la Figure 2.6



**Figure 2.6** – Boîtes à moustaches (boxplots) des paramètres de cure.

### 2.3.3.5 Le paramètre environnemental $k_e$

La teneur en humidité relative ( $Rh$ ) de l'environnement d'exposition joue un rôle important dans l'intrusion du  $CO_2$  dans les pores du béton. Selon les recherches de BALEN et GEMERT 1994, la carbonatation est optimale lorsque le taux d'humidité relative est maintenu à environ 65%. À cette valeur, la diffusion du  $CO_2$  n'est pas significativement entravée par l'humidité, mais elle permet néanmoins une dissolution rapide des hydrates dans la solution interstitielle.



Le paramètre  $k_e$  est introduit dans le modèle *fib* (CEB-FIP 2010) pour prendre en compte l'influence de l'humidité relative sur le processus de carbonatation. Il est défini par l'équation (2.4) suivante.

$$k_e = \left( \frac{1 - \left(\frac{Rh}{100}\right)^{f_e}}{1 - \left(\frac{Rh_{ref}}{100}\right)^{f_e}} \right)^{g_e}, \quad (2.4)$$

où  $Rh$  est l'humidité relative du lieu et  $Rh_{ref} = 65\%$  est prise comme valeur de référence de l'humidité relative. les quantités  $f_e$  et  $g_e$  sont des paramètres de régression tels que :  $f_e = 5$  et  $g_e = 2.5$  CEB-FIP 2010. Pour calculer les valeurs de  $k_e$ , nous avons pris comme valeur de  $Rh$ , la valeur médiane sur toute la période d'observation (i.e de l'année d'observation à l'année 2021). En se référant à l'équation (2.4), on peut conclure que les structures pour lesquelles la valeur de  $k_e$  est inférieure à 1 correspondent aux endroits où la teneur en humidité relative ( $Rh$ ) est supérieure à la valeur de référence  $Rh_{ref}$ . En revanche, si la valeur de  $Rh$  du lieu d'exposition est inférieure à  $Rh_{ref}$ , alors  $k_e$  est supérieur à 1. Les valeurs obtenues pour le paramètre  $k_e$  varient entre 0.43 et 1.23, avec une médiane de 0.66 (voir le boxplot (A) des valeurs obtenues représenté dans la Figure 2.7).

### 2.3.3.6 Le paramètre environnemental $W_t$

Le paramètre de pluviométrie  $W_t$  est utilisé dans le modèle de *fib* pour quantifier l'effet des précipitations sur la profondeur de carbonatation du matériau. D'après CEB-FIP 2010, il est calculé selon l'équation :

$$W_t = \left( \frac{t_0}{t} \right)^{(p_{SR} \cdot ToW)^{b_w}} \quad (2.5)$$

où :

- $t_0 = 0.0767$  est un temps de référence en années ;
- $p_{SR}$  est la probabilité que le matériau soit mouillé par la pluie pendant les jours de pluie.
- $ToW$  (Time of Wetness) est la proportion de jours de pluie avec une précipitation minimale de  $h_{Nd} = 2.5\text{mm}$ .
- $b_w$  est un paramètre aléatoire. Ses valeurs sont distribuées selon une loi de distribution normale de paramètre moyen  $m = 0.0446$  et d'écart type  $sd = 0.163$ .

La valeur de  $p_{SR}$  est déterminée en fonction des propriétés géométriques de la structure et de la direction du vent pendant les jours de pluie. Dans notre étude, nous avons fait l'hypothèse que tous les matériaux non abrités sont mouillés pendant les jours de pluie, ce qui correspond à  $p_{SR} = 1$  pour ces matériaux, et  $p_{SR} = 0$  pour les matériaux protégés de la pluie.

En ce qui concerne la valeur de  $ToW$ , nous l'avons déterminée en utilisant le rapport entre le nombre médian de jours de pluie sur la période d'observation et le nombre total de jours dans une année (soit 365 jours).

Quant à la valeur de  $b_w$ , nous avons supposé qu'elle est égale à  $m = 0.0446$  pour notre étude.

Le Boxplot (B) des valeurs obtenues, représenté dans la Figure 2.7, montre que toutes les valeurs de  $W_t$  sont naturellement inférieures à 1 pour les structures non abritées. En revanche, pour les matériaux abrités, qui représentent 46.6% des points de mesure de la base de données, la valeur de  $W_t$  est égale à 1.

### 2.3.3.7 Le paramètre environnemental $C_s$

L'effet du  $\text{CO}_2$  sur la valeur de la profondeur de carbonatation est modélisé dans le modèle de la *fib* à travers le paramètre  $C_s$  exprimé [ $\text{kg}/\text{m}^3$ ] qui est tel que :

$$C_s = C_{s.atm} + C_{s.emi.} \quad (2.6)$$

où  $C_{s.atm}$  [ $\text{kg}/\text{m}^3$ ] est la concentration en  $\text{CO}_2$  du lieu d'exposition et  $C_{s.emi.}$  [ $\text{kg}/\text{m}^3$ ] une concentration supplémentaire due aux sources d'émission. D'après CEB-FIP 2010, la valeur est égale à  $0.00049\text{kg}/\text{m}^3$  et est utilisée pour les structures confinées telles que les tunnels. Pour notre étude, nous avons considéré  $C_s = C_{s.atm}$  qui correspond à la médiane des valeurs de concentrations atmosphériques annuelles de  $\text{CO}_2$  (converties en [ $\text{kg}/\text{m}^3$ ]) observées sur la période 2003-2021. Le Boxplot des valeurs obtenues est également représentée à la Figure 2.7 (c).

## 2.4 La base de données de la modélisation du temps de carbonatation

### 2.4.1 Présentation générale

Dans le cadre de notre étude sur le temps de carbonatation des matériaux en béton armé, nous avons appliqué des modèles de durées de vie à notre base de données. Comme décrit dans le Chapitre 1 (Section 1.5.3), les modèles de régression en survie permettent de modéliser l'effet des caractéristiques des matériaux qui dépendent du temps. Afin de pouvoir appliquer ces modèles, nous avons adapté la base de données initiale.

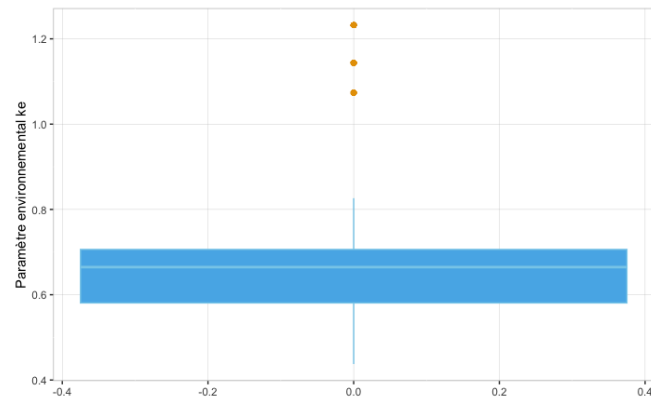
La nouvelle base de données est composée de 793 points de mesure de la profondeur de carbonatation, avec 12 variables présentées dans le Tableau 2.4, que nous détaillerons dans les sections suivantes. Pour éviter toute ambiguïté, nous avons renommé certaines variables. Par exemple, l'âge du matériau est désormais appelé  $X$  pour le distinguer du temps de carbonatation noté  $T$ .

Variable	Description	Unité
$X$	L'âge du matériau	[année]
$Y$	Profondeur de carbonatation à l'âge $X$	[mm]
$C$	L'épaisseur de l'enrobage	[mm]
$\delta$	Indicateur de la carbonatation à l'âge $X$ (1 si le matériau est déjà carbonaté et 1 sinon)	[-]
$f_{cm}$	La résistance moyenne à la compression à $X$	[Mpa]
$t_c$	La durée de la cure	[jour]
$P_{co2}$	Concentration atmosphérique de $\text{CO}_2$	[ppm]
$Rh$	L'humidité relative du lieu d'exposition	[%]
$Tp$	La température du lieu d'exposition	[°C]
$Dp$	Proportion annuelle de jours de pluie	[-]
$Tom$	L'année de l'enquête	[-]

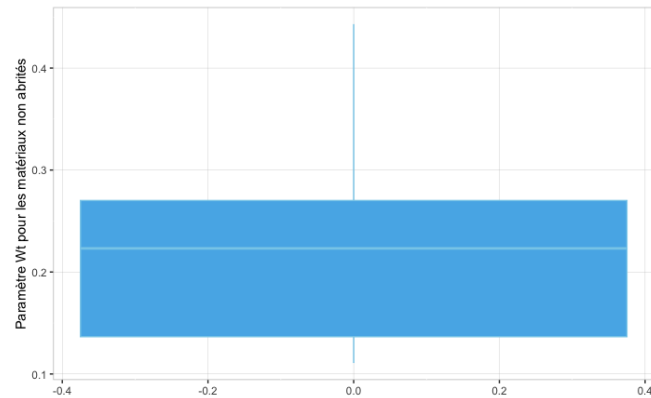
Table 2.4 – Description des variables de la base de données de l'analyse de survie

### 2.4.2 La nouvelle variable cible

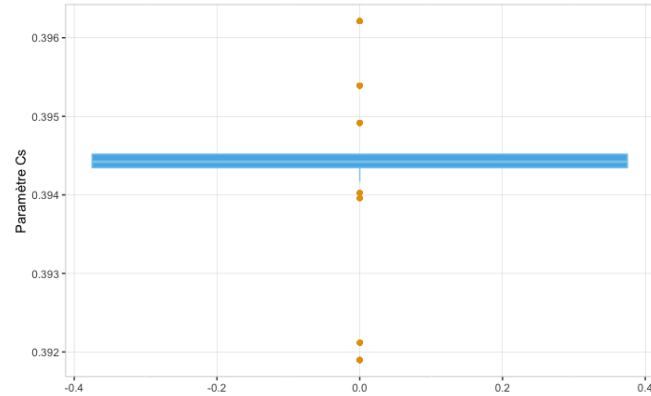
On rappelle que dans cette deuxième approche, nous nous intéressons au temps de carbonatation qui correspond au moment où la valeur de la profondeur de carbonatation désignée par



(A) Variable  $k_e$



(B) Variable  $W_t$



(C) Variable  $C_s$

Figure 2.7 – Boîtes à moustaches (boxplots) des paramètres environnementaux

la variable  $Y$  sera égale à la valeur de l'enrobage  $C$ . Comme mentionné précédemment dans le premier chapitre, ce moment n'est jamais observé : c'est donc une donnée censurée. Ainsi, à partir des variables  $Y$  et  $C$ , nous avons défini la variable de censure  $\delta$  qui prend la valeur 1 si le matériau est déjà carbonaté au moment de l'observation et 0 sinon. Mathématiquement, pour un matériau  $i$  donné, nous avons  $\delta_i = 1$  si  $Y_i > C_i$  et  $\delta_i = 0$  si  $Y_i < C_i$ . Ainsi, l'information disponible sur le temps de carbonatation, qui est notre variable cible, est donnée à travers la variable  $(X, \delta)$ . Comme illustré à la Figure 2.8, la base de données est constituée de 158 mesures carbonatées (soit environ 19.48% des points de mesure) et 653 non carbonatées, soit environ 80.52%.

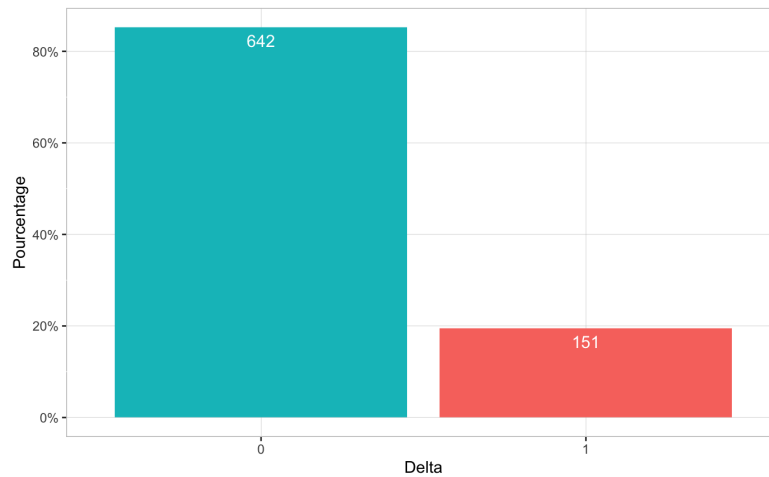


Figure 2.8 – Proportions des valeurs de la variable  $\delta$ .

### 2.4.3 Prise en compte des variables explicatives dans les modèles de durée de vie

Dans notre étude, nous avons analysé l'effet des facteurs suivants sur le temps de carbonatation, en nous concentrant sur les caractéristiques facilement observables par un ingénieur génie civil :

1. Enrobage  $C$  [mm] : il s'agit de la couche de béton recouvrant l'armature. Une valeur plus élevée d'enrobage implique un temps de carbonatation plus long.
2. Résistance moyenne à la compression  $f_{cm}$  [MPa] : ce paramètre, déjà mentionné, est fourni par toutes nos sources de données et reflète la qualité globale du matériau.
3. Temps de cure  $t_c$  [jours] : bien que cette information ne soit pas systématiquement fournie, elle peut être facilement obtenue en suivant les recommandations de la norme **NF EN 13670**, comme expliqué dans la section 2.3.3.3.
4. Température du lieu d'exposition  $T_p$  [°C] : cette donnée peut être observée à l'aide des stations météorologiques les plus proches du lieu d'exposition. Elle représente une caractéristique environnementale qui peut influencer le temps de carbonatation.
5. Humidité relative du lieu  $Rh$  [%] : l'humidité relative est l'un des facteurs environnementaux les plus importants. Les valeurs de  $Rh$  peuvent être obtenues à partir des stations météorologiques les plus proches.
6. Proportion annuelle de jours de pluie  $Dp$  [-] : cette variable permet d'évaluer l'effet de la pluviométrie sur le temps de carbonatation.
7. Pression atmosphérique de  $CO_2$  ( $P_{co2}$  en [ppm]) : la concentration de  $CO_2$  dans l'atmosphère est également l'un des principaux facteurs de la carbonatation. Des concentrations plus élevées de  $CO_2$  peuvent accélérer le processus de carbonatation.

L'objectif de notre étude était d'identifier, parmi ces 7 caractéristiques, celles qui ont une influence significative sur le temps de carbonatation.

Afin de pouvoir prendre en compte la variabilité temporelle de certaines caractéristiques, en utilisant toutes les informations recueillies, nous avons constituée une base de type longitudinale. Il s'agit d'une base dans laquelle un individu (point de mesure) apparaît autant de fois que ses informations sont mises à jours.

Rappelons que pour chaque matériau considéré, nous avons collecté, chaque année de la date de construction à 2021, les valeurs des caractéristiques environnementales. Cet intervalle d'une année se justifie par le fait que le processus de carbonatation, dans les conditions naturelles,

est très lent. Ainsi, l'hypothèse selon laquelle l'évolution de la carbonatation entre deux années successives est négligeable ne semble donc pas aberrante.

La base de donnée, ci-dessus décrite a donc été modifiée afin de l'adapter aux modèles de régression en survie présentés à la au Chapitre 1.1. Cette modification a été faite selon l'Algorithme 1 ci-dessous.

La Table 2.5 donne 7 lignes extraites de la base obtenue. Les trois premières lignes correspondent à la représentation des informations recueillies sur un matériau âgé de 3 ans et localisé à Rome. Les quatre dernières lignes quand à elles donnent résumé les informations recueillies sur un matériau vivant à Bhopal en Inde. Comme on peut le constater, la valeur 0 est assignée à la variable  $Dp$  pour les matériaux protégés de la pluie (slt= oui).

La base de données décrite ci-dessus a été modifiée afin de l'adapter aux modèles de régression en survie présentés au Chapitre 1.1. Cette adaptation a été réalisée en suivant l'Algorithme 1 ci-dessous et la Table 2.5 présente 7 lignes extraites de la base obtenue. Les trois premières lignes correspondent à la représentation des informations recueillies sur un matériau âgé de 3 ans et localisé à Rome. Les quatre dernières lignes, quant à elles, résument les informations recueillies sur un matériau se trouvant à Bhopal en Inde. Comme on peut le constater, la valeur 0 est assignée à la variable  $Dp$  pour les matériaux protégés de la pluie (slt = oui).

---

**Algorithme 1 :** Construction de la base de donnée adaptée aux modèles de régression en survie

---

**Entrées :** Les informations de la base décrite à la Table 2.4

**Sorties :** La base de donnée pour l'analyse du temps de carbonatation

**Initialisation :** Créer une base d'une ligne avec 11 colonnes nommées respectivement Id, X, delta,  $t_c$ , C,  $f_{cm}$ , Rh, Tp,  $P_{co_2}$ , Dp et  $X_2$  ;

N= nombre de lignes de la base d'entrée ;

**tant que** ( $i \leq N$ ) **faire**

toc=tom<sub>i</sub>-X<sub>i</sub> (correspond à la date de construction du matériau i);

j = 1;

**tant que** ( $1 \leq j \leq X_i$ ) **faire**

Rajouter une ligne à la base d'entrée de telle sorte que :

Id=i ;

delta=delta<sub>i</sub> ;

C=C<sub>i</sub>;

$f_{cm} = f_{cm_i}$  ;

$X_2 = j$  (l'âge du matériau à la  $j^{i\text{me}}$  visite) ;

t=toc+j (l'année correspondante à la  $j^{i\text{me}}$  visite);

Rh=Rh(t) (valeur moyenne de Rh pour l'année t);

Tp = Tp(t) (valeur moyenne de Tp pour l'année t) ;

Dp = Dp(t) (valeur moyenne de Dp pour l'année t) ;

$P_{co_2} = P_{co_2\text{median}}$  (valeur médiane sur la période 2003-2021 en ppm);

j = j + 1

**fin**

i = i + 1

**fin**

---

## 2.5 Conclusion

Dans ce deuxième chapitre, nous avons présenté les bases de données indispensables pour analyser la carbonatation (profondeur de carbonatation et temps de carbonatation). Ces bases

id	X	t_c	f_cm	delta	C	Tp	Rh	Pco2	Dp	Slt	ToM	X2	Location
221	3.19	3	37.0	0	29.89	18.01	76.82	393.26	0.33	Non	2013	1	Rome
221	3.19	3	37.0	0	29.89	17.77	73.10	393.26	0.25	Non	2013	2	Rome
221	3.19	3	37.0	0	29.89	18.14	74.15	393.26	0.32	Non	2013	3	Rome
457	4.00	2	32.5	0	30.00	25.34	63.60	400.05	0.00	Oui	2017	1	Bhopal
457	4.00	2	32.5	0	30.00	25.79	65.07	400.05	0.00	Oui	2017	2	Bhopal
457	4.00	2	32.5	0	30.00	25.74	59.66	400.05	0.00	Oui	2017	3	Bhopal
457	4.00	2	32.5	0	30.00	26.14	60.06	400.05	0.00	Oui	2017	4	Bhopal

**Table 2.5** – Extraction de 7 lignes de la base de données de modélisation du temps de carbonatation

de données sont constituées de mesures de carbonatation provenant de structures en béton armé localisées dans lieux avec des climats variés. Ces bases sont utilisées pour l'étude respective de la profondeur de carbonatation dans le chapitre 3 et le temps de carbonatation dans le chapitre 4. Notons que les variables explicatives sélectionnées pour les bases de données sont des informations facilement disponibles ou accessibles par les ingénieurs, ce qui donne une grande portée pratique aux modèles que nous allons ensuite proposer.

## Engineering statistical models for carbonation depth

---

Ce troisième chapitre, rédigé en anglais, représente la version publiée dans la revue *European Journal of Environmental and Civil Engineering* d'une analyse statistique de la carbonatation. Une analyse statistique de la profondeur de carbonatation est réalisée, et deux modèles de profondeurs de carbonatation sont proposés. Le premier modèle repose sur la régression linéaire généralisée (voir la Section 1.5.1 du Chapitre 1), et le second modèle est basé sur les processus de dégradation Gamma (voir la Section 1.5.2 du Chapitre 1). Ces modèles sont ajustés sur un ensemble de données constitué de mesures de carbonatation provenant de structures exposées à la carbonatation naturelle.

### 3.1 Introduction

When exposed to atmospheric conditions, reinforced concrete structures undergo deterioration over time, mainly due to corrosion of the reinforcing steel bars embedded in the concrete. This deterioration lessens the serviceability and safety of the r.c structures. Ingress of chloride and carbonation are the main causes of corrosion. According to Tuuti's works (TUUTTI 1992b), corrosion process is a process that takes place in two stages : the corrosion initiation stage and the corrosion propagation stage. Our study is pertains to the first stage.

Carbonation is unavoidable for most structures (except underwater ones), whereas chloride ingress is associated with proximity to the sea, the use of de-icing salts or a particular industrial environment. The present study is focused on carbonation.

Carbonation is a chemical reaction between carbon dioxide gas in its ionic aqueous dissolved form, and the hydrates of the cement paste. The atmospheric carbon dioxide gas ( $CO_2$ ) penetrates the porous network of the concrete and reacts with the calcium hydroxide (Portlandite) (J et F. P. GLASSER 1991-1992; OLLIVIER et VICHOT 2008c). This reaction produces water and calcium carbonate that lower the pH of the pore solution. The drop of pH activates the initiation stage of the corrosion process. Thus, the level of deterioration of r.c structures induced by carbonation can be related to the carbonated part of concrete cover, which constitutes the protective barrier against  $CO_2$  ingress.

Knowledge of the carbonation depth at a given time during the operation of the concrete structural assets is highly desirable for the owners and managers in that it allows corrective maintenance to be started. Substantial endeavours have been made in the literature to explain and model carbonation based on physical considerations. Among the modelling approaches, carbonation models can be divided into three categories : numerical models, empirical models and analytical models

In the numerical models, authors try to describe the carbonation process as precisely as possible by integrating all the physical and chemical phenomena (OLLIVIER et VICHOT 2008c). Naturally, they need a large number of input parameters for the computation. Due to the diffusive nature of carbonation, with a fairly high rate of consumption compared to the ingress velocity, the carbonation depth,  $X_c$ , can generally be related to the exposure age,  $t$ , (OLLIVIER et VICHOT 2008d) by :

$$X_c = At^B, \quad (3.1)$$

where  $A$  and  $B$  are calibration parameters.

The authors of empirical modelling use data to estimate calibration parameters. The wetting-drying effect, which has an important impact on the carbonation level, is disregarded in these models.

An analytic solution of the Fick's first law can be derived by assuming that carbonation is only caused by the diffusion of carbon dioxide gas within the concrete, which immediately reacts with the portlandite and other hydrated calcium carbonates. By comparison with some experimental studies of indoor and outdoor carbonation, it can be concluded that the depth of carbonation  $X_c$  is reasonably proportional to the square root of the exposure age,  $t$ , of the concrete :

$$X_c = K\sqrt{t}, \quad (3.2)$$

where  $K$  is the carbonation coefficient. This coefficient is a function of factors that influence the carbonation process. Three categories of factors are identified :

- Material factors (water to cement ratio, amount of hydrates, 28-day compressive strength);



- Curing factors (curing conditions and curing time);
- Environmental factors (relative humidity, temperature, amount of CO<sub>2</sub> in the ambient air).

Various semi-empirical models based on equation (5.4) have been developed. They differ from each other in the expression of the carbonation coefficient  $K$ . By carrying out experimental studies on concrete microstructure and by resolving a complex nonlinear system of differential equations, Papadakis V.G. et al. 1991 was one of the first to propose an expression of  $K$  that is related only to the concrete microstructure and CO<sub>2</sub> concentration. To incorporate relative humidity and temperature in carbonation modelling, Bakker BAKKER 1994 later proposed a carbonation model on the same basis, which additionally included the wetting-drying cycles of concrete by considering the diffusion to be hindered in moist concrete.

Several other models for  $K$  such as ATİŞ 2004; CEB-FIP 2010; LIU et al. 2008; P.HILL et al. 2015; TAFFESE, E. SISTONEN et PUTTONEN 2015; TAFFESE, Esko SISTONEN et PUTTONEN 2015 have been developed using mathematical tools. A model based on equation 5.4 was elaborated in the framework of the *fib-Model Code 2010* in CEB-FIP 2010. This model, inspired by the Duracrete project (**Duracrete**), takes the environmental and execution factors into account as well as the quality of the concrete (see equation (3.3)).

$$X_c(t) = \sqrt{2k_e k_c C_s R_{NAC,0}^{-1} t W(t)}. \quad (3.3)$$

In this equation :

- $X_c(t)$  ([mm]) is the carbonation depth at exposure age  $t$  (in years),
- $k_e$  is an environmental factor ([-]). It is a function of relative humidity;
- $k_c$  is an execution parameter ([-]);
- $C_s$  is CO<sub>2</sub> concentration in the air [ $kg/m^3$ ];
- $R_{NAC,0}^{-1}$  is the inverse carbonation resistance of concrete expressed in  $(mm^2/year)/(kg/m^3)$ ;
- $W(t)$  is the weather function ([-]).

These factors will be described in detail in the Section 3.3.2.

Regarding statistical approaches, some important works are worthy of note. ATİŞ 2004 applies statistical methods to laboratory data to evaluate the influence of compressive strength and porosity on accelerated carbonation depth for fly ash concretes. To expand statistical methods to modelling the carbonation of outdoor exposed concretes, Hills *et al.* (in P.HILL et al. 2015) evaluated multivariate analysis and linear regression methods on natural carbonation data by modelling the carbonation coefficient  $K$ . Some other modern methods, such as machine learning and neural networks have also been used to identify factors affecting the carbonation depth from both laboratory testing and in-situ data TAFFESE, E. SISTONEN et PUTTONEN 2015; TAFFESE, Esko SISTONEN et PUTTONEN 2015.

The models mentioned above can be used to predict the carbonation depth in concrete-but under some specific conditions. Most of these models were calibrated and validated by using laboratory experimental data and their use cannot be simply extended to in-situ exposed structures without a minimum of re-calibration. The models of P.HILL et al. 2015 assume the normality of the target (i.e. the logarithmic transformation of the carbonation coefficient  $K$ ). However, it is mentioned that the logarithmic of the measured carbonation depth :age ratio ( $\log(K) = \log\left(\frac{x_c}{\sqrt{t}}\right)$ ) does not have a sufficiently Gaussian (normal) distribution. In addition, these authors fit two models with different predictors (the first is based on the origin of concrete and its age while the second is focused on the concrete strength) and there is a lack of clarity concerning the conditions under which the first or the second model should be used. For carbonation studies of European r.c structures, the *fib* model is recommended. Inconveniently, laboratory tests are needed to determine input parameters such as  $R_{NAC,0}^{-1}$  for almost any different concrete, so this model somewhat tricky to handle.

Therefore, the main goal of the present work is to build and propose a statistical model that :

- includes the main factors affecting carbonation process;
- is efficient and easy for engineers to use.

We present two carbonation models, one based on generalized linear models and the other on Gamma degradation processes for concretes exposed to natural carbonation. In the second part of this paper, we will present the scope of our study in detail. The two statistical carbonation models proposed are presented by describing the methods applied each in the third section. In the final part of the paper, performances of our proposed models when applied to sheltered r.c structures are compared to those of the *fib* model.

## 3.2 Review on carbonation models

Main models of carbonation are based on the first Fick's law of diffusion. It states that the flux of mass ( $J_x$ ) through the unit area of a section  $x$  is proportional of the concentration gradient  $\partial c/\partial x$  and the diffusion coefficient  $D$  :

$$J_x = -D \cdot \frac{\partial c}{\partial x}, \quad (3.4)$$

where  $c$  is the concentration of the gas or the liquid.

Considering the modeling approaches, carbonation models can be divided into three categories :

- Empirical models (ALEXANDER, MACKECHNIE et AM 2007)
- Numerical models (BARY et SEILLIER 2003; THIERY 2006)
- Semi-empirical (ATIŞ 2004; CEB-FIP 2010; P.HILL et al. 2015; TAFSESE, Esko SISTONEN et PUTTONEN 2015)

The depth of carbonation (denoted here  $X_c(t)$  and expressed in *mm*) is related to the exposure age  $t$  by :

$$X_c(t) = At^B,$$

where  $A$  and  $B$  are calibration parameters.

In empirical models, authors use data to estimate calibration parameters. These models disregard the wetting–drying effect on the carbonation level which has an important impact on it.

For numerical models, authors try to describe as precisely as possible the carbonation process by integrating all the physical and chemical phenomenon of the process (see OLLIVIER et VICHOT 2008c). Naturally, they need a large number of input parameters for computation.

The analytic solution of the first Fick's law (solved by assuming that carbonation is only caused by the diffusion of carbon dioxide gas within concrete) combined to some indoor and outdoor experimental studies on carbonation lead to the conclusion that depth of carbonation  $X_c(t)$  (expressed here in [*m*]) is proportional to the square root of the exposure age  $t$  (in [*years*]) of concrete :

$$X_c(t) = K\sqrt{t}, \quad (3.5)$$

where  $K$  ( $[m/s^{1/2}]$ ) is the carbonation coefficient.  $K$  is a function of factors which have an impact on the carbonation depth. Three categories of factors are identified :

- Material factors (water-cement ratio, amount of hydrates, 28 days compressive strength of concrete)
- Curing factors (curing conditions and curing time)
- Environmental factors (relative humidity, Temperature, amount of  $CO_2$  in ambient air).

An exhaustive list of semi-empirical models based on equation (5.4) have been developed. They differ from each other in the expression of the carbonation coefficient  $K$ .

By using experimental studies on the microstructure of concrete and resolving complex nonlinear system of differential equations, V. G. PAPADAKIS et al. 1991 elaborates an expression of the carbonation coefficient related to the microstructure of the concrete and to the carbon dioxide concentration. BAKKER 1994 conceived a carbonation model which also includes the wetting-drying cycles by supposing the diffusion of the  $CO_2$  within wet concretes were insignificant. Other several models for  $K$  have been developed using mathematical tools. A full semi-empirical model based on equation 5.4 is performed by a panel of researchers in CEB-FIP 2010. This model, called *Duracrete*, considers environmental and execution factors, as well as the quality of the concrete, as given in Equation (3.3). The various factors in this equation are clearly described in Section 3.3.2.

Regarding statistical approaches, we can mention some important works. ATIŞ 2004 applies statistical methods to laboratory data to evaluate the influence of compressive strength and porosity on accelerated carbonation depth of Fly Ash concretes.

Moreover, to expand statistical methods to carbonation of outdoor exposed concretes modelling, P.HILL et al. 2015 evaluated multivariate analysis and linear regression methods on natural carbonation data by modelling the carbonation coefficient  $K$ . Some other modern methods such as machine learning and neural networks have also been used to identify factors affecting carbonation depth using both laboratory testing and in situ data TAFFESE, Esko SISTONEN et PUTTONEN 2015.

### 3.3 Scope of the present study

#### 3.3.1 The data set

The first step of our work was to build a dataset (i.e. carbonation measurements obtained from concretes exposed in the field) of 793 measured carbonation depths. This is a compilation of published data sets available in EKOLU 2020; EKOLU et van EDE 2014; GUIGLIA et TALIANO 2013; LIANG et al. 2013; NEVES et al. 2013; REVIRON et URRUTIA 2016. The remaining data were collected on private r.c constructions exposed in situ in France. Except those from NEVES et al. 2013 which should remain confidential, these data are provided in the tables of Appendix A.

The information collected about concretes differs from one subset to another. Each item has been completed with information about environmental conditions, such as the average value over 10 years for relative humidity, temperature, amount of the atmospheric  $CO_2$  concentration, and number of rainy days (defined by a minimum amount of precipitation water of  $h_{Nd} = 2.5 \text{ mm}$  (see CEB-FIP 2010) at the exposure sites. These values, provided in Appendix B, were collected from the nearest weather stations. In addition, other complementary information (whether the structure is sheltered or is located in a confined place) was also added to data sets. Finally, in addition to carbonation depth and age of concrete, the database contains the values of the 6 other variables described in Table 3.1.

Most of structures included in the database were erected more than 20 years ago. The type of cement employed in the concrete mixes is therefore assumed to be CEM I, which is the most usual for concrete bridges and was also widely used for ordinary buildings at this time.

#### 3.3.2 Variables selected for the modelling

As mentioned earlier, the *fib* model is an useful carbonation model provided that the parameter  $R_{NAC,0}^{-1}$  is given. Thus, we kept the following covariates in our study.

<b>Id</b>	<b>Description</b>	<b>Unit</b>
$Y$	Concrete carbonation depth	[cm]
$t$	Age of concrete	[Year]
$f_{cm}$	Compressive strength at $t$	[Mpa]
$RH$	Relative humidity of the exposure area	[%]
<i>Confined</i>	Indicates if concrete is located in a confined place	[-]
<i>Sheltered</i>	Variable indicates if concrete is sheltered	[-]

**Table 3.1** – Variables of the dataset obtained.

- *The parameter  $k_e$*

The relative humidity  $RH$  of the exposure environment impacts the intrusion of  $CO_2$  into the concrete pores. The carbonation process is optimum when  $RH$  remains around 65%. For such a value, gaseous diffusion is not hindered to any great extent by the moisture of concrete, which is nonetheless enough to allow easy dissolution of hydrates in the pore solution (BALEN et GEMERT 1994). The  $k_e$  parameter is proposed in *fib-Model Code 2010* and allows the influence of the relative humidity on the carbonation to be considered. It is defined by :

$$k_e = \left( \frac{1 - \left( \frac{Rh_{real}}{100} \right) f_e}{1 - \left( \frac{Rh_{ref}}{100} \right) f_e} \right)^{g_e}, \quad (3.6)$$

where  $Rh_{ref} = 65\%$  is taken as the reference value of the relative humidity and  $Rh_{real}$  is the relative humidity of the area. For our study, for each item  $Rh_{real}$  is equal to the 10-years average value of the relative humidity from <https://www.fr.climate-data.org/>. The parameters  $f_e$  and  $g_e$  are regression parameters and we have :  $f_e = 5$  and  $g_e = 2.5$  (CEB-FIP 2010).

- *The parameter  $C_s$*

The  $CO_2$  concentration of the ambient air is denoted by  $C_s$  and expressed in  $[kg/m^3]$ . It is defined by

$$C_s = C_{S.atm} + C_{S.emi.}, \quad (3.7)$$

where  $C_{S.atm}$  is the  $CO_2$  concentration of the atmosphere  $[kg/m^3]$  and  $C_{S.emi.}$  is the additional  $CO_2$  concentration due to emission sources  $[kg/m^3]$ . The atmospheric  $CO_2$  concentration value of each area is taken to be equal to the average value of the annual  $CO_2$  concentration from 2003 to 2019. These data are download from A et al. 2019. The additional  $CO_2$  concentration ( $C_{S.emi.} = 0.00049 kg/m^3$ ) is used for confined structures like tunnels. For usual structures, we use  $C_s = C_{S.atm}$  (CEB-FIP 2010).

- *The weather parameter  $W(t)$*

The weather  $W(t)$  function quantifies the effect of rain events on the carbonation depth of the concrete. According to CEB-FIP 2010, it is expressed by :

$$W(t) = \left( \frac{t_0}{t} \right)^{\frac{(P_{SR} \cdot ToW)^{b_{tw}}}{2}}, \quad (3.8)$$

where :

- $t_0 = 0.0767$  is time of reference [*year*];
- $p_{SR}$  is the probability of driving rain. It is equal to 1 for horizontal materials and 0 for sheltered materials. In our study, we consider that all unsheltered concretes are already wet during rain events (i.e.  $p_{SR} = 1$ );
- $ToW$  (Time of Wetness) is equal to the ratio between the number of annual rainy days with a minimum water precipitation of  $h_{Nd} = 2.5 \text{ mm}$  and the number of days in a year (CEB-FIP 2010). These values were obtained from the nearest weather station;
- $b_w$  is a random parameter. Its values are obtained from a normal distribution with mean  $m = 0.446$  and standard deviation  $sd = 0.163$ .

#### - The curing parameter

Curing has an important influence on carbonation. Insufficient curing decreases the quality of the surface layer, which promotes carbonation (see BALAYSSAC et al. 1994). The *fib* model includes the effect of concrete curing time on modelled carbonation by considering the random parameter  $k_c$  [-] defined as follows. Its randomness is given by  $b_c$ , which follows a normal distribution with mean  $m = -0.567$  and standard deviation  $sd = 0.024$  (CEB-FIP 2010).

$$k_c = \left( \frac{t_c}{7} \right)^{b_c}, \quad (3.9)$$

where  $t_c$  [*day*] is the curing time. In this paper, except for data from GUIGLIA et TALIANO 2013 where curing time is equal to 3 days (according to Italian standard), the values of  $t_c$  are determined according to 13670 2013 by taking into account the surrounding temperature on the work site.

Hence, this curing parameter only expresses the influence of the curing duration. However, according to Balayssac's works, the curing effect is also dependent on the concrete design (28-day compressive strength, water-cement ratio) (see BALAYSSAC et al. 1994). Consequently, we consider in this study, the present study also uses the curing parameter (denoted by  $k_{cfmt}$ ) proposed by KARI 2011 and that includes the average value of the 28-day compressive strength (denoted by  $f_{cm28}$  and expressed in [*Mpa*]) :

$$k_{cfmt} = \frac{1.9 \times 10^{-2}}{10^{-0.025f_{cm28}}} + \left( 1 - \frac{1.9 \times 10^{-2}}{10^{-0.025f_{cm28}}} \right) \sqrt{\frac{28}{0.01 \times t_c}}. \quad (3.10)$$

In this work, we assume that the ratio of 2-day mean compressive strength to 28-day mean compressive strength is in the interval [0.3, 0.5]. Moreover, the mean compressive strength at 28 days can be derived from by using the experimental compressive strength  $f_{cm}$  according to CEB-FIP 2010 as follows :

$$f_{cm28} = e^{-s \left( 1 - \sqrt{\frac{28}{t}} \right)} f_{cm}, \quad (3.11)$$

where  $s$  is a parameter depending on the strength class of the cement. Here, the value of 0.25 is given to  $s$ , corresponding to the most usual cement employed in bearing structures.

- *Material factor*

Carbonation is caused by the ingress of  $\text{CO}_2$  into pores of the concrete. Several works have shown that the porous network of the concrete is most correlated to the content and type of cement. As mentioned above, only CEM I cement is considered in this study. However, obtaining information about cement contents from experiment is rather complex and may be inaccurate, if not reported in documentation related to the structure. In contrast, the compressive strength, also depending on the cement content of the concrete (see V. PAPADAKIS et DEMIS 2011), is information that is systematically provided for each structure of our database. According to CEB-FIP 2010, the compressive strength of the concrete at age  $t$  (denoted  $f_{cm}$ ) is a function of its 28-day compressive strength. So, in our work, material parameters are represented by  $t$  and  $f_{cm}$ .

Finally, the variables presented in the Table 3.2 were chosen for modelling.

Id	Description	Unit
$Y$	Concrete carbonation depth	[cm]
$t$	Age of concrete	Years
$f_{cm}$	Compressive strength at $t$	[MPa]
$w_t$	Weather parameter	[-]
$k_e$	Environmental parameter	[-]
$C_s$	Amount of atmospheric $\text{CO}_2$ gas	$\text{Kg}/\text{m}^3$
$k_{cfmt}$	Curing parameter	[-]

Table 3.2 – Variables of the final dataset.

### 3.4 Statistical models for carbonation depth prediction

As mentioned previously, the carbonation process is governed by material, environmental and technological factors characterizing the disposition of concrete to carbonation. In this section, we will introduce two carbonation models, a generalized linear regression model and a gamma degradation model.

#### 3.4.1 Carbonation Model based on Generalized Linear Regression Methods

##### 3.4.1.1 A brief presentation of GLM

The generalized linear models (GLMs) are extensions of the famous linear regression model (LM). LMs are used to predict a quantitative outcome  $\mathbf{Y}$  on the basis on a vector of  $p$  predictor variables also called covariates and denoted by  $\mathbf{X} = (1, X^1, \dots, X^p)^T$ . It is assumed that  $\mathbf{Y}$  is the sum of a linear combination of predictors and an *error term*. In other words, we have :

$$\mathbf{Y} = \sum_{j=0}^p \beta_j X^j + \varepsilon, \quad (3.12)$$

where :

- $\beta = (\beta_0, \beta_1, \dots, \beta_p)^T$  is the unknown vector of the regression coefficients ;
- $\varepsilon$  is the *error term* (the part of  $\mathbf{Y}$  unexplained by predictors). It is a random variable assumed to have a Gaussian (normal) distribution with zero mean and standard deviation  $\sigma$ .

From the above, it is clear that  $\mathbf{Y}$  has a Gaussian distribution. Thus, LMs are only appropriated for Gaussian distributed targets. GLM allows LM to be extended to other types of outcomes (non-normal targets like discrete quantitative or qualitative variables) but still assumes the linear combination of predictors. In a GLM,  $\mathbf{Y}$  is assumed to have a probability distribution with density  $f_{\mathbf{Y}}$  and a function of its expected value expressed as a linear combination of covariates  $\mathbf{X}$ . More precisely, if we denote by  $\{(\mathbf{Y}_i, \mathbf{X}_i); i = 1, \dots, n\}$  an independent replication of the random variable  $(\mathbf{Y}, \mathbf{X})$ , we have :

$$\begin{cases} Y_i \sim f_{\mathbf{Y}_i} \\ g(\mu_i) = \sum_{j=0}^p \beta_j X_i^j \Leftrightarrow \mu_i = g^{-1} \left( \sum_{j=0}^p \beta_j X_i^j \right), \end{cases} \quad (3.13)$$

where

- $\mu_i = \mathbf{E}[Y_i]$  is the expected value of  $\mathbf{Y}_i$ ;
- $f_{\mathbf{Y}_i}$  is the density function of  $\mathbf{Y}_i$ . It is assumed known;
- $g$  is a monotonic and differential function called the *link function*. It is also assumed known;
- $\beta = (\beta_0, \beta_1, \dots, \beta_p)^T$  is the unknown vector of the regression coefficients.

Once the most suitable probability distribution and link function are found (as described in the next subsection), the coefficients of the model are estimated with maximum likelihood methods (see McCULLAGH et J. NELDER 1989).

### 3.4.1.2 GLM fitting of carbonation depth

The aim of our study is to predict the carbonation depth,  $\mathbf{Y}$ , of concretes with given characteristics. For this purpose, let us assume that the carbonation depth,  $\mathbf{Y}_i$ , of the concrete  $i$  follows a GLM with features  $\mathbf{X}_i = (1, \sqrt{t^i}, f_{cm}^i, k_e^i, C_s^i, W_t^i, k_{cfmt}^i, k_c^i)^T$  for  $i = 1, \dots, n$ .

Note that the we have assumed that the carbonation depth is related to the square root of the concrete's age, in accordance with to KLOPPER 1978.

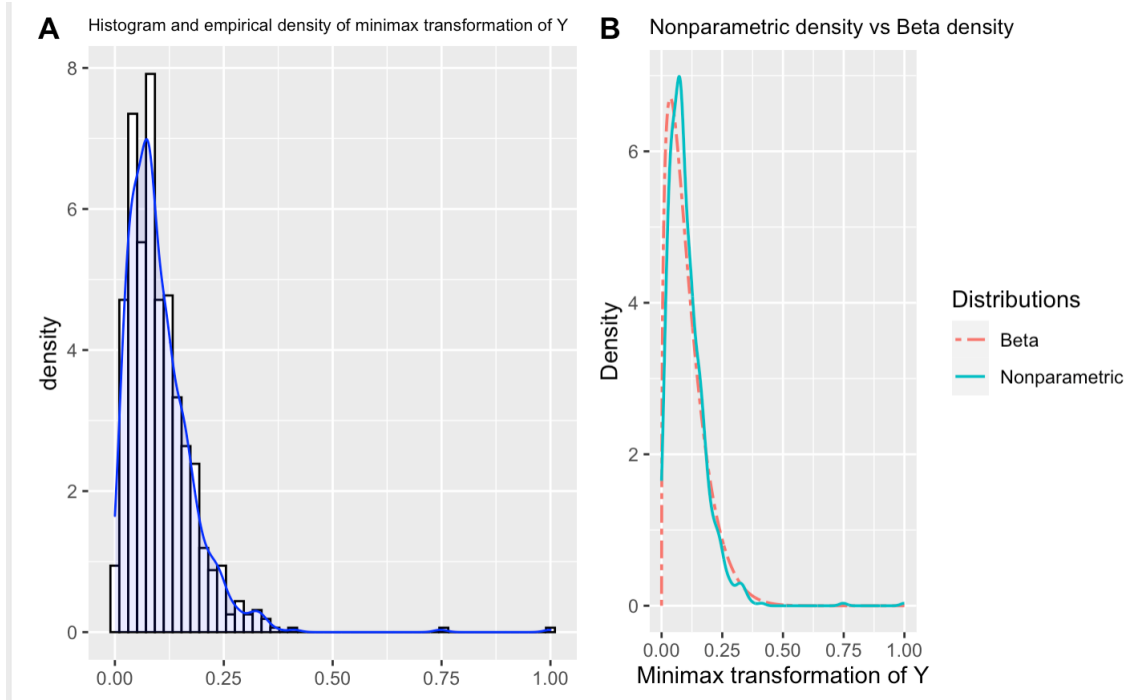
Furthermore, it is important to recall that the variable  $k_{cfmt}$  is a function of the 28-days mean compressive strength and the curing time. As a result, there is a possible correlation between  $k_{cfmt}$  and  $f_{cm}$ , and between  $k_{cfmt}$  and  $k_c$ . This information is accounted in models by introducing interaction variables  $k_{cfmt} \times f_{cm}$  and  $k_{cfmt} \times k_c$ .

The first step in the fitting procedure is to identify the adequate distribution law and link function from our dataset. Then, the corresponding regression parameters are estimated.

The choice of the probability distribution is based on both graphical analysis and nonparametric hypothesis testing. A graphical analysis of the empirical density function of  $\mathbf{Y}$  (also of its *log* and *minimax* transformations) shows that the best fit is given by a Beta distribution on the *minimax* transformation (denoted by  $\tilde{\mathbf{Y}}$ ). This can be seen on the Figure 5.2, where the empirical density of  $\tilde{\mathbf{Y}}$  and the Beta distribution are depicted. It can be seen that these two density curves are very close. We recall that the density function of the Beta distribution is defined by :

$$f(y; a, b) = \begin{cases} \frac{y^{a-1}(1-y)^{b-1}}{\int_0^1 u^{a-1}(1-u)^{b-1} du} & \text{for } y \in ]0, 1[ \\ 0 & \text{otherwise,} \end{cases} \quad (3.14)$$

where  $a$  and  $b$  are two positive reals called shape parameters. Since the carbonation depth  $\mathbf{Y}$  may have values greater than 1, the Beta distribution is fitted on the transformed variable  $\tilde{\mathbf{Y}}$



**Figure 3.1** – Graphical description and graphical adequacy test between the minimax transformation of  $\mathbf{Y}$  and the Beta distribution.

defined by  $\tilde{\mathbf{Y}} = \frac{\mathbf{Y} - \mathbf{Y}_{\min}}{\mathbf{Y}_{\max} - \mathbf{Y}_{\min}}$  where  $\mathbf{Y}_{\min}$  and  $\mathbf{Y}_{\max}$  are the minimum and maximum values of  $\mathbf{Y}$  respectively.

A Kolmogorov-Smirnov goodness of fit test (see BAGDONAVIČIUS et al. 2011) was performed and the hypothesis that  $\tilde{\mathbf{Y}}$  is from a Beta distribution is not rejected with the highest significance level ( $p\text{-value} = 0.1028$ ).

Thus, in the second part of this section, let us perform the minimax transformation  $\tilde{\mathbf{Y}}$  of  $\mathbf{Y}$  and let us assume that this random variable is Beta distributed. The next step is to find which link function best fits the observed data. We considered the following link functions :

- *logit* :  $g(\mu) = \log\left(\frac{\mu}{1-\mu}\right)$
- *probit* :  $g(\mu) = \Phi^{-1}(\mu)$ , where  $\Phi(\cdot)$  is the cumulative density function of the standard normal distribution.
- *loglog* :  $g(\mu) = -\log(-\log(\mu))$
- *complementary log-log (cloglog)* :  
 $g(\mu) = \log(-\log(1-\mu))$
- *cauchy* :  $g(\mu) = \tan(\pi(\mu - 0,5))$ .

The performance of a GLM model can be measured by its *MSE* defined by :

$$MSE = \frac{1}{n} \sum_i (y_i - \hat{y}_i)^2, \quad (3.15)$$

where  $n$  is the sample size and  $\hat{y}_i$  is the predicted value of  $y_i$  given by the model. The Lower the *MSE* is, the better is the model.

To avoid overfitting phenomenon, model selection was performed by using the *K-fold cross-validation* (see FUSHIKI 2009). This consist of splitting the dataset into  $K$  subsets called folds. Each model is trained on  $K - 1$  folds and then tested (*MSE* calculation) on the remaining fold. The procedure is repeated as many times as there are folds (i.e.  $K$  times).



In the fitting of models, the predictor  $k_e$  is first considered as a continuous variable. From a practical viewpoint this is not fully satisfactory because it is easier to consider the relative humidity of the ambient air where structures are exposed as falling into one of three categories pertaining to dry, wet or medium conditions, as recommended in 206/CN 2014. Therefore the predictor  $k_e$  was subsequently replaced by its categorical version, denoted  $K_{ecat}$  :

- *dry condition* for concretes with  $k_e > 1$  (i.e.  $RH < 65\%$ );
- *medium condition* for concretes with  $k_e \in [0.5, 1]$  (i.e.  $65\% \leq RH \leq 80\%$ );
- *wet condition* for concretes with  $k_e < 0.5$  (i.e.  $RH > 80\%$ ).

Figure 3.2 gives boxplots of  $MSE$  for each model by using  $K = 50$  folds. It can be clearly seen that  $MSE$  values are better when  $K_{ecat}$  is considered. In this case, except for the *cauchy* link,  $MSE$  values obtained are similar for the other link functions. However, it can be seen that the *Cloglog* link function gives the lowest  $MSE$  value. Thus, the model with the *Cloglog* link function, which consider  $K_{ecat}$ , was selected.

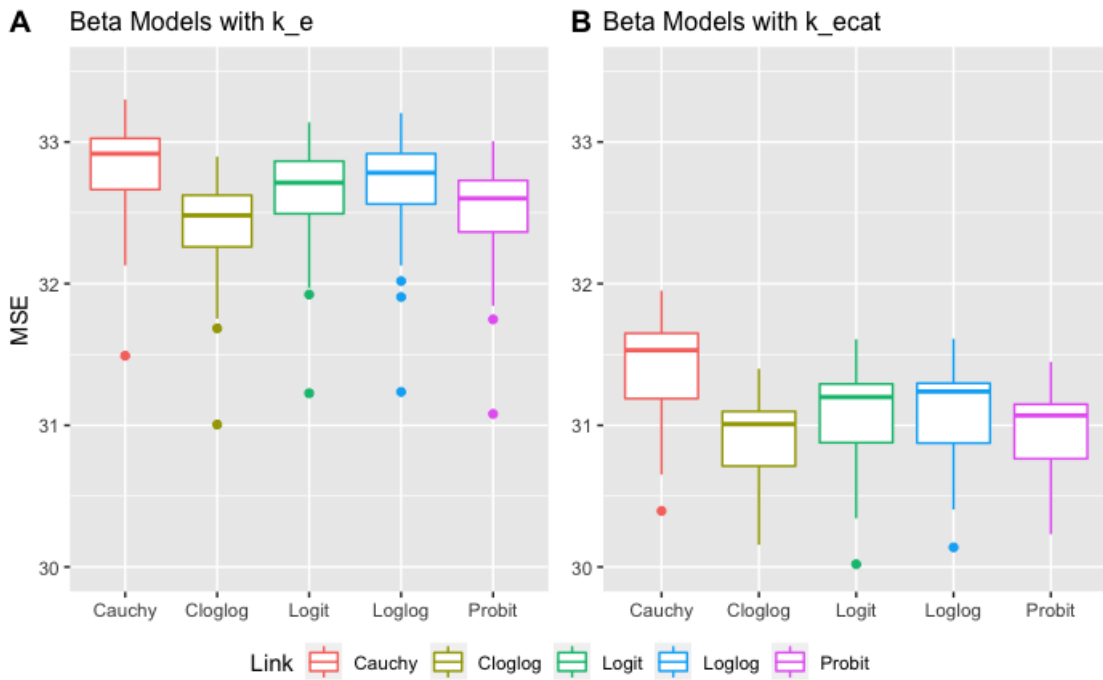


Figure 3.2 – Boxplots of model's  $MSE$ s obtained on  $K = 50$  folds of carbonation models based on GLM.

Table 3.3 gives the estimated value of the regression parameter  $\beta_j$  associated with each component of the vector  $\mathbf{X}$ , its standard deviation and the corresponding p-values of Wald tests. The Wald test (see McCULLOCH et SEARLE 2001b) is used to examine significance of predictors. Let us recall that this is a test of  $\beta_j = 0$  (i.e. the covariate does not significantly affect the carbonation depth  $\mathbf{Y}$ ). In particular, the test of  $\beta_{K_{ecat}^{dry}} = 0$  (respectively  $\beta_{K_{ecat}^{Medium}} = 0$ ) is equivalent to examining whether the carbonation depth is significantly different under dry conditions (respectively medium conditions) as opposed to wet conditions.

It can be seen that, except for  $k_{cfmt} \times f_{cm}$  (p-value  $> 5\%$ ), all variables are highly significant.

Finally, after excluding this non-significant variable from the predictors, regression coefficients were re-estimated. These estimated coefficients and their standard deviations are presented in the Table 3.4.

Our final model assumes that, for a concrete with known characteristics  $\mathbf{X}$  equals to the vector

	Estimated	Std. error	<i>p</i> -value
$\beta_0$	72.32	12	$8.7 \times 10^{-10}$
$\beta_{\sqrt{t}}$	0.2	$1.38 \times 10^{-2}$	$4.7 \times 10^{-48}$
$\beta_{f_{cm}}$	-0.025	$2.05 \times 10^{-2}$	$4.04 \times 10^{-2}$
$\beta_{K_{ecat}^{dry}}$	1.13	0.026	$4.5 \times 10^{-9}$
$\beta_{K_{ecat}^{Medium}}$	1.28	0.058	$6.32 \times 10^{-9}$
$\beta_{k_{cfmt}}$	-0.268	0.01	0.008
$\beta_{k_c}$	-4.8	0.916	$1.28 \times 10^{-9}$
$\beta_{C_s}$	170	29	$5.07 \times 10^{-9}$
$\beta_{W_t}$	0.119	0.0512	0.0202
$\beta_{k_{cfmt} \times k_c}$	0.169	0.0348	$1.21 \times 10^{-6}$
$\beta_{k_{cfmt} \times f_{cm}}$	$-1.55 \times 10^{-5}$	0.025	0.9799

**Table 3.3** – Estimated regression coefficients and the corresponding p-values (Wald test) for the carbonation model based on GLM.

	Estimated	Std. error	<i>p</i> -value
$\beta_0$	72.4	11.8	$8.7 \times 10^{-10}$
$\beta_{\sqrt{t}}$	0.2	0.0136	$4.73 \times 10^{-48}$
$\beta_{f_{cm}}$	-0.026	$4.8 \times 10^{-3}$	$7.58 \times 10^{-8}$
$\beta_{K_{ecat}^{dry}}$	1.135	0.0225	$4.9 \times 10^{-7}$
$\beta_{K_{ecat}^{Medium}}$	1.28	0.020	$6.3 \times 10^{-9}$
$\beta_{k_{cfmt}}$	-0.27	0.04	$1.13 \times 10^{-10}$
$\beta_{k_c}$	-4.84	0.912	$1.1 \times 10^{-7}$
$\beta_{C_s}$	-170	28.7	$3.08 \times 10^{-9}$
$\beta_{W_t}$	0.12	0.0511	0.02
$\beta_{k_{cfmt} \times k_c}$	0.17	0.057	$4.39 \times 10^{-11}$

**Table 3.4** – Estimated regression coefficients and the corresponding p-values (Wald test) for carbonation model based on GLM (Final model).

$(\sqrt{t}, f_{cm}, k_e, k_{cfmt}, k_c, C_s, W_t)^T$ , the mean value  $\mu$  (in *mm*) of the carbonation depth is given by the following relationship :

$$\begin{aligned}
 \mu = & 51 - 50 \exp\{-\exp(72.4 + 0.2\sqrt{t} \\
 & - 0.025f_{cm} + 1.13I_{\{k_e > 1\}} \\
 & + 1.28I_{\{k_e \in [0.5; 1]\}} - 0.27k_{cfmt} \\
 & - 4.84k_c - 170C_s \\
 & + 0.12W_t + 0.168k_c \times k_{cfmt})\}
 \end{aligned} \tag{3.16}$$

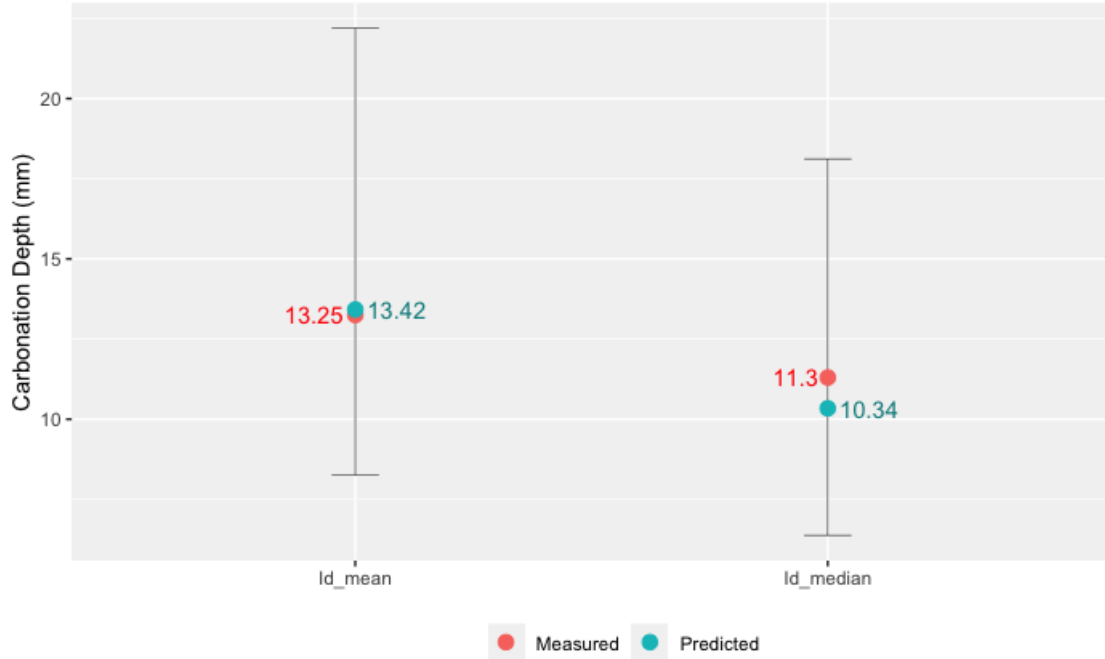
where  $I_A$  is the indicator function of the set  $A$ .

As an example, we can use this model to predict the mean carbonation value of two artificial concretes that we have respectively named  $Id_{mean}$  and  $Id_{median}$ . The features of  $Id_{mean}$  (resp. of  $Id_{median}$ ) are the average (resp. the median) values of  $t; f_{cm}; k_e; k_{cfmt}; k_c; C_s$  and  $W_t$  obtained with our data set (i.e. by using all the 793 carbonation measurements). Table 3.5 provides the values of these features for each artificial concrete.

	$t$	$f_{cm}$	$k_e$	$k_{cfmt}$	$k_c$	$C_s$	$W_t$	$Y$
$Id_{median}$	5.07	39.6	0.69	25.37	1.61	0.39442	1	11.3
$Id_{mean}$	16.16	39.5	0.77	25.52	1.54	0.39454	0.55	13.24

**Table 3.5** – Covariates values of  $Id_{median}$  and  $Id_{mean}$ .

The results of the prediction show that predicted and measured values of  $Id_{mean}$  are almost equal. For  $Id_{median}$ , the predicted value of the carbonation depth is smaller than the measured value by one unit as seen on Figure 3.3.



**Figure 3.3** – Predicted values of carbonation depths and their 95% confident intervals for  $Id_{mean}$  and  $Id_{median}$  gives by equation (3.16).

### 3.4.2 Model based on Gamma degradation models

#### 3.4.2.1 Overview of Gamma degradation process

Let  $Y_t$  be the random value of degradation at the measurement time  $t$ . A classical model for the degradation process  $(Y_t)_{t \geq 0}$  is given by a Gamma process (MEEKER et ESCOBAR 1998) if :

- the increments  $\Delta Y_t = Y_t - Y_s, \forall 0 \leq s < t$  are independent of one another ;
- $\forall t, \Delta Y_t$  is Gamma distributed. In this model, the density function of a Gamma distribution with shape parameters  $A(t)$  (assumed to be time dependent) and scale parameter  $\theta$  (assumed to be constant over time) is defined as follows :

$$f(z, A(t), \theta) = \frac{z^{A(t)-1} e^{-\frac{z}{\theta}}}{\Gamma(A(t)) \theta^{A(t)'}}$$

where  $A(\cdot)$  is a positive and increasing function,  $\theta$  a positive real and  $\Gamma$  the standard gamma function (defined by  $\Gamma(h) = \int_0^{+\infty} s^{h-1} e^{-s} ds, \forall h \geq 0$ ).

Covariates  $\mathbf{X} = (1, X_1, \dots, X_p)^T$ , where  $p$  is the number of predictors, can be incorporated in degradation Gamma models by integrating them in the expression of the shape function  $A(\cdot)$

as in BAGDONAVICIUS et NIKULIN 2000b or in the scale parameter  $\theta$  as proposed by LAWLESS et M.CROWDER 2004b.

### 3.4.2.2 Gamma degradation model fitting of carbonation depth

As it can be seen above, Gamma degradation models are applied in general cases where degradation is measured several times on each component. In the rest of our work, we will use Gamma degradation models to build a carbonation model in the simple case where a single degradation measurement is taken on each concrete.

Moreover, as done in BAGDONAVICIUS et NIKULIN 2000b, we assume that covariates effects are taken into account in the shape function  $A(\cdot)$  such that  $A(t | \mathbf{X}) = t^\lambda e^{\beta' \mathbf{Z}}$  where  $\beta = (\beta_1, \dots, \beta_p)^T$  is the vector of unknown parameters,

$$\mathbf{Z} = g(\mathbf{X}) \text{ and } g : \mathbb{R}^p \mapsto \mathbb{R}^p$$

is some specified function. In other words, it is assumed that, at any time  $t$ , the carbonation depth  $Y_t(\mathbf{X})$  has a Gamma distribution with shape function  $t^\lambda e^{\beta' \mathbf{Z}}$  and scale parameter  $\theta$ .

The fitting procedure consists of using our dataset to find the most suitable  $g$  function and then estimating model's parameters ( $\lambda$ ,  $\beta$  and  $\theta$ ). In addition, the following  $g$  functions are tried according to BAGDONAVICIUS et NIKULIN 2000b.

- $g((X_1, \dots, X_p)) = (X_1, \dots, X_p)^T$ ,
- $g((X_1, \dots, X_p)) = (\log(X_1), \dots, \log(X_p))^T$  (power rule model),
- $g((X_1, \dots, X_p)) = (1/X_1, \dots, 1/X_p)^T$  (Arrhenius model).

Let us respectively denote these models *Gamma.exp*, *Gamma.power* and *Gamma.arrhenius*. Once the models had been calibrated on our data base (estimating model's parameters), their performance and adequacy were evaluated with the *K-fold cross validation* method.

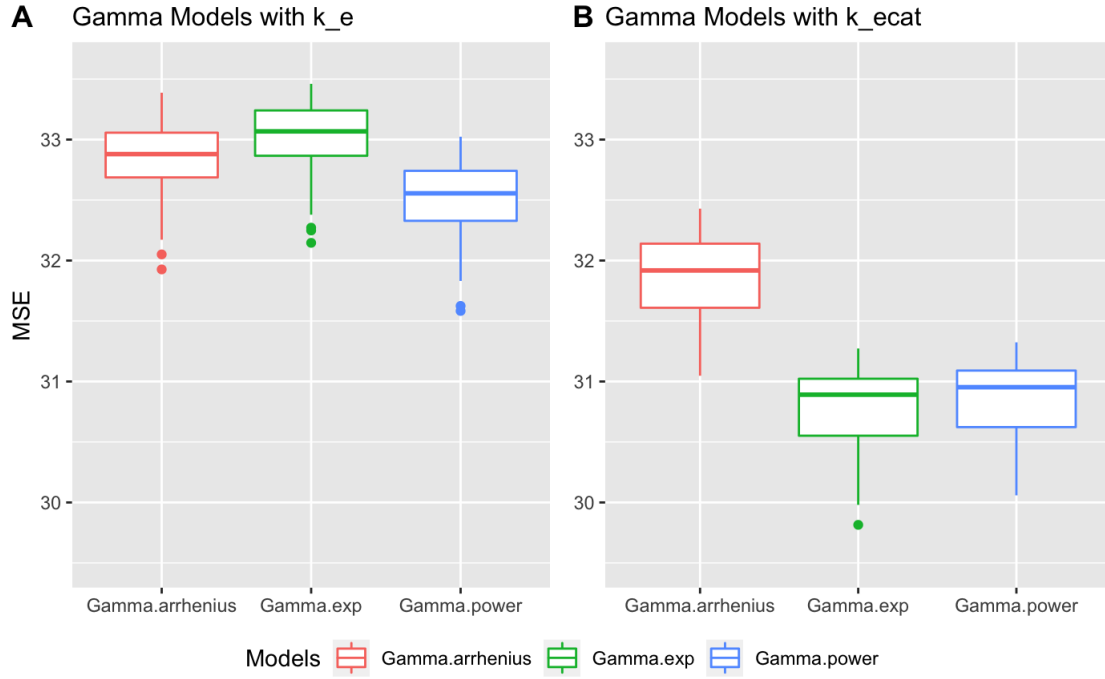
Figure 3.4 gives boxplots of *MSE* obtained from this method with  $K = 50$  folds. As can be seen, the *MSE* values are lowest when the categorical version the environmental parameter related to the relative humidity is considered (i.e.  $K_{ecat}$ ). However, in this case, the models *Gamma.exp* and *Gamma.power* have almost the same *MSE* value.

Moreover, the Wald test is performed to evaluate predictors significance of these models (*Gamma.exp* and *Gamma.power*). Table 3.6 provides the p-values of the Wald test for each predictor. In both models, it can be seen that the interaction variable  $k_{cfmt} \times f_{cm}$  does not significantly affect the carbonation process (p-value  $> 5\%$ ). In addition, the weather parameter  $W_t$  is considered as significant predictor in the *Gamma.exp* model while it is not in the *Gamma.power* model.

From this, it can be concluded that the *Gamma.exp* model reveals a satisfactory fit of the data set when  $K_{ecat}$  is used. Table 3.7 gives the estimated coefficients, their standard errors, and the p-values of the Wald test performed for each predictor of the *Gamma.exp* model re-estimated without the non-significant variable  $k_{cfmt} \times f_{cm}$ .

The obtained model (*Gamma.exp*) allows the mean value of the carbonation depth (in *mm*) to be estimated at a given time  $t$  and features  $\mathbf{X} = (f_{cm}, k_e, k_{cfmt}, k_c, C_s, W_t)^T$  by the following expression :

$$\begin{aligned} \mu = t^{0.29} \exp(71.3 - 0.02f_{cm} + 1.1I_{\{k_e > 1\}} \\ 1.22I_{\{k_e \in [0.5; 1]\}} - 0.21k_{cfmt} - 4.25k_c \\ - 161C_s + 0.15W_t + 0.14k_{cfmt} \times k_c) \end{aligned} \quad (3.17)$$



**Figure 3.4** – Boxplots of model'  $MSEs$  obtained on  $K = 50$  folds for carbonation based on Gamma degradation process.

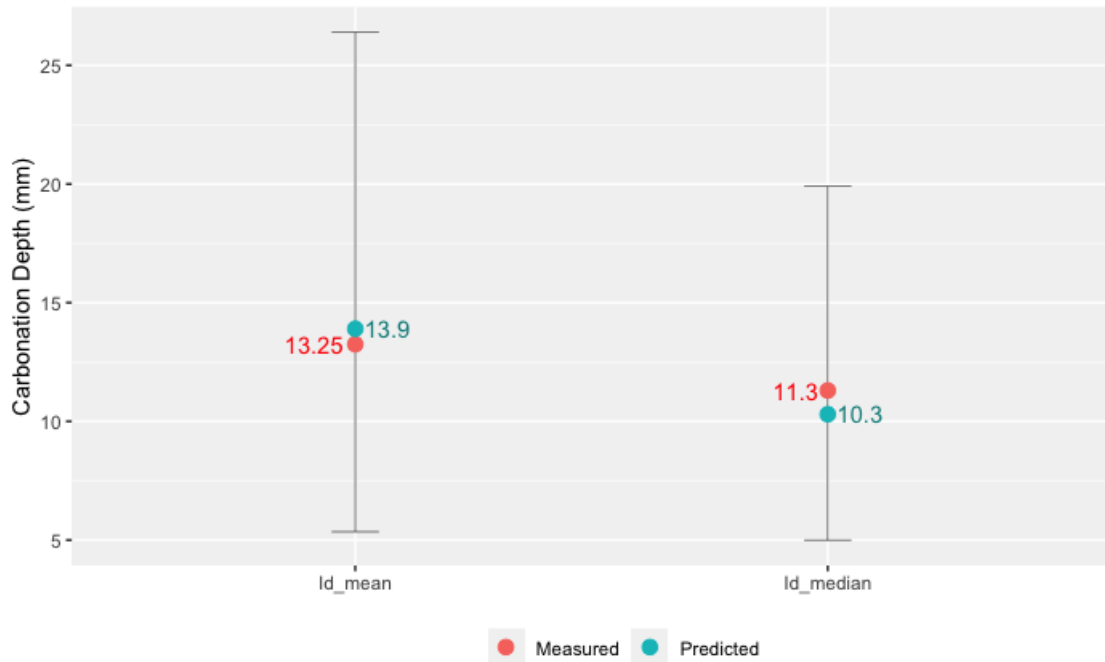
	<i>Gamma.power</i>	<i>Gamma.exp</i>
$\lambda$	$9.9 \times 10^{-54}$	$2.3 \times 10^{-57}$
$\beta_0$	$5.9 \times 10^{-10}$	$2.3 \times 10^{-12}$
$\beta_{f_{cm}}$	$2.75 \times 10^{-9}$	$3.9 \times 10^{-7}$
$\beta_{K_{ecat}^{dry}}$	$2.9 \times 10^{-16}$	$1.3 \times 10^{-16}$
$\beta_{K_{ecat}^{Medium}}$	$2.9 \times 10^{-21}$	$1.3 \times 10^{-21}$
$\beta_{k_{cfmt}}$	$3.5 \times 10^{-6}$	$3.1 \times 10^{-9}$
$\beta_{k_c}$	$2.2 \times 10^{-9}$	$2.12 \times 10^{-8}$
$\beta_{C_s}$	$6.2 \times 10^{-12}$	$8.6 \times 10^{-11}$
$\beta_{k_{cfmt} \times k_c}$	$3.17 \times 10^{-9}$	$4.5 \times 10^{-10}$
$\beta_{k_{cfmt} \times f_{cm}}$	0.35	0.96
$\beta_{W_i}$	0.096	0.0004

**Table 3.6** – P-values of the Wald's test performed for each predictor of *Gamma.power* model and *Gamma.exp* model when  $K_{ecat}$  is used.

This second model is also used to predict carbonation depth of the  $Id_{median}$  and  $Id_{mean}$  described in the previous section. The results of predictions (predicted values of the carbonation depths and their 95% confidence intervals) are shown on Figure 3.5. It is seen that the predicted carbonation depth value for  $Id_{median}$  is the same as the one obtained with the GLM model. For  $Id_{mean}$ , the predicted value provided by the Gamma model is greater than the measured carbonation depth but with an acceptable deviation (less than 1 mm).

	Estimated	Std. error	<i>p</i> -value
$\lambda$	0.29	0.016	$2.3 \times 10^{-57}$
$\beta_0$	71.3	10	$2.35 \times 10^{-12}$
$\beta_{f_{cm}}$	-0.02	0.004	$3.9 \times 10^{-7}$
$\beta_{K_{ecat}^{dry}}$	1.22	0.124	$1.3 \times 10^{-16}$
$\beta_{K_{ecat}^{Medium}}$	1.1	0.13	$1.3 \times 10^{-21}$
$\beta_{k_{cfmt}}$	-0.21	0.035	$3.1 \times 10^{-9}$
$\beta_{k_c}$	-4.26	0.75	$2.12 \times 10^{-8}$
$\beta_{C_s}$	-161	24	$8.6 \times 10^{-11}$
$\beta_{W_t}$	0.15	0.02	0.0004
$\beta_{k_{cfmt} \times k_c}$	0.14	0.022	$4.5 \times 10^{-10}$

**Table 3.7** – Estimated regression coefficients, their standard errors and the corresponding p-values (Wald test) for carbonation model based on Gamma degradation model (Final model).



**Figure 3.5** – Predicted values of carbonation depths and their 95% confident intervals for  $Id_{mean}$  and  $Id_{median}$  gives by equation (3.17).

### 3.5 Application of proposed models to real sheltered structures and comparison with the *fib* model

GUIGLIA et TALIANO 2013 proposes an empirical relationship between the inverse resistance to carbonation  $R_{NAC,0}^{-1}$  and the compressive strength of concrete at the 28<sup>th</sup> day ( $f_{cm28}$ ). For some sheltered concretes (abutments, piers and tunnels), the following expression is suggested :

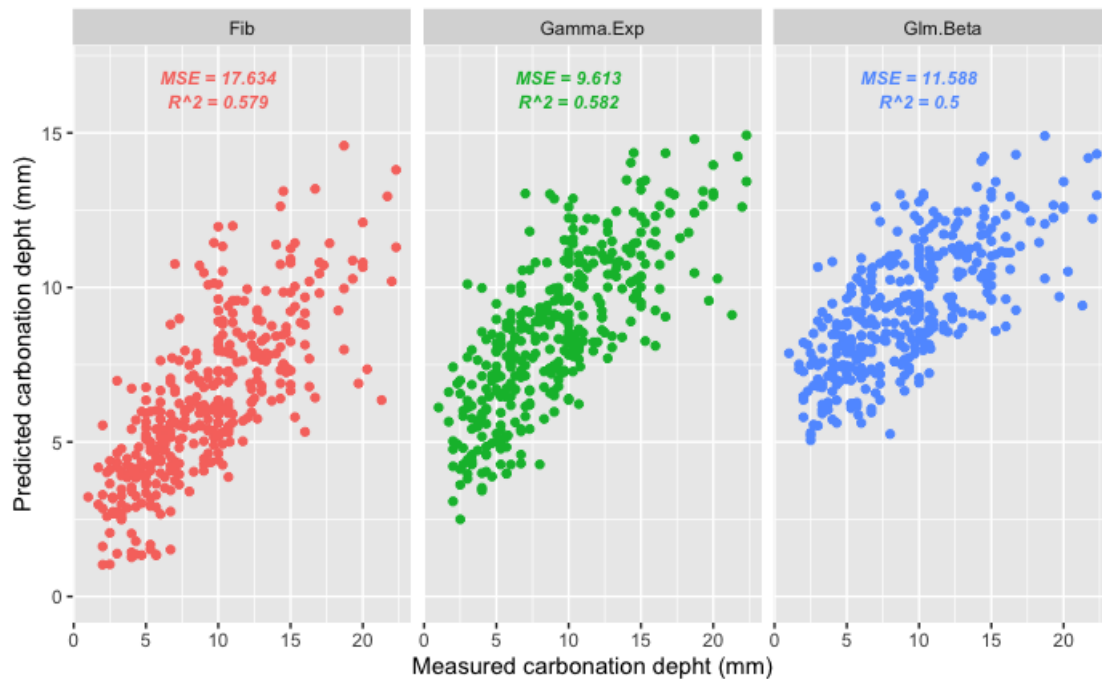
$$R_{NAC,0}^{-1} = 10^7 f_{cm28}^{-2.1} \quad (3.18)$$

where  $f_{cm28}$  is in  $N/mm^2$  and  $R_{NAC,0}^{-1}$  in  $mm^2/year/(kg/m^3)$ .

If the expression (3.18) is introduced into equation (3.3), the resulting carbonation depths given by the *fib* model for data considered in GUIGLIA et TALIANO 2013 are best fitted.

For ease of notation, the GLM model is referred as *Glm.Beta* model in the remainder of this paper.

Our proposed models are used to predict carbonation depths of these same data and prediction errors (*MSE*) and the determination coefficient of Pearson ( $R^2$ ) are calculated. Figure 3.6 shows the scatter plot of measured versus predicted values of carbonation depths and gives the *MSE* and  $R^2$  values for *Glm.beta*, *Gamma.exp* and the *fib* models respectively.



**Figure 3.6** – Measured versus Predicted carbonation depths obtained from *Glm.Beta*, *Gamma.exp* and *Fib* models evaluated on Guiglia & Taliano's data ; with the corresponding values of *MSE* and  $R^2$ .

A comparison of the results suggests that statistical models calibrated from a wide and varied database give a better *MSE* value, and with similar  $R^2$  values, although the dataset used for the comparison is the one used to fit the *fib* model (3.3).

The accuracy of the models is assessed by the coefficient of determination between the predicted and measured carbonation depths given by each model. Pearson's  $R^2$  coefficients show an acceptable level of adequacy of the proposed models. In addition, we see that the *Gamma.exp* and the *fib* models have almost the same accuracy according to the  $R^2$  coefficients.

By using model (3.16), we can predict the carbonation depth of a reinforced concrete exposed to natural carbonation when its compressive strength, curing time, age and some environmental features, such as relative humidity, the  $CO_2$  concentration of the ambient air and information about rain events are known. However, this model overfits small carbonation depths (values less than 5 mm) as seen on Figure 3.6. The *Gamma.exp* model based on a Gamma degradation process, uses the same parameters as *Glm.Beta* model and best fits small values of carbonation depths (by using the model (3.17)).

### 3.6 Conclusion

In this paper, two carbonation models based on statistical models and calibrated on data from worldwide and real-life functioning reinforced concrete structures have been proposed.

The first one, i.e. the *Glm.Beta* model, is based on a generalized linear regression model and the second model, i.e. the *Gamma.exp* model, is based a Gamma degradation model. They constitute useful empirical expressions to predict carbonation depths of exposed R.C structures in the field, with usually available input parameters such as the compressive strength, curing time, age, relative humidity, average CO<sub>2</sub> concentration of the ambient air and number of rainy days (with height of precipitation > 2.5mm) in the year. From our analysis, it is drawn that, the *Glm.Beta* model overfits small carbonation depth values. Thus, it is advisable to use the *Gamma.exp* model for the prediction of the carbonation depth of rather new structures.

Moreover, carbonation depths of in-situ r.c structures from GUIGLIA et TALIANO 2013 have been predicted by using the proposed, and the *fib* models. The results show that the proposed models give better mean squared errors (*MSE*), and may thus be employed as valuable alternative models.



## Chapitre 4

---

# Survival analysis of the time-to-carbonation

---

Comme le précédent, ce quatrième chapitre constitue une version provisoire d'une publication sur la modélisation du temps de carbonatation à l'aide des modèles de durée de vie. Cette publication sera soumise à un comité de lecture international.

## 4.1 Introduction

Corrosion is one of the main causes of deterioration in reinforced concrete (r.c) structures. This two-stage process, consisting of the initiation stage and the propagation stage, compromises the serviceability and safety of the structure. With the exception of underwater r.c structures, carbonation is the main trigger for the initiation stage of corrosion.

Carbonation is a chemical reaction between diffused carbon dioxide (CO<sub>2</sub>) gas and calcium hydroxide present in the cement paste. Atmospheric carbon dioxide infiltrates the concrete pores through diffusion, dissolving in the interstitial solution and forming a weak acid that reacts with the hydrates in the cement paste. Simultaneously, the pH of the interstitial solution drops, initiating the corrosion process of the steel reinforcements embedded in the concrete. The carbonation process is influenced by three main groups of factors :

- Material factors (such as water-to-cement ratio, amount of hydrates, 28-day compressive strength, etc.);
- Curing factors (including curing conditions and curing time);
- Environmental factors (such as relative humidity, temperature, amount of CO<sub>2</sub> in the ambient air, etc.).

The knowledge of the carbonation level at a given time during the operation of the reinforced concrete structure is highly desirable for the owners and managers, as it allows them to initiate corrective maintenance. Therefore, various carbonation models are available in the literature. Among the modelling approaches, the carbonation models can be categorized into three main categories : numerical models, empirical models, and analytical models. This paper focuses on the third category.

Based on the solution of the the Fick's first law, analytical models are the most frequently used by engineers to asses carbonation levels of r.c structures. Assuming that carbonation is only caused by the diffusion of carbon dioxide within the concrete and comparing several experimental studies under indoor or outdoor condition on the carbonation, numerous authors (see e.g. V. G. PAPADAKIS et al. 1991) have concluded that the depth of carbonation  $Y_c(t)$  [mm] is reasonably proportional to the square root of the exposure age  $t$  [year] of the concrete. In other words :

$$Y_c(t) = K\sqrt{t}, \quad (4.1)$$

where  $K$  [mm/year<sup>0.5</sup>] is the carbonation coefficient. This coefficient is a function of factors that influence the carbonation process. The mathematical tools used and derived expressions of  $K$  form the main differences among the analytical models proposed in the literature.

By carrying out experimental studies on concrete micro-structure and solving a complex nonlinear system of differential equations, Papadakis (see V.G. et al. 1991) was one of the first to propose an expression of  $K$  that is related only to the properties of the concrete and the atmospheric CO<sub>2</sub> concentration. On the same basis, BAKKER 1994 proposed a model that additionally includes the effect of wetting-drying cycles on carbonation.

Several other models for  $K$  based on machine learning and deep learning tools are available. Some notable works in this category include LIU et al. 2008; P.HILL et al. 2015; TAFFESE, E. SISTONEN et PUTTONEN 2015; TAFFESE, Esko SISTONEN et PUTTONEN 2015. However, these models have too many input parameters or require calibration on laboratory carbonation data, making their use for assessing in-situ exposed reinforced concrete challenging.

A famous analytical model is the one proposed in the *fib-Model Code 2010* (see CEB-FIP 2010). This model takes environmental and curing factors into account, as well as the quality of the concrete . However, the main pitfall in using the fib model lies in assessing the effective carbonation resistance of dry concrete under natural conditions (the parameter  $R_{NAC,0}^{-1}$ ).

Regarding statistical approaches, some analytical models based on linear regression or on multivariate analysis methods are worth noting. Some of these models focus on investigating the effect of specific factors on carbonation (e.g., ATIŞ 2004; KIM et LEE 1999), while others concentrate on studying particular types of concrete (e.g., LEEMANN et al. 2015). Additionally, there are other statistical carbonation models that have been calibrated using both accelerated and natural carbonation data, yet it should be considered that these models contain parameters that are not always readily available to engineers in practical applications.

In their study, KOUASSI et al. 2022 proposed two carbonation models that incorporate key influencing factors using generalized linear regression and Gamma degradation models, respectively. However, one limitation of these models is that they do not account for the time-dependent nature of environmental factors. It is important to consider this aspect to accurately quantify the impact of these factors on carbonation depth.

Therefore, in this study, we will conduct a survival analysis of carbonation and develop a time-to-carbonation model that takes into account the main influencing factors, including time-dependent factors. Survival analysis, also known as time-to-event analysis, encompasses a range of statistical methods used to analyze the time elapsed before an event occurs. To the best of our knowledge, survival analysis has not been previously applied in carbonation studies. Throughout this paper, we define *time-to-carbonation* as the duration until carbonation reaches the steel reinforcement within the concrete structure.

Our work is presented in three main parts. The first one (Section 4.2) is devoted to the brief presentation of survival analysis and some survival models. In the section 3, we apply survival models to a dataset of carbonation measurements taken from r.c structures exposed to natural carbonation. This part is achieved by an application of the proposed model to assess the time-to-carbonation of a structure located in Paris under some climate change scenarios.

## 4.2 A brief presentation of survival analysis

Survival analysis is a set of statistical methods used to study the random time  $T$  before the occurrence of an event of interest. This event can be, for example, in the medical field, the recovery of a patient from treatment; in sociology, the loss of employment; or in reliability studies, the failure of a system or the carbonation of the reinforced concrete as in our study.

The interest in the survival study are also the factors that influence the occurrence of the event. Factors are characteristics of the studied object that can have a significant effect on its lifetime. For instance, in the case of patients, factors may include their sex, blood pressure, weight, etc. In the context of reinforced concrete structures, a factor could be the relative humidity of the surrounding area. It is important to note that factors can be either time-dependent or time-independent.

One of the strengths of survival methods lies in their ability to handle censoring. In a survival experiment, censoring occurs when the time-to-failure is not precisely observed. The most common types of censoring are right censoring and left censoring. An observation is considered right-censored (resp. left-censored) at a given time when it is only known that the item is still functioning at that time (resp. has experienced failure before that time). In a survival experiment where each item is observed only once, and the resulting information is either left-censored or right-censored, the data are referred to as *current status data*.

Let  $T$  be the non-negative random variable representing the time-to-failure. In the following sections, we will present some basic quantities in survival analysis and the most commonly used approaches to assess the relationship between  $T$  and its factors.

### 4.2.1 Some basics quantities of the survival analysis

In survival analysis, the time-to-failure  $T$  is mainly characterized by its hazard function (or risk function) and its reliability or survival function.

The hazard rate at time  $t$  (or instantaneous risk at  $t$ ), denoted by  $\lambda(t)$ , is approximately the probability that an item of age  $t$  without failure at this time fails in the next small interval of time. More precisely, for all  $t \geq 0$ , the hazard rate  $\lambda(t)$  is defined by :

$$\lambda(t) = \lim_{h \rightarrow 0} \frac{\mathbb{P}(t \leq T < t+h \mid T \geq t)}{h}. \quad (4.2)$$

On the other hand, the reliability or survival function, denoted by  $R(t)$ , represents the probability that an item survives beyond time  $t$  without experiencing failure. For all  $t \geq 0$ , the reliability function  $R(t)$  is given by :

$$R(t) = \mathbb{P}(T > t) = \exp\left(-\int_0^t \lambda(s) ds\right). \quad (4.3)$$

The quantity  $\Lambda(t) = \int_0^t \lambda(s) ds$  is called cumulative hazard at time  $t$ . The survival function is the complementary function to the well-known cumulative distribution function (CDF) of the time-to-failure  $T$ .

When an item is still functioning at a given time  $x > 0$ , one is often interested in the remaining time before the failure. The conditional reliability function at the age  $x$ , denoted by  $R(\cdot \mid x)$ , allows to assess the residual life time. Specifically, the probability that this remaining time is greater than a duration  $t$  is given by :

$$R(t \mid x) = \mathbb{P}(T - x > t \mid T > x) = \frac{R(x+t)}{R(x)}. \quad (4.4)$$

Modelling the survival time  $T$  involves estimating functions  $\lambda(\cdot)$  (often it is preferred to estimate  $\Lambda(\cdot)$  or  $R(\cdot)$ ). Two categories of models are available : parametric models and non-parametric models. Parametric models assume a known shape for the underlying survival function (such as Exponential, Weibull, Log-Normal, Logistic, Log-Logistic, Gamma, etc.), with only the parameter(s)  $\theta$  of finite dimension remaining unknown. Thus,  $\Lambda(\cdot)$  (resp.  $R(\cdot)$ ) is denoted by  $\Lambda_\theta(\cdot)$  (resp.  $R_\theta(\cdot)$ ). The well-known Maximum Likelihood method can be used to estimate the model parameter  $\theta$  even under censoring mechanisms.

In contrast to parametric models, non-parametric models do not assume any specific shape for the function of interest, and the latter is fully estimated without making strong distributional assumptions.

### 4.2.2 Some survival regression models

As aforementioned, survival analysis also examines the effects of factors or covariates affecting the hazard rate. A variety of models, referred as survival regression models, have been developed to handle covariates effects. Survival regression models include proportional hazard models or Cox models (PH models), proportional odds models (PO models), additive hazard models, and accelerated lifetime models, among others. In this paper, we will primarily focus on the most used models which are PH models and PO models. These models are well-known for their ease of interpretation of regression coefficients.

Let us denote by  $Z(t) = (Z^1(t), \dots, Z^p(t))^T$ , the  $p$ -dimensional vector of time-dependent covariates measured at  $t$ . If the  $j^{\text{th}}$  component of  $Z(t)$  is time-independent, then  $Z^j(t)$  is noted  $Z^j$ . We notice that  $u^T$  is the transpose of the vector  $u$ .

#### 4.2.2.1 Proportional hazard models

The Cox PH models, introduced by Cox 1972 assume the hazard rate at  $t$ , conditionally to the history  $Z^*(t) = \{Z(u), 0 \leq u < t\}$  of the covariates vector, to be given by

$$\lambda(t | Z^*(t)) = \lambda_0(t) \exp(\beta^T Z(t)), \quad (4.5)$$

where

- $\lambda_0(\cdot)$  is the baseline hazard function. It is an unknown function that corresponds to hazard when  $Z(t) = 0_{\mathbb{R}^p}$ .
- $\beta = (\beta_1, \dots, \beta_p)$  is the unknown coefficients of regression. Each  $\beta_j, j = 1, \dots, p$ , quantifies the effect size of  $Z^j$  on the hazard.

When the baseline is assumed to be parametric (i.e  $\lambda_0(\cdot) = \lambda_\theta(\cdot)$ ), the resulted model is called parametric Cox model. The model parameters  $\theta$  and  $\beta$  are both estimated by Maximum Likelihood. In the case where no assumption is made on  $\lambda_0(\cdot)$  (this is the semi-parametric Cox model), the finite dimension parameter  $\beta$  and the function  $\Lambda_0(\cdot)$  are fully estimated.

In this study we will use the first category of models which allows to propose an explicit formula for the hazard and the reliability functions.

As mentioned above, regression coefficients of Cox models are simple to interpret. In fact, for a given time  $t$ , we have :

$$\frac{\lambda(t | Z^j(t) = z^j + 1)}{\lambda(t | Z^j(t) = z^j)} = \exp(\beta_j(z^j + 1) - \beta_j z^j) = \exp(\beta_j).$$

This means that if the covariates  $Z^k$  (for all  $k \neq j$ ) are assumed to be constant, an increase of one unit in  $Z^j$  leads to a multiplication of the hazard rate by  $\exp(\beta_j)$ . This implies that a positive (resp. negative) value of  $\beta_j$  indicates that the increasing of one unit in  $Z^j$  leads to the increasing (resp. decreasing) of the hazard rate.

#### 4.2.2.2 Proportional odds models

Generally speaking, the odds of a random event with two outcomes are defined by the ratio of the probability of success to the probability of failure. Based on the same principle, in the PO model, the ratio of the probability  $1 - R(t)$  that the event occurs on the interval  $[0, t]$  to the probability  $R(t)$  that it survives past  $t$  is studied. It assumes the odds for a given value of covariates and the odds with null covariates value to be proportionals. In other words, the PO models assumes that :

$$\frac{1 - R(t | Z^*(t))}{R(t | Z^*(t))} = \frac{1 - R_0(t)}{R_0(t)} \exp(\beta^T Z(t)). \quad (4.6)$$

In this expression :

- The function  $R_0(\cdot)$  is the baseline survival function. As for PH models, PO models can be fully parametric ( $R_0(\cdot) = R_\theta(\cdot)$ ) or semi-parametric. Naturally, the baseline odds is given by  $\frac{1 - R_0(t)}{R_0(t)}$ .
- $\beta = (\beta_1, \dots, \beta_p)$  is the vector of regression coefficients. Each  $\beta_j, j = 1, \dots, p$ , quantifies the effect of  $Z^j$  on the odds. Thus, if  $\beta_j > 0$ , the odds increases when values of  $Z^j$  increase (conditionally that  $Z_i^k, \forall k \neq j$  are assumed to be constant). In contrast, if  $\beta_j < 0$ , the odds decreases when values of  $Z^j$  increase.

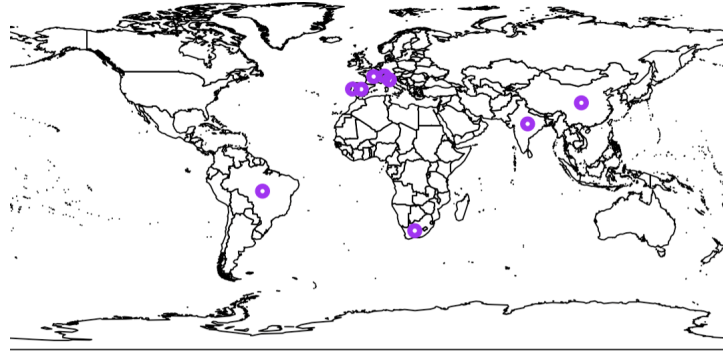


Figure 4.1 – Geographical localisation of studied reinforced concretes.

## 4.3 Survival model of carbonation

In this section, we will apply the models presented above to study the time-to-carbonation and identify the most influential factors affecting carbonation. We will proceed by describing the dataset used, the fitting and model selection procedures, and conclude with an application of the obtained model.

### 4.3.1 The time-to-carbonation dataset

The dataset used for the analysis consists of 793 carbonation measurements obtained from reinforced concrete (r.c) structures operating under real-exposure conditions. These structures are located in various cities around the world, as depicted by the purple dots in Figure 4.1. The dataset, compiled from published carbonation data, is presented and provided in KOUASSI et al. 2022. Along with the carbonation depth ( $Y$ ), the dataset includes additional information, such as the cover depth of steel reinforcement ( $C$ ), the concrete's mean compressive strength ( $f_{cm}$ ), its age ( $X$ ), and the location of the r.c structure.

As seen in equations (4.5) and (4.6), the survival regression models require the histories of time-dependent covariates to model their effects. Therefore, we collected annual mean values of the relative humidity ( $Rh$ ), temperature ( $Tp$ ), and atmospheric concentration of CO<sub>2</sub> ( $P_{CO_2}$ ) from the nearest weather stations of the r.c structures. Additionally, we gathered data on the variable  $Dp$ , which represents the annual proportion of rainy days (the annual number of rainy days with a minimum amount of precipitation water of  $h_{Nd} = 2.5 \text{ mm}$  on 365 days). Since carbonation is a long-term process, for each concrete, we collected values of  $P_{CO_2}$ ,  $Rh$ ,  $Tp$ , and  $Dp$  from its construction year ( $ToM - X$ ) up to the year 2021. These values have been incorporated into the database to consider their influence in the modeling process. Table 4.2 presents a subset of the dataset containing 7 rows. The first three rows represent the collected information for a 3-year-old r.c located in Rome, while the remaining four rows summarize the collected information for a r.c located in Bhopal (India). It is evident that time-independent covariates have the same value over time, and for sheltered concrete, all  $Dp$  values are equal to zero.

The meteorological data was collected from the Copernicus web site (<https://cds.climate.copernicus.eu/>). Unfortunately, it is important to note that, according to our investigation, historical values of the  $P_{CO_2}$  variable are only available from 2003 to 2022.

Finally, the data set used for the analysis contains the variables described in the Table 4.1.

Variable	Description	Unit
$X$	Age of the concrete	[Year]
$Y$	Carbonation depth value at $X$	[mm]
$C$	The cover of the concrete	[mm]
$\delta$	The censoring indicator	[-]
$f_{cm}$	Mean compressive strength	[Mpa]
$t_c$	Curing duration of the concrete	[day]
$P_{co_2}$	Atmospheric $CO_2$ concentration <sup>1</sup>	[ppm]
$Sl$	Indication of concrete's sheltering	[-]
$Rh$	Relative Humidity	[%]
$Dp$	Annual proportion of raining day <sup>1</sup>	[-]
$Tp$	Temperature <sup>1</sup>	[ °C ]
$ToM$	Year of the carbonation measurement	[Year]

<sup>1</sup>Annual average values collected from nearest weather station

**Table 4.1** – Description of the variables in the dataset.

id	X	t_c	f_cm	delta	C	Tp	Rh	Pco2	Dp	Sl	ToM	Location
221	3.19	3	37.0	0	29.89	18.01	76.82	393.26	0.33	NO	2013	Rome
221	3.19	3	37.0	0	29.89	17.77	73.10	393.26	0.25	NO	2013	Rome
221	3.19	3	37.0	0	29.89	18.14	74.15	393.26	0.32	NO	2013	Rome
457	4.00	2	32.5	0	30.00	25.34	63.60	400.05	0.00	Yes	2017	Bhopal
457	4.00	2	32.5	0	30.00	25.79	65.07	400.05	0.00	Yes	2017	Bhopal
457	4.00	2	32.5	0	30.00	25.74	59.66	400.05	0.00	Yes	2017	Bhopal
457	4.00	2	32.5	0	30.00	26.14	60.06	400.05	0.00	Yes	2017	Bhopal

**Table 4.2** – Extract of the dataset used for the analysis

### 4.3.2 Time-to-carbonation model : model fitting and model selection

For each concrete  $i$ , the time-to-carbonation  $T_i$  corresponds to the first moment when the carbonation reaches the reinforcing steel bar. In other words,

$$T_i = \min_{t>0} (Y_i(t) \geq C_i). \quad (4.7)$$

Unfortunately, this time is never observed. In fact, each concrete structure of our dataset was observed only once at the time  $X_i$  and at  $X_i$ , only information that  $i$  is already carbonated (i.e.  $Y_i(X_i) > C_i$ ) or not yet (i.e.  $Y_i(X_i) < C_i$ ) is available. Thus, obtained data are current status data. Survival regression models have been adapted to account for current status data. Let  $\delta = \mathbb{1}(T > X)$  be the censoring indicator variable. Then the available data has the form  $\{(X_i, \delta_i, Z_i^*(X_i)), i = 1, \dots, n\}$ , where  $n$  is the sample size. If  $X_i$  and  $T_i$  are assumed to be independent given the history of covariates, the logarithm of the likelihood function of the sample is given by

$$l_n(\theta, \beta) = \sum_{i=1}^n \delta_i \log \left( R_{(\theta, \beta)}(X_i | Z_i^*(X_i)) \right) + (1 - \delta_i) \log \left( 1 - R_{(\theta, \beta)}(X_i | Z_i^*(X_i)) \right), \quad (4.8)$$

where

$$R_{(\theta, \beta)}(X_i | Z_i^*(X_i)) = \exp \left( -\Lambda_{(\theta, \beta)}(X_i | Z_i^*(X_i)) \right).$$

Various methods have been proposed in the literature to estimate parameters ( $\theta$  and  $\beta$ ) of models that include time-dependent covariates. The main challenge handled in parameters estimation, is to compute the cumulative hazard

$$\Lambda_{(\theta, \beta)}(X_i | Z_i^*(X_i)) = \int_0^{X_i} \lambda_{(\theta, \beta)}(s | Z_i^*(s)) ds$$

when the expression of the covariate process is known.

In our study, as time-dependent covariates are updated at discrete time, we will use procedure similar to the one of SPARLING et al. 2006 to compute the cumulative hazard at the monitoring time  $X_i$ . In fact, for the  $i$ th item, let  $\tau_{i0}, \dots, \tau_{ik_i}$  be the  $k_i$  update times of  $Z_i$  such that  $Z_i(t)$  is assumed to be constant on the interval  $]\tau_{il}, \tau_{i(l+1)}]$  (also equivalent to assume that the effect of the covariate on  $]\tau_{il}, \tau_{i(l+1)}]$  is constant)  $\forall l = 0, 1, \dots, (k_i - 1)$ . Then,  $\forall t \geq 0$ ,

$$\Lambda(t | Z_i(t)) = \sum_{l=0}^{k_i-1} \mathbb{1}(t \leq \tau_{i(l+1)}) \int_{\tau_{il}}^{\tau_{i(l+1)}} \lambda(s | Z_i(s)) ds \quad (4.9)$$

It is important to note that for each item  $i$ ,  $k_i = \lfloor X_i \rfloor$ , and in the sequence of times  $\{\tau_{i0}, \dots, \tau_{ik_i}\}$ ,  $\tau_{i0}$  corresponds to the age of item  $i$  at its year of construction (i.e.,  $\tau_{i0} = 0$ ). In this case, for all  $j = 1, \dots, p$ , the quantity  $Z^j(\tau_{il})$  represents the value of covariate  $j$  at age  $l$ , for all  $l = 0, \dots, k_i$ , which corresponds to the value of covariate  $j$  at the year  $ToM_i - k_i + l$ .

For example, let us consider item 221 (i.e.,  $i = 221$ ), which is  $X_{221} = 3.19$  years old, with the year of measurement being  $ToM_{221} = 2013$  (see Table 4.1). For this item, we have  $k_{221} = k_i = \lfloor X_i \rfloor = 3$ ,  $\{\tau_{i0}, \dots, \tau_{ik_i}\} = \{0, 1, 2, 3\}$ , and  $\{Z^j(\tau_{i0}), \dots, Z^j(\tau_{ik_i})\}$  corresponds to  $\{Z^j(2010), Z^j(2011), Z^j(2012), Z^j(2013)\}$ , for all  $j = 1, \dots, p$ . If we consider the  $Rh$  covariate, for example, we have  $\{Z^j(\tau_{i0}), \dots, Z^j(\tau_{ik_i})\} = \{76.82, 73.10, 74.15\}$  for this item. In addition, as shown in Table 4.1, the values of time-independent covariates remain the same among the years (see e.g., the values for the  $t_c$  covariate).

Since the covariate process remains constant within each interval  $]\tau_{il}, \tau_{i(l+1)}]$ , computing the value of the integral in equation (4.9) is straightforward.

As previously mentioned, our main objective is to identify the covariates ( $C$ ,  $f_{cm}$ ,  $t_c$ ,  $P_{CO_2}$ ,  $Rh$ ,  $Tp$ , and  $Dp$ ) that have the most significant impact on the time-to-carbonation. In the next section, we will successively employ both the PH model and PO models for this purpose.

#### 4.3.2.1 Proportional hazard model of the time-to-carbonation

Let us assume that  $\lambda(s | Z_i(s))$  is given according a PH model. Under this assumption, we have :

$$\begin{aligned} \Lambda_{(\theta, \beta)}(X_i | Z_i^*(X_i)) &= \sum_{l=0}^{k_i-1} \mathbb{1}(X_i \leq \tau_{i(l+1)}) \int_{\tau_{il}}^{\tau_{i(l+1)}} \lambda_{\theta}(s) e^{\beta^T Z_i(s)} ds \\ &= \sum_{l=0}^{k_i-1} \mathbb{1}(X_i \leq \tau_{i(l+1)}) (\Lambda_{\theta}(\tau_{i(l+1)}) - \Lambda_{\theta}(\tau_{il})) e^{\beta^T Z_i(\tau_{i(l+1)})} \end{aligned}$$



Thus, when the cumulative hazard is replaced by the above expression in the equation (4.8), the fitting procedure consists of estimating  $\theta$  and  $\beta$  by maximizing the log-likelihood function  $l_n$  after identifying the shape of the cumulative baseline hazard (i.e.  $\Lambda_\theta(\cdot)$ ) that best fit our data.

To identify the best shape of  $\Lambda_\theta(\cdot)$ , we fit a semi-parametric PH model. That is to consider the model with cumulative hazard function given by

$$\Lambda_{(\Lambda_0, \beta)}(X_i) = \sum_{l=0}^{k_i-1} \mathbb{1}(X_i \leq \tau_{i(l+1)}) (\Lambda_0(\tau_{i(l+1)}) - \Lambda_0(\tau_{il})) e^{\beta^T Z_i(\tau_{i(l+1)})}$$

where  $\Lambda_0(\cdot)$  is a non-negative and increasing function.

The estimator  $\hat{\Lambda}_0(\cdot)$  of  $\Lambda_0(\cdot)$  is obtained by using the estimation procedure proposed in WELLNER et ZHAN 1997 and then, both graphical methods and Kolmogorov-Smirnov goodness of fit test (BAGDONAVIČIUS et al. 2011) have been used to identify the best parametric shape of  $\hat{\Lambda}_0(\cdot)$ .

On the Figure 4.2, the graphic (a) shows the Weibull plot (i.e  $\log(\hat{\Lambda}_0(t))$  versus  $\log(t)$ ) and the graphic (b) shows the empirical survival  $R_0(\cdot) = \exp(-\hat{\Lambda}_0(\cdot))$  curve and the Weibull survival curve. As it can be seen, empirical functions are well fitted by Weibull functions. Indeed, the Weibull plot shows a linear relationship between  $\log(\hat{\Lambda}_0(t))$  and  $\log(t)$ . Additionally, the graph (b) demonstrates that the curves of the empirical reliability function and the Weibull reliability function are close to each other. These graphical results are confirmed by the Kolmogorov-Smirnov goodness-of-fit test (p-value = 0.89).

Thus, in the remainder of our analysis, we will consider  $\Lambda_0(\cdot)$  to be the cumulative hazard function of the Weibull distribution, denoted as  $\Lambda_{(a,b)}(\cdot)$ . We recall that for all  $t \geq 0$ ,

$$\Lambda_{(a,b)}(t) = \left(\frac{t}{b}\right)^a,$$

where  $a$  and  $b$  are two positive real values. The parameters  $a$  and  $b$  are called the shape parameter and the scale parameter, respectively.

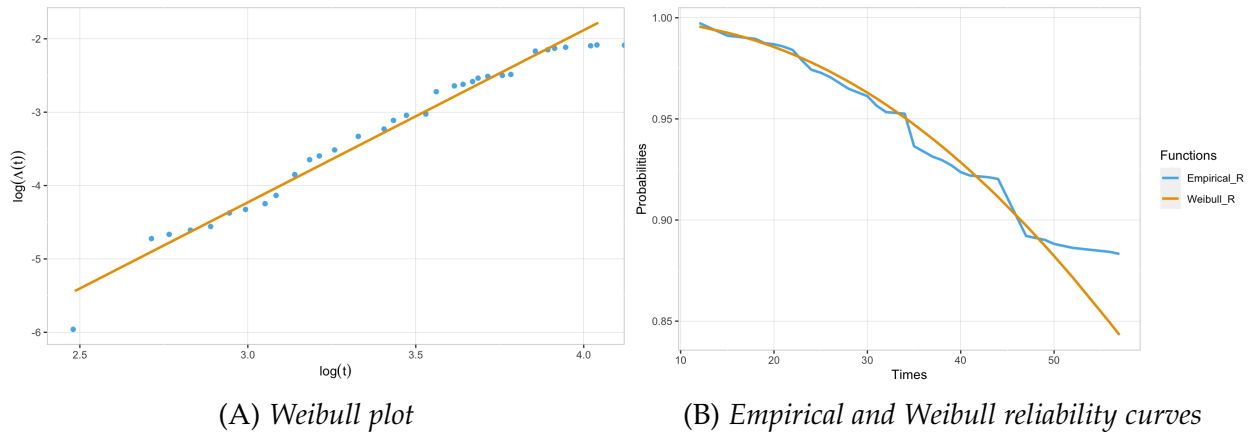


Figure 4.2 – Graphical goodness-of-fit analysis

Once the most appropriated baseline has been determined, the next objective is to identify the most influential factors. For our study, the available covariate vector for a concrete  $i$  at a given time  $t$  is  $Z(t) = (C_i, f_{cm_i}, t_{c_i}, Rh_i(t), Dp_i(t), Tp_i(t), P_{co_2i}(t))^T$ . We recall that  $u^T$  is the transposed vector of  $u$ .

Also, since the complete history of  $P_{co_2}$  covariate is unavailable, we adopt two modelling approaches.

The first approach treats the  $P_{CO_2}$  covariate as time-dependent and its history before 2003 is estimated using a model. Indeed, the history of  $P_{CO_2}$  for each locality (city) is not available before this date, but the USA's National Oceanic and Atmospheric Administration provides the monthly mean trend of  $P_{CO_2}$  from 1980 to 2003 (available on <https://gml.noaa.gov/ccgg/trends/global.html>). Thus, the scaled trend of  $P_{CO_2}$  can be expressed as

$$\tilde{P}_{CO_2} = \log \left( \frac{P_{CO_2} - \bar{P}_{CO_2}}{sd(P_{CO_2})} \right),$$

where  $\bar{P}_{CO_2}$  is the mean value and  $sd(P_{CO_2})$  is the standard deviation of  $P_{CO_2}$  of values obtained on the period 1980-2003. To model the temporal variability of atmospheric CO<sub>2</sub> trends, we use the transformed variable  $\tilde{P}_{CO_2}$  where  $t' = t - 1980$  represents the difference between 1980 and the observation time  $t$ . By fitting different polynomial regression models, including the first order (*Poly1* model), second order (*Poly2* model), and third order (*Poly3* model), we find that the *Poly2* and *Poly3* models provide the best fit for  $\tilde{P}_{CO_2}$  (see Figure 4.3). Upon analyzing the Mean Squared Error (MSE) values of these models, we observe that the *Poly2* model outperforms the *Poly3* model. Hence,  $\tilde{P}_{CO_2}$  can be estimated as follows :

$$\hat{\tilde{P}}_{CO_2}(t') = -1.31 \times 10^{-2} + 1.07 \times 10^{-2}t' + 9.6 \times 10^{-5}t'^2. \quad (4.10)$$

Thus, for each observation time  $t < 2003$ , the estimated annual mean value of  $P_{CO_2}$  for each locality is given by  $\hat{P}_{CO_2}(t)$  :

$$\hat{P}_{CO_2}(t) = sd(P_{CO_{2loc}}) \exp \left( \hat{\tilde{P}}_{CO_2}(t') \right) + \bar{P}_{CO_{2loc}}, \quad (4.11)$$

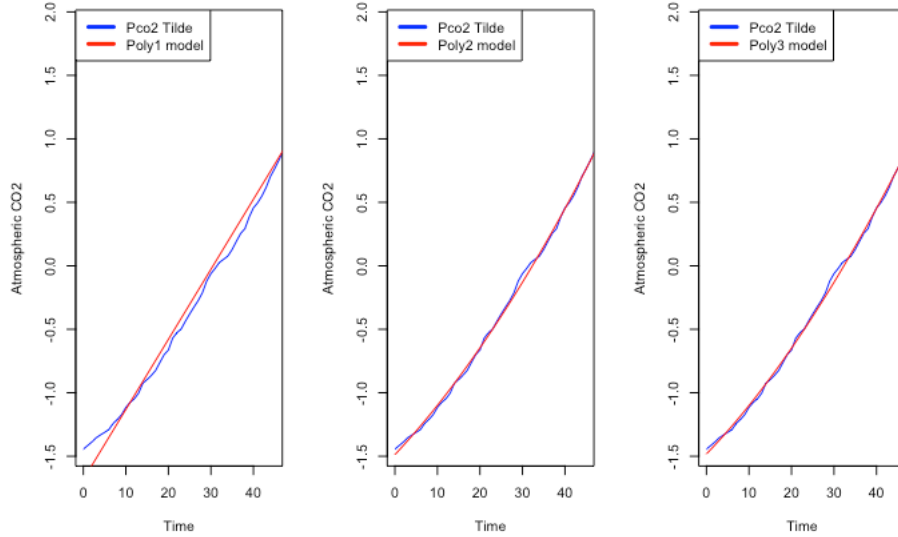
where  $\bar{P}_{CO_{2loc}}$  and  $sd(P_{CO_{2loc}})$  are respectively the mean value and the standard deviation of the available atmospheric CO<sub>2</sub> values of the locality (i.e. observed values from 2003 to 2022) and  $t' = t - 1980$ .

By using the equation (4.11), the history of the  $P_{CO_2}$  variable for each locality has been provided and a PH model has been fitted.

Furthermore, since our aim is to identify the most influencing factors, a model selection has been performed. That was to find the model containing only variables with significant effect on the time-to-carbonation. In survival regression models, the influence of a variable can be assessed by a Wald test with null hypothesis  $\mathbf{H}_0 : \beta_j = 0$ . Thus, the effect of the  $j$ th covariate is considered as significant if the p-value of the test is less than a threshold value which is usually equals to 5%. When model parameters have been estimated and a model selection has been performed, the PH model that best fits our data has been identified. The estimated parameters, their standard deviation and p-values of Wald tests are presented in the Table 4.3. The  $\beta_0$  parameter is such that  $b = e^{-\frac{\beta_0}{a}}$ .

Parameter	Estimate	Std Error	p-value
$a$	2.27	0.5	$10^{-4}$
$\beta_0$	-83.3	18.32	$5 \times 10^{-4}$
$\beta_C$	-0.15	0.04	$5 \times 10^{-5}$
$\beta_{Tp}$	0.18	0.05	$10^{-3}$
$\beta_{Cs}$	0.2	0.06	$2 \times 10^{-3}$

**Table 4.3** – The estimated values of the PH model parameters, their standard errors and the corresponding p-values of Wald test when  $P_{CO_2}$  is considered as time-dependant.



**Figure 4.3** – Polynomials model fitting of  $\tilde{P}_{CO_2}$  : from left to right, the blue lines represent the observed values of  $\tilde{P}_{CO_2}$ , and the red lines represent the predicted values obtained using the Poly1, Poly2, and Poly3 models, respectively.

The results indicate that the cover, temperature, and atmospheric concentration of  $CO_2$  are the only variables that significantly affect the time-to-carbonation. Interestingly, the variable  $Rh$ , which is important in the chemical reactions related to carbonation, is found to be insignificant in this model. To ensure that the results obtained from the model were not influenced by the procedure used to estimate the missing history of the  $P_{CO_2}$  variable, we fitted a proportional hazards (PH) model with the significant three variables mentioned above, by using the estimated history of  $P_{CO_2}$  obtained with the *Poly3* model (which performs similarly to the *Poly2* model). The analysis shows that the effect of  $P_{CO_2}$  is not considered significant in this case. This indicates that the procedure used to estimate the missing history of  $P_{CO_2}$  has an impact on the PH model obtained earlier.

Therefore, in the second approach, the  $P_{CO_2}$  variable is considered as time-independent. For each concrete  $i$  and for an observation time  $t$ ,  $P_{CO_2i}(t) = P_{CO_2i} = p_{CO_2i}$ , with  $p_{CO_2i}$  the median value of the available history for the  $i$ th concrete (i.e. observed values of  $P_{CO_2}$  from 2003 to 2022). Under this assumption, the PH model has been fitted and a model selection has been performed. The results reveal that the effect of the  $Dp$  variable is most significant when it is considered as time independent. i.e  $Dp_i(t) = Dp_i = Dp_{m_i}$ , where  $Dp_{m_i}$  is the median value of the available history for the  $i$ th concrete. The estimated parameters, their standard error and p-values of Wald test of the most suitable PH model are reported in Table (4.4).

It is revealed that excepted for the curing variable  $t_c$ , all features have significant effect on the time-to-carbonation. This latest model thus considers the effect of multiple carbonation factors.

#### 4.3.2.2 Proportional Odds Model of the Time-to-carbonation

In this section we assume that the odd of the carbonation, given the history  $Z^*(t)$  of covariates is estimated according a PO model. The equation (4.6) can be written as :

$$-\text{logit}(R(t | Z^*(t))) = -\text{logit}(R_0(t)) \exp(\beta^T \mathbf{Z}(t)), \quad (4.12)$$

Parameter	Estimate	Std Error	p-value
$a$	1.16	0.21	$2.28 \times 10^{-8}$
$\beta_0$	-80	34.1	$1.9 \times 10^{-2}$
$\beta_C$	-0.32	0.03	$5.55 \times 10^{-16}$
$\beta_{f_{cm}}$	-0.04	0.01	$6 \times 10^{-4}$
$\beta_{Dp}$	14.5	2.19	$2.26 \times 10^{-11}$
$\beta_{P_{co_2}}$	0.21	0.08	$1.2 \times 10^{-2}$
$\beta_{Rh}$	-0.1	0.02	$1.56 \times 10^{-11}$
$\beta_{Tp}$	0.28	0.04	$7.08 \times 10^{-7}$

**Table 4.4** – The estimated values of PH model parameters, their standard errors and the corresponding p-values of Wald test when  $P_{co_2}$  and  $Dp$  are considered as time-independent.

where the *logit* function is defined on  $]0,1[$  such that  $logit(p) = \log(\frac{p}{1-p})$ . Let  $H(t) = -logit(R_0(t))$ . Then the log likelihood function under proportional odds assumption with a parametric baseline is given by

$$\ln(\beta, \theta) = \sum_i^n \delta_i \left( H(X_i) + \beta^T \mathbf{Z}(X_i) \right) - \log \left( 1 + \exp(H(X_i) + \beta^T \mathbf{Z}(X_i)) \right) \quad (4.13)$$

As previously shown, we assume the baseline  $R_0(\cdot)$  to have a Weibull shape. In other words,  $R_0(t) = R_\theta(t) = \exp\left(-\left(\frac{x}{b}\right)^a\right)$ , where  $\theta = (a, b)$  with  $a$  and  $b$  two positive reals. Similarly to the PH models, the fitting procedure consists of estimating parameters  $\beta$  and  $\theta$  and performing a model selection.

As aforementioned, the complete history of the  $P_{co_2}$  variable is unavailable and the survival models fitted with estimated histories are unreliable because the results change depending on the estimated values of  $P_{co_2}$  and the effect of  $Rh$  is not revealed as significant. Thus, the  $P_{co_2}$  variable is considered as time-independent. The Table (4.5) gives the estimated parameters, their standard errors and their Wald test p-values. As it can be seen, the environmental factors  $Rh$  and  $P_{co_2}$  which play an important role in the carbonation process (chemical point of view) are not considered as significant factors by PO models.

Parameter	Estimate	Std Error	p-value
$a$	1.6	0.3	$9.2 \times 10^{-6}$
$\beta_0$	-85.6	49.3	$4.2 \times 10^{-2}$
$\beta_C$	-0.35	0.04	$5.9 \times 10^{-14}$
$\beta_{f_{cm}}$	-0.06	0.01	$3 \times 10^{-4}$
$\beta_{t_c}$	-0.5	0.28	$3 \times 10^{-3}$
$\beta_{Dp}$	7.05	1.35	$1.98 \times 10^{-7}$
$\beta_{Tp}$	0.35	0.07	$1.6 \times 10^{-6}$

**Table 4.5** – PO model : estimated parameters, their standard errors and their p-values of the Wald test

### 4.3.2.3 Selection of the best time-to-carbonation model

Both models (PH and PO models) fit data well. However, considering the objective of our analysis, which is to study the effect of environmental characteristics on the time-to-carbonation, as well as the ease of interpretation of the PH model, the latter will be preferred over the PO model. Thus, for a concrete  $i$  with known history  $Z^*(t)$  of covariate vector  $Z(t)$  is equals to  $(C_i, f_{cm_i}, P_{co_2_i}, Dp_i, Rh_i(t), Tp_i(t))^T$ , the hazard rate at  $t$  will be computed by using the formula from Equation (4.14).

$$\lambda_i(t | Z^*(t)) = at^{a-1}e^{(-80-0.32C_i+14.5Dp_i-0.04f_{cm_i}+0.2P_{co_2_i}-0.1Rh_i(t)+0.28Tp_i(t))}, \quad (4.14)$$

where  $a = 1.5$ .

Let's consider two materials,  $i$  and  $i'$ , located in the same place (i.e., having the same historical values of  $Dp$ ,  $Tp$ ,  $Rh$ , and  $P_{co_2}$ ). Additionally, let's assume that these concretes have the same value of  $f_{cm}$  at a given time  $t$ . In this scenario, the material with the largest cover depth will have the lowest hazard risk (or highest reliability probability) at time  $t$ . Specifically, if  $C_{i'} = C_i + 1$ , then at time  $t$ , the hazard of  $i'$  is approximately  $e^{-0.32}$  times the hazard of  $i$ . Similarly, if we swap the place of  $f_{cm}$  and  $C$ , we will see that the failure risk at time  $t$  of  $i'$  will be approximately  $e^{-0.04}$  times the failure rate of  $i$ .

Furthermore, if we assume that  $i$  and  $i'$  have the same values of cover and compressive strength at time  $t$ , and the only difference between the two items is the value of  $P_{co_2}$ , then the material with the largest value of  $P_{co_2}$  will have the largest hazard rate. More precisely, if, for example,  $P_{co_2_{i'}} = P_{co_2_i} + 1$ , then  $\lambda_{i'} \approx e^{0.2}\lambda_i$ .

However, the analysis of coefficients  $\beta_{Dp}$ ,  $\beta_{Rh}$ , and  $\beta_{Tp}$  requires careful consideration. These factors are inherently interconnected, and assuming, for example, that materials  $i$  and  $i'$  have the same values of  $Rh$  and  $Dp$  but significantly different values of  $Tp$ , or vice versa, would be oversimplifying the complex relationships between them.

### 4.3.3 Some applications of the proposed model

In this section, we apply the proposed PH model mentioned above to assess the reliability of an unsheltered reinforced concrete structure subjected to natural carbonation. We consider three reinforced concretes assumed to be built in 1986, and their values of cover ( $C$ ) and compressive strength ( $f_{cm}$ ) are provided in Table 4.6. Using our model, we will sequentially analyze the impact of carbonation factors and assess the reliability under various climate change scenarios.

	Cover [mm]	$f_{cm}$ [MPa]
<b>Concrete 1</b>	36	43.28
<b>Concrete 2</b>	36	33.28
<b>Concrete 3</b>	26	43.28

Table 4.6 – Cover and compressive strength values of the concretes.

#### 4.3.3.1 Effect of carbonation factors on the reliability of the reinforced concrete under real climates.

To assess the impact of carbonation factors, we compare the reliability curves of three reinforced concrete structures (concretes 1, 2, and 3) under different climatic conditions in

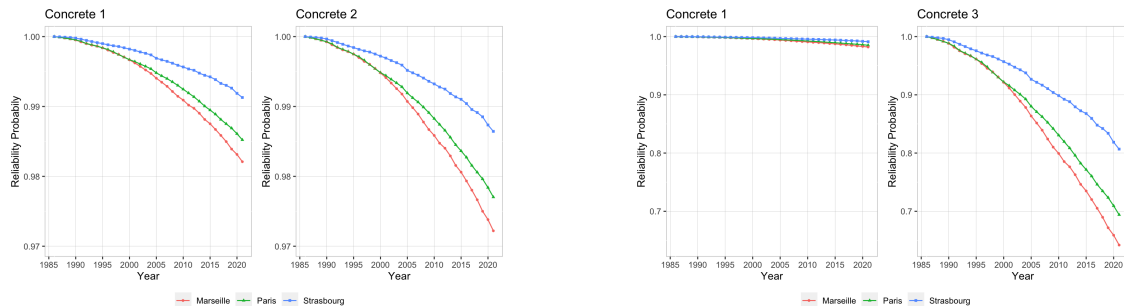
three French cities : Paris, Marseille, and Strasbourg. The mean values of environmental characteristics for these cities during the period from 1986 to 2021 are presented in Table 4.7. The reliability (or survival) curves have been computed by calculating the hazard rates using the proposed relationship (Equation (4.14)), and they are depicted in Figure 4.4.

	$T_p$ [°C]	$R_h$ [%]	$P_{CO_2}$ [ppm]	$D_p$ [day]
<b>Marseille</b>	15.58	69.2	394.2	81.3
<b>Paris</b>	14.45	80.8	394.3	117.0
<b>Strasbourg</b>	10.3	77.5	394.0	124.3

Table 4.7 – Mean values of environmental characteristics of the three cities.

The reliability (or survival) curves are computed by calculating the hazard rates using the proposed relationship (equation (4.14)), and they are depicted in Figure (4.4).

The analysis of the results indicates that the reliability curve is consistently higher for the concrete placed in Strasbourg compared to Marseille and Paris, regardless of the concrete characteristics. This can be attributed to the unique environmental conditions in Strasbourg, where the concrete pores become saturated with water due to the humid climate and moderate temperatures. As a result, the diffusion of CO<sub>2</sub> gas in the porous network is slowed down, leading to a delayed carbonation process. On the other hand, in Marseille, the diffusion of CO<sub>2</sub> into the pores of the reinforced concrete is faster, resulting in a more rapid carbonation process. This explains why the survival curve of this city is the lowest among the three locations.



(A) Concrete 1 Vs Concrete 2

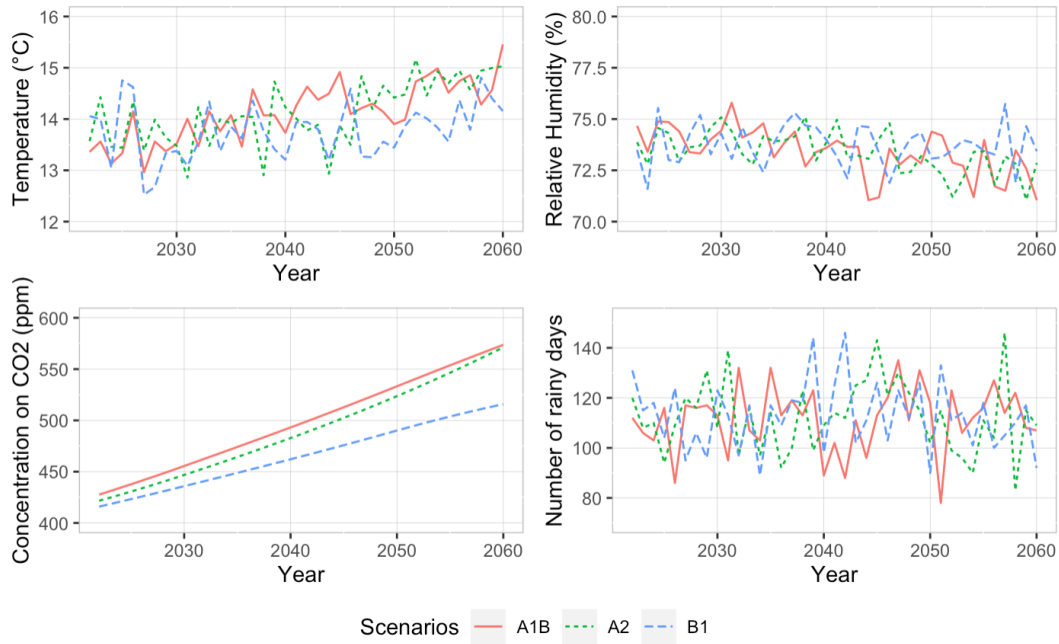
(B) Concrete 1 Vs Concrete 3

Figure 4.4 – Reliability curves of the reinforced concretes *concrete 1*, *concrete 2* and *concrete 3* when they are located in Marseille, Paris and Strasbourg respectively.

Furthermore, by analyzing the graphs on the left side of the Figure 4.4 (A), it is shown that concretes with the lowest compressive strength values have the lowest survival probabilities at 35 years (i.e. in 2021). For instance, in the city of Marseille, the survival probability after 35 years of operation for *concrete 1* is approximately 0.98, while it is 0.97 for *concrete 2* (indicating a 1% drop in reliability). In contrast, the graphs on the right side (B) demonstrate the impact of the cover value on the reliability of reinforced concrete structures subjected to carbonation. It is observed that in the city of Marseille, a reduction of 10mm in the cover value leads to a decrease of approximately 34.7% in reliability.

#### 4.3.3.2 Analysis of the carbonation of a reinforced concrete under climate change scenarios

In light of the findings from the previous section, it is evident that the reliability of *concrete 1* remains relatively high regardless of the city. Thus, we are interested in estimating the residual



**Figure 4.5** – Projection of the temperature, the relative humidity, the number of rainy days and the atmospheric concentration on  $\text{CO}_2$  for Paris under the scenarios **A1B**, **A2** and **B1**.

life time of this concrete, which refers to the additional time that the concrete can remain operational before carbonation occurs. To achieve this, we will use our model to evaluate the residual reliability at the age of  $x = 35$  years (using Equation (4.4)) for this reinforced concrete over the next 20 years (i.e., the period from 2022 to 2042).

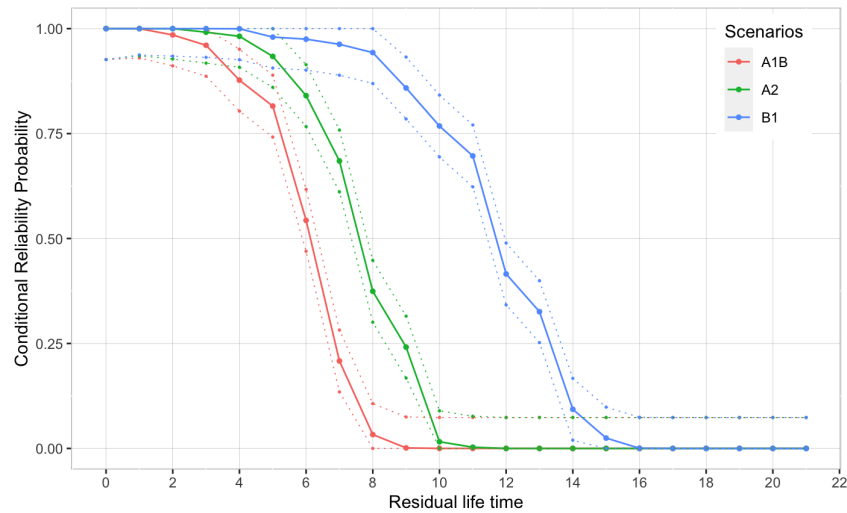
As the values of the environmental factors are not currently observable, we are considering climate warming scenarios labelled as **A1B**, **A2**, and **B1**, which have been proposed by the Intergovernmental Panel on Climate Change (IPCC) in IPCC 2007. These scenarios offer projections for future climate conditions and are based on various assumptions, including population growth, economic development, technology transfer, and other factors that contribute to predicting climate change. We have chosen these scenarios because the estimates for the values of environmental factors (temperature, relative humidity, rainy days, and atmospheric  $\text{CO}_2$  concentration) are available for Paris until 2100.

In brief, the **A2** scenario represents a pessimistic future world, while the **B1** scenario represents an optimistic future world. The **A1B** scenario falls between these two extremes and represents a future world with moderate assumptions. Each scenario provides a different trajectory for the evolution of environmental factors. Figure 4.5 illustrates the predicted values of  $Rh$ ,  $Tp$ ,  $Dp$ , and  $P_{\text{CO}_2}$  from 2022 to 2060 based on these scenarios.

The reinforced concrete (r.c) is assumed to be placed in Paris, and the graphs in Figure 6 show the conditional reliability functions at an age of  $x = 35$  years for the concrete under the scenarios **A2**, **A1B**, and **B1**. We observe that the median residual life time of failure (the time at which the reliability curve intersects the horizontal line with equation  $y = 0.5$ ) is 6 years under **A1B**, 8 years under **A2**, and 12 years under **B1**.

The median residual life time under the pessimistic scenario **A2** is greater than the median residual life time under the moderate scenario **A1B**. These results are attributed to the differences in the mean values of  $P_{\text{CO}_2}$  between these scenarios during the period from 2022 to 2050, while the values of  $Rh$  and  $Tp$  remain approximately the same. In particular, the mean value of  $P_{\text{CO}_2}$  is higher under the **A1B** scenario compared to the **A2** scenario during this interval. This higher  $\text{CO}_2$  concentration in the atmosphere accelerates the carbonation process, leading

to a shorter median residual life under the **A1B** scenario compared to the **A2** scenario, which represents a more pessimistic view of the future.



**Figure 4.6** – Conditional Reliability curves at age  $x=35$  years old of the considered concrete (assumed located in Paris) from 2022 to 2042 under climate change scenarios **A1B**, **A2** and **B1**. The dotted lines represent the corresponding piecewise confident intervals

## 4.4 Conclusion

The survival models provide a set of statistical methods for analyzing carbonation time and its main influencing factors. By using the Cox proportional hazards models, we have developed a carbonation time model that takes into account the temporal variability of environmental characteristics, allowing for a better quantification of their impact on the time-to-carbonation. This model enables the assessment of the reliability of reinforced concrete exposed to natural carbonation under observed climatic conditions, as well as under climate warming scenarios. However, due to the unavailability of complete historical data on atmospheric CO<sub>2</sub> concentrations at the exposure sites, this variable is treated as independent of time in the models. Therefore, if it were possible to include this aspect, the performance of our model would likely be improved.



---

## An Extended Geometric Failure Rate Reduction model with covariates

---

Comme mentionné précédemment, la modélisation de la maintenance imparfaite continue d'être un domaine de recherche actif. Étant donné que les moments de défaillance sont de nature aléatoire, les processus stochastiques sont largement utilisés pour étudier ces événements. Ce travail vise à améliorer les performances de ces modèles en introduisant les covariables dans les modèles à réduction géométrique du taux de défaillance (ou Geometric Failure Rate Reduction en anglais). Ce chapitre, qui décrit les résultats de ce travail, constitue une version provisoire d'une publication qui sera également soumise à une revue à comité de lecture international.

## 5.1 Introduction

Engineers are constantly seeking to improve the reliability of repairable systems by analyzing and understanding the effectiveness of repair actions or corrective maintenance. This crucial issue has been the subject of numerous reliability studies. The most known maintenance assumptions among authors are minimal maintenance, known as "As Bad As Old" (ABAO) and perfect maintenance, referred as "As Good As New" (AGAN). In minimal maintenance, the system is brought back to its state before the failure occurred. On the other extreme case, perfect maintenance involves restoring the system to a brand-new condition, which may be the case if one replace it with an identical and new one.

Both types of maintenance have been extensively studied in the literature. Under the minimal repair assumption, the recurrence of failures is modeled by a nonhomogeneous Poisson process (NHPP), while under the perfect maintenance assumption, a renewal process (RP) is used. These models provide valuable insights on the behavior of repairable systems and their maintenance strategies.

However, in practical situations, maintenance actions are often imperfect, resulting in a state for the item that lies between "as good as new" and "as bad as old". To address this reality, researchers have proposed numerous imperfect maintenance models. Some of the most well-known models include the BP model introduced by BORDES et MERCIER 1983, the virtual age models introduced by KIJIMA 1989, the arithmetic reduction models (Arithmetic Reduction of failure Intensity (ARI) and the Arithmetic Reduction of Virtual Age (ARA)) introduced by DOYEN et GAUDOIN 2004, as well as the Geometric Process (GP) model proposed by LAM 1998. These models provide valuable tools for analyzing the reliability of repairable systems under imperfect maintenance scenarios and understanding the behavior of these systems.

In this work, we will focused on the Geometric Process models. A nonparametric inference for this model, which involves geometric reduction of the times between failures (or interarrival times), has been performed by LAM 1992. One of the simplest GP models is the Geometric Failure Rate Reduction model (GFRR) proposed by FINKELSTEIN 2008. This model is built upon the same concept as the ARI model, with the key difference being its reliance on the assumption of a geometric reduction in the intensity process rather than an arithmetic reduction, as in the ARI model. However, the GFRR model, as defined by FINKELSTEIN 2008, encounters the "explosion" problem. To address this issue, DAUXOIS et al. 2018 introduce an extended version of this model known as Extended Geometric Failure Rate Reduction (EGFRR) model. The EGFRR model allows to reduce explosion and better capture the behavior of real systems. Drawing inspiration from the approach used by BORDES et MERCIER 2013 to extend the GP model, the EGFRR model incorporates a functional parameter that allows to control the influence of past repairs on the future reliability of the system.

Integrating the effect of covariates is a promising approach to increase the performance of the imperfect maintenance models. It is observed that seemingly identical systems have different behaviors, which can be explained by various factors such as environmental conditions. For example, a wind turbine functioning in a windy region is likely to have shorter failure times compared to one located in a less windy region. These characteristics or covariates, which may be time-dependent or time-invariant, provide crucial information to improve the modeling of the failure process of repairable systems.

This aspect (the study of covariate effects), has already been considered by several authors in the reliability assessment of repairable system. In the simplest case of the ABAO model, one can cite NOVÁK 2015, LI et al. 2016, and NING et al. 2020. The first work proposes an ABAO model with time-independent covariates, while the second incorporates the influence of time-dependent covariates. The last work examines the impact of covariates on a bivariate ABAO. In the broader context of imperfect maintenance, the works of WU et SCARF 2015

introduce an ARA model with covariates and BRENIÈRE 2021 presents a Virtual Age model with covariates.

However, to the best of our knowledge, the incorporation of covariate effects in GP models has not been explored in the existing literature. Therefore, in this study, we propose to introduce and perform statistical inference on an Extended Geometric Failure Rate Reduction model with covariates. The first two sections of this paper are devoted to the introduction of the EGFR model without (Section 5.2) and with (Section 5.3) covariates. A nonparametric estimation procedure for estimating model parameters is presented in Section 5.4. Subsequently, the asymptotic properties study and a Monte Carlo simulation study of the estimators are performed in Sections 5.5 and 5.6. Finally, in Section 5.7, we draw conclusions from our study and discuss potential future directions for research.

## 5.2 Geometric Failure Rate Reduction models without covariates

Let us consider repairable systems that undergoes both perfect and imperfect maintenance. Specifically, we assume that perfect maintenance is carried out after a predetermined number  $K$  of imperfect maintenance actions. We assume that the maintenance period is negligible and we denote the successive failure times by

$$0 = T_0 < T_1 < T_2 < \dots < T_k < \dots < .$$

Let  $(N_t)_{t \in \mathbb{R}^+}$  be the counting process of the number of failures up to  $t$  and  $\lambda_t$  its intensity process. Recall that, for all  $t > 0$ , the intensity process is defined by

$$\lambda_t = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \mathbb{P} \left( N_{(t+\Delta t)^-} - N_{t^-} = 1 \mid \mathcal{H}_{t^-} \right),$$

where  $\mathcal{H}_{t^-}$  represents the history of the process up to the time  $t$ .

For this sequence of successive failure times, let  $X_k = T_i - T_{k-1}$ , for all  $k \in \mathbb{N}^*$ , be the interfailure times. We denote by  $\lambda_{X_k}(\cdot)$  the instantaneous hazard function of the random variable  $X_k$ . The instantaneous hazard rate or failure rate  $\lambda_{X_k}(t)$  at a given time  $t$  is defined by :

$$\lambda_{X_k}(t) = \lim_{h \rightarrow 0} \frac{1}{h} \mathbb{P} (t \leq X_k < t + h \mid X_k > t).$$

Imperfect maintenance actions are carried out according to a Geometric Failure Rate Reduction (GFRR) model as proposed by FINKELSTEIN 2008 when the intensity process is defined as follows :

$$\lambda_t = \rho^{N_{t^-}} \lambda_{X_1} (t - T_{N_{t^-}}). \quad (5.1)$$

The  $\lambda_{X_1}(\cdot)$  function represents the instantaneous hazard function of the first interarrival time  $X_1 = T_1 - T_0$  and  $\rho$  is a strictly positive parameter helps to quantify the effect size of the repairs actions  $N_{t^-}$  carried out before  $t$ . Thus, as seen in (5.1), the GFRR model assumes that the failure intensity of the failure process at a given time  $t$  depends not only on the time elapsed since the last repair before  $t$  but also on the number of past repairs. This is achieved by introducing the  $\rho$  parameter. From Equation (5.1), it can be seen that the perfect repair model (AGAN) is obtained with  $\rho = 1$ , indicating that maintenance action performed before  $t$  has restored the system to a new state.

As aforementioned, the GFRR model, as defined in equation (5.1), suffers from the "explosion" phenomenon. If the value of  $\rho$  deviates notably from 1, the time intervals between failures

either rapidly decrease or increase, which fails to accurately capture the behavior of actual systems. Thus, following a similar approach to that proposed by BORDES et MERCIER 2013, DAUXOIS et al. 2018 suggest an extension of these models referred to as the Extended Geometric Failure Rate Reduction (EGFRR) model. The intensity process of an EGFRR model is given by :

$$\lambda_t = \rho^{a(N_{t^-}+1)} \lambda_{X_1}(t - T_{N_{t^-}}), \quad (5.2)$$

where :

- $k \mapsto a(k)$  is a known and non-decreasing function such that  $a(1) = 0$ ;
- the parameter  $\rho$  is defined as in (5.1).

Under the assumption that the intensity of the failure process is given according to an EGFRR model, it is interesting to note from ledoux that one has

$$\lambda_{X_k}(t) = \rho^{a(k+1)} \lambda_{X_1}(t). \quad (5.3)$$

This clearly illustrates the geometric reduction of the interarrival times intensity. The relation (5.3) can also be expressed using cumulative hazard functions as follows :

$$\Lambda_{X_k}(t) = \rho^{a(k+1)} \Lambda_{X_1}(t).$$

Based on the observations of the interfailure times  $X_k$ , DAUXOIS et al. 2018 adopts a semiparametric approach and proposes estimators  $\hat{\rho}$  and  $\hat{\Lambda}_{X_k}(\cdot)$  for the Euclidean parameter  $\rho$  and the functional parameters  $\Lambda_{X_k}(\cdot)$ , respectively. Furthermore, they establish the asymptotic normality and study the large sample properties of these estimators.

### 5.3 Extended Geometric Failure Rate Reduction with covariates

As previously stated, the reliability and safety of a system can be affected by its specific characteristics. To evaluate how these attributes (also called covariates) influence the incidence of system failures, we suggest to integrate the effects of covariates into the extended GFRR model. Let us consider a given number  $S$  of independent systems that are experiencing both perfect and imperfect maintenances according to a GFRR model. We still assume that, after  $K$  imperfect repairs, an AGAN (As Good As New) maintenance is performed. For a specific system indexed by  $s$ , we denote by

- $0 = T_0^s < T_1^s < T_2^s < \dots < T_k^s < \dots$ , the distinct failure times of system  $s$ ;
- $X_k^s = T_k^s - T_{k-1}^s, \forall k = 1, \dots, K$ , the interarrival failure times of system  $s$ ;
- $\lambda_{X_k^s}(\cdot)$ , the instantaneous hazard function of the interarrival time  $X_k^s$  for system  $s$ ;
- $Z_s = (Z_s^1, \dots, Z_s^p)^T$ , the vector of  $p$  time-fixed covariates of the system  $s$ . Here, the vector  $v^T$  represents the transpose vector of  $v$ .

Drawing inspiration from the Cox models (see Cox 1972), our model with covariates, referred as *EGFRR model with covariates*, assumes that the instantaneous hazard function of the interarrival time  $X_k$  of the system  $s$  is given by

$$\lambda_{X_k^s}(\cdot) = \rho^{a(k,\theta)} \lambda_{X_1}(\cdot) \exp(\beta^T Z_s), \quad (5.4)$$

where :

- $\rho \in \mathbb{R}_+^*$  is still a positive real number characterizing the maintenance effectiveness;
- for a fixed real number  $\theta$ , the function  $(k, \theta) \mapsto a(k, \theta)$  is a non decreasing function of the number of imperfect maintenance  $k$  such that  $a(1, \theta) = 0$ ;
- $\beta \in \mathbb{R}^p$  is the vector of regression coefficients;

—  $\lambda_{X_1}(t)$  represents the baseline instantaneous hazard at  $t$ .

In this model, the real parameter  $\theta$  is unknown and the baseline hazard function  $\lambda_{X_1}(\cdot)$ , which is also unknown, represents the hazard of a new system with all covariates having zero values (i.e,  $Z = 0_{\mathbb{R}^p}$ ). Furthermore, for all  $j = 1, \dots, p$ , the  $j$ -th component  $\beta^j$  of the unknown regression parameter  $\beta$  characterizes the effect of the covariate  $Z^j$  on the hazard function  $\lambda_{X_k^s}(\cdot)$ .

From Equation (5.4), one can see that

$$\Lambda_{X_k^s}(\cdot) = \Lambda_{X_1}(\cdot) \exp\left(\gamma a(k, \theta) + \beta^T Z_s\right), \quad (5.5)$$

where  $\gamma = \log(\rho)$  and  $\Lambda_{X_k^s}(\cdot)$  is the cumulative hazard function of  $X_k^s$ . It is recalled that  $\Lambda_{X_k^s}(t) = \int_0^t \lambda_{X_k^s}(s) ds$ .

It can be seen that the standard EGFR model is obtained when  $Z = 0_{\mathbb{R}^p}$  and  $a(k, \theta) = a(k)$ . As discussed in DAUXOIS et al. 2018, the shape of the function  $a(\cdot)$  must be chosen in a way that dissipates the explosion phenomenon. Specifically, it is mentioned that when  $a(k) = k - 1$ , the value of the parameter  $\rho$  has to be close to 1.

In our model, the parameter  $\theta$  provides more flexibility in the form of the function  $a(\cdot)$  in the shape of the intensity process. The choice of the function  $a(\cdot)$ , in conjunction with the inclusion of covariate effects via the parameter  $\beta$ , increases the ability of the EGFR model with covariates to address diverse scenarios and better modelizes the system reliability. Moreover, for a given system  $s$ , we have the relationship :

$$\frac{\Lambda_{X_{k+1}^s}(\cdot)}{\Lambda_{X_k^s}(\cdot)} = e^{\gamma(a(k+1, \theta) - a(k, \theta))}. \quad (5.6)$$

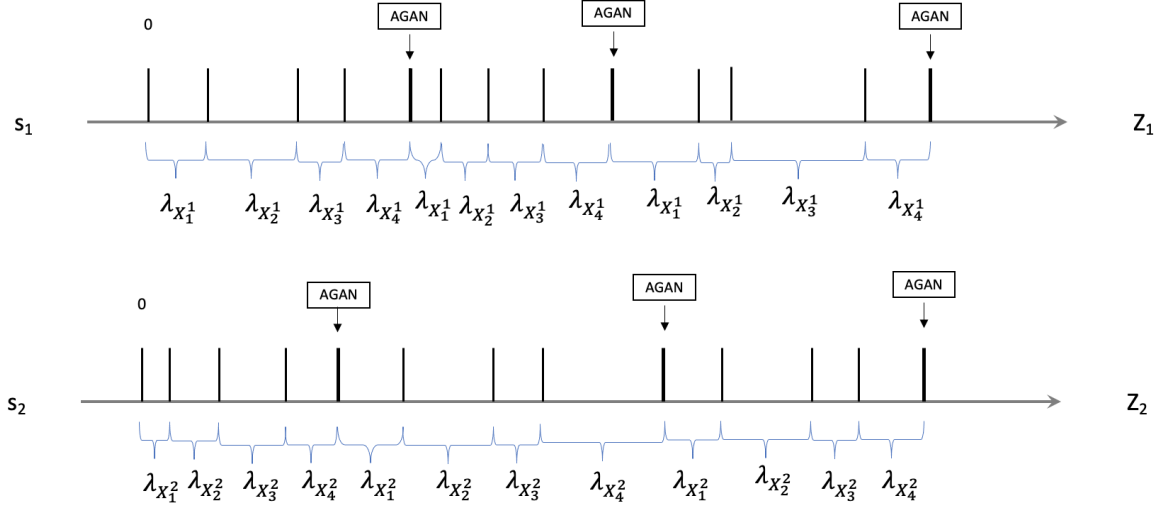
Equation (5.6) highlights that, when  $\gamma < 0$  (or  $\rho < 1$ ), the ratio  $\frac{\Lambda_{X_{k+1}^s}(\cdot)}{\Lambda_{X_k^s}(\cdot)}$  will be lower than 1, indicating that  $X_{k+1}^s$  will be stochastically greater than  $X_k^s$ . This suggests that the maintenance action improves the reliability of the system. On the other hand, in the case of imperfect repairs, we expect  $\frac{\Lambda_{X_{k+1}^s}(\cdot)}{\Lambda_{X_k^s}(\cdot)}$  to be greater than 1. This indicates that  $X_{k+1}^s$  is stochastically smaller than  $X_k^s$ , which leads to shortening of the interarrival times between failures. Then, we assume the  $\gamma$  parameter to be greater than 0.

It is worth noting that in the EGFR model with covariates, the regression coefficients  $\beta_j$  for all  $j = 1, \dots, p$  are interpretable, similar by what happens in the Cox model. This interpretability allows us to analyze the impact of each covariate on the hazard functions of the interarrival times  $X_k^s$  for different systems. For instance, consider two systems  $r$  and  $s$  such that  $Z_r^j = Z_s^j + 1$ , where  $Z^j$  represents the  $j$ -th covariate. If we assume that all other covariates  $Z^l$ , where  $l \neq j$ , are fixed and equal for both systems, then it can be shown that the hazard functions of the interarrival times  $X_k^s$  and  $X_k^r$  are proportional. The proportionality factor associated with the  $j$ -th covariate is equal to  $e^{\beta_j}$ . This allows us to quantify the impact of this covariate on the hazard function.

## 5.4 Estimation in the EGFR model with covariates

Our main objective is to estimate the Euclidean parameters of the EGFR model with covariates, which are the repair effect parameter  $\rho$ , the adjustment parameter  $\theta$  and the regression parameter  $\beta$ . To achieve this, we will follow a semiparametric approach as proposed by DAUXOIS et al. 2018.

Let us consider  $S$  independent systems, each of which undergoes a series of failure and maintenance actions. For a given system  $s$ , a perfect maintenance is carried out after  $K$  imperfect maintenances, these latter operated according to an EGFRR model with covariates. Let us consider that the number of perfect repairs operated on each system is fixed and equal  $M$ . This maintenance procedure is repeated for each system, resulting in  $M$  independent and identically distributed (i.i.d.) series of the  $K$  interfailure times for each system. In other words, for each system  $s$ , we observe  $M$  sequences of  $X_k^s$  for all  $k = 1, \dots, K$ . Figure 5.1 visually illustrates this situation for the specific case where  $S = 2$ ,  $K = 4$ , and  $M = 3$ . In this illustration, the interarrival times are represented by their hazard rate functions.



**Figure 5.1** – Example of alternation between perfect imperfect maintenances in the sequential maintenance policy with  $S = 2$ ,  $K = 4$  and  $M = 3$ .

Let us consider the order statistics  $X_{k,(1)}^s < X_{k,(2)}^s < \dots < X_{k,(m)}^s < \dots < X_{k,(M)}^s$  of the  $M$  observations of the  $k$ -th inter-failure time of the  $s$ -th system, denoted as  $X_k^s$ . The cumulative hazard function  $\Lambda_{X_k^s}(\cdot)$  of  $X_k^s$  can be estimated using the Nelson-Aalen estimator  $\hat{\Lambda}_{X_k^s}(\cdot)$  (see e.g. ANDERSEN, BORGAN et al. 1993), in the case where no censoring occurs. More precisely, for all  $k = 1, \dots, K$  and  $s = 1, \dots, S$ :

$$\hat{\Lambda}_{X_k^s}(t) = \sum_{m=1}^M \frac{1}{M - m + 1} I\left(X_{k,(m)}^s \leq t\right),$$

where  $I(\cdot)$  is the indicator function, which equals 1 if the condition inside the parentheses is true and 0 otherwise.

#### 5.4.1 Estimation of the regression parameter $\beta$

Using equation (5.5), for two distinct systems  $r$  and  $s$  and for all  $t > 0$ , we can derive the following equation for all  $k = 1, \dots, K$ :

$$\log\left(\frac{\Lambda_{X_k^s}(t)}{\Lambda_{X_k^r}(t)}\right) = \beta^T (Z_s - Z_r).$$

If  $\tau_1$  and  $\tau_2$  are two positive numbers, with  $\tau_1$  sufficiently small and  $\tau_2$  sufficiently high, by

integrating the above relation and summing over all systems and failures, we obtain :

$$\sum_{r,s=1}^S \sum_{k=1}^K \int_{\tau_1}^{\tau_2} \left( \log \left( \frac{\Lambda_{X_k^s}(t)}{\hat{\Lambda}_{X_k^r}(t)} \right) - \beta^T (Z_s - Z_r) \right)^2 dt = 0.$$

Drawing inspiration from the the approach used in DAUXOIS et al. 2018, an estimator of  $\beta$  can be

$$\hat{\beta} = \operatorname{argmin}_{\beta \in \mathbb{R}^p} \sum_{r,s=1}^S \sum_{k=1}^K \int_{\tau_1}^{\tau_2} \left( \log \left( \frac{\hat{\Lambda}_{X_k^s}(t)}{\hat{\Lambda}_{X_k^r}(t)} \right) - \beta^T (Z_s - Z_r) \right)^2 dt. \quad (5.7)$$

In order to derive an explicit expression of this estimator, let the function  $F$  be defined by

$$F_{\hat{f}}(\beta) = \sum_{r,s=1}^S \sum_{k=1}^K \int_{\tau_1}^{\tau_2} \left( \log \left( \frac{\hat{\Lambda}_{X_k^s}(t)}{\hat{\Lambda}_{X_k^r}(t)} \right) - \beta^T (Z_s - Z_r) \right)^2 dt,$$

for all  $\beta \in \mathbb{R}^p$ . We can easily show that  $F_{\hat{f}}$  is a  $C^\infty$  and convex function of  $\mathbb{R}^p$ . If we denote by  $\nabla F_{\hat{f}}(\beta) = \left( \frac{\partial F_{\hat{f}}(\beta)}{\partial \beta_j} \right)_{1 \leq j \leq p}$  the gradient of  $F_{\hat{f}}$ , then we have  $\nabla F_{\hat{f}}(\hat{\beta}) = 0_{\mathbb{R}^p}$ . Moreover,

$$\begin{aligned} \nabla F_{\hat{f}}(\hat{\beta}) = 0_{\mathbb{R}^p} &\Leftrightarrow -2 \sum_{r,s=1}^S \sum_{k=1}^K (Z_s - Z_r) \int_{\tau_1}^{\tau_2} \left( \log \left( \frac{\hat{\Lambda}_{X_k^s}(t)}{\hat{\Lambda}_{X_k^r}(t)} \right) - \hat{\beta}^T (Z_s - Z_r) \right) dt = 0_{\mathbb{R}^p} \Leftrightarrow \\ &\sum_{r,s=1}^S \sum_{k=1}^K (Z_s - Z_r) \int_{\tau_1}^{\tau_2} \left( \log \frac{\hat{\Lambda}_{X_k^s}(t)}{\hat{\Lambda}_{X_k^r}(t)} \right) dt = \sum_{r,s=1}^S \sum_{k=1}^K (\tau_2 - \tau_1) (Z_s - Z_r) \hat{\beta}^T (Z_s - Z_r). \end{aligned}$$

Since  $\hat{\beta}^T (Z_s - Z_r) \in \mathbb{R}$ , then  $\hat{\beta}^T (Z_s - Z_r) = (Z_s - Z_r)^T \hat{\beta}$ . Integrating this in the above relationship, we have :

$$\sum_{r,s=1}^S \sum_{k=1}^K (Z_s - Z_r) \int_{\tau_1}^{\tau_2} \left( \log \frac{\hat{\Lambda}_{X_k^s}(t)}{\hat{\Lambda}_{X_k^r}(t)} \right) dt = \sum_{r,s=1}^S \sum_{k=1}^K (\tau_2 - \tau_1) (Z_s - Z_r) (Z_s - Z_r)^T \beta.$$

Let us consider the matrix  $D$  of  $\mathcal{M}_p(\mathbb{R})$  such that  $D = \sum_{r,s=1}^S (Z_s - Z_r)(Z_s - Z_r)^T$ . Under the assumption of  $D$  being invertible, we deduce that :

$$\hat{\beta} = \frac{D^{-1}}{(\tau_2 - \tau_1)K} \sum_{r,s=1}^S \sum_{k=1}^K (Z_s - Z_r) \int_{\tau_1}^{\tau_2} \left( \log \frac{\hat{\Lambda}_{X_k^s}(t)}{\hat{\Lambda}_{X_k^r}(t)} \right) dt, \quad (5.8)$$

where  $D^{-1}$  is the inverse matrix of  $D$ .

In the standard regression model, a condition that ensures the invertibility of the matrix  $D$  is that the variables  $Z_s^1, \dots, Z_s^p$ , for all  $s = 1, \dots, S$ , must be linearly independent. Therefore, we consider the hypothesis ( $\mathbf{H}_1$ ) below, which ensures the invertibility of  $D$  :

$\mathbf{H}_1$  : For all couple of non identical elements  $r$  and  $s$  in  $\{1, \dots, S\}$  the variables  $Z_r^1 - Z_s^1, \dots, Z_r^p - Z_s^p$  are linearly independent.

#### 5.4.2 Estimation of the parameters $\gamma$ and $\theta$

Now, for all failure  $k = 1, \dots, K - 1$ , using again (5.5), we have

$$\log \left( \frac{\Lambda_{X_{k+1}^s}(t)}{\Lambda_{X_k^s}(t)} \right) = \gamma (a(k+1, \theta) - a(k, \theta)).$$

Thus, with  $\tau_1$  and  $\tau_2$  being two positive real numbers as defined above, we have

$$\int_{\tau_1}^{\tau_2} \left( \log \left( \frac{\Lambda_{X_{k+1}^s}(t)}{\Lambda_{X_k^s}(t)} \right) - \gamma (a(k+1, \theta) - a(k, \theta)) \right)^2 dt = 0.$$

Finally, summing over the number of failures and systems, we obtain :

$$\sum_{s=1}^S \sum_{k=1}^{K-1} \int_{\tau_1}^{\tau_2} \left( \log \left( \frac{\Lambda_{X_{k+1}^s}(t)}{\Lambda_{X_k^s}(t)} \right) - \gamma (a(k+1, \theta) - a(k, \theta)) \right)^2 dt = 0.$$

Thus, we can estimate both  $\gamma$  and  $\theta$  by  $\hat{\gamma}$  and  $\hat{\theta}$  defined as follows :

$$(\hat{\gamma}, \hat{\theta}) = \underset{(\gamma, \theta) \in \mathbb{R} \times \mathbb{R}_+}{\operatorname{argmin}} \sum_{s=1}^S \sum_{k=1}^{K-1} \int_{\tau_1}^{\tau_2} \left( \log \left( \frac{\hat{\Lambda}_{X_{k+1}^s}(t)}{\hat{\Lambda}_{X_k^s}(t)} \right) - \gamma (a(k+1, \theta) - a(k, \theta)) \right)^2 dt. \quad (5.9)$$

Unlike  $\hat{\beta}$ , the expressions of  $\hat{\gamma}$  and  $\hat{\theta}$  are not easy to specify. However, they can be obtained using iterative numerical methods such as Newton-Raphson. Moreover, it is worth noting that in the special case where  $a(k, \theta) = a(k)$ , the model parameters  $\beta$  and  $\gamma$  can be estimated separately. The estimation of  $\beta$  can be performed using Equation (5.8), while an estimator for  $\gamma$  can be obtained following the approach proposed in DAUXOIS et al. 2018.

## 5.5 Large sample behaviour of estimators

Throughout the rest of this paper, we adopt the following notations :

- $I_k^s = \int_{\tau_1}^{\tau_2} \left( \log \Lambda_{X_{k+1}^s}(t) - \log \Lambda_{X_k^s}(t) \right) dt$  and  $\hat{I}_k^s = \int_{\tau_1}^{\tau_2} \left( \log \hat{\Lambda}_{X_{k+1}^s}(t) - \log \hat{\Lambda}_{X_k^s}(t) \right) dt$ ;
- $I = (I_1^1, \dots, I_K^1, \dots, I_1^S, \dots, I_K^S)$  and  $\hat{I} = (\hat{I}_1^1, \dots, \hat{I}_K^1, \dots, \hat{I}_1^S, \dots, \hat{I}_K^S)$ ;
- $J_I(\gamma, \theta) = \sum_{s=1}^S \sum_{k=1}^{K-1} \int_{\tau_1}^{\tau_2} \left( \log \left( \frac{\Lambda_{X_{k+1}^s}(t)}{\Lambda_{X_k^s}(t)} \right) - \gamma (a(k+1, \theta) - a(k, \theta)) \right)^2 dt$ ;
- $J_{\hat{I}}(\gamma, \theta) = \sum_{s=1}^S \sum_{k=1}^{K-1} \int_{\tau_1}^{\tau_2} \left( \log \left( \frac{\hat{\Lambda}_{X_{k+1}^s}(t)}{\hat{\Lambda}_{X_k^s}(t)} \right) - \gamma (a(k+1, \theta) - a(k, \theta)) \right)^2 dt$ ;
- $\nabla J_I(\gamma, \theta)$  (resp.  $\nabla J_{\hat{I}}(\gamma, \theta)$ ) is the gradient function of  $J_I(\cdot, \cdot)$  (resp. the gradient function of  $J_{\hat{I}}(\cdot, \cdot)$ );
- $H_I(\gamma, \theta)$  (resp.  $H_{\hat{I}}(\gamma, \theta)$ ) is the Hessian matrix the function  $J_I(\cdot, \cdot)$  (resp. the Hessian matrix of  $J_{\hat{I}}(\cdot, \cdot)$ );
- $(\gamma_0, \theta_0) = \underset{(\gamma, \theta) \in \mathbb{R}^2}{\operatorname{argmin}} J_I(\gamma, \theta)$ ;
- $(\hat{\gamma}, \hat{\theta}) = \underset{(\gamma, \theta) \in \mathbb{R}^2}{\operatorname{argmin}} J_{\hat{I}}(\gamma, \theta)$ ;
- $F_I(\beta) = \sum_{r,s=1}^S \sum_{k=1}^K \int_{\tau_1}^{\tau_2} \left( \log \left( \frac{\Lambda_{X_k^s}(t)}{\Lambda_{X_k^r}(t)} \right) - \beta^T (Z_s - Z_r) \right)^2 dt$ ;
- $F_{\hat{I}}(\beta) = \sum_{r,s=1}^S \sum_{k=1}^K \int_{\tau_1}^{\tau_2} \left( \log \left( \frac{\hat{\Lambda}_{X_k^s}(t)}{\hat{\Lambda}_{X_k^r}(t)} \right) - \beta^T (Z_s - Z_r) \right)^2 dt$ ;
- $\beta_0 = \underset{\beta \in \mathbb{R}^p}{\operatorname{argmin}} F_I(\beta)$ ;
- $\hat{\beta} = \underset{\beta \in \mathbb{R}^p}{\operatorname{argmin}} F_{\hat{I}}(\beta)$ ;
- $d_k(\theta) = a(k+1, \theta) - a(k, \theta)$
- $\dot{d}_k(\theta) = \frac{\partial a(k+1, \theta)}{\partial \theta}(\theta) - \frac{\partial a(k, \theta)}{\partial \theta}(\theta)$

We also assume that the following hypotheses are verified.



- $\mathbf{H}_2$  : For all fixed value of  $v$ , the function  $(\gamma, \theta) \mapsto J_v(\gamma, \theta)$  is of  $C^3$  class functions.
- $\mathbf{H}_3$  : For all fixed value of  $v$ , the function  $(\gamma, \theta) \mapsto J_v(\gamma, \theta)$  is a strictly convex function with respect to both  $\gamma$  and  $\theta$ .
- $\mathbf{H}_4$  : The function  $\forall \theta \mapsto d_k(\theta)$  is a  $C^1$  class function.

### 5.5.1 Asymptotic behaviour of $\hat{\Lambda}_{X_k^s}(\cdot)$

The properties of the Nelson-Aalen estimator have been the subject of numerous work. For example, ANDERSEN, BORGAN et al. 1993 provides the asymptotic properties of this estimator. Building on the results presented in DAUXOIS et al. 2018, we adapt these results to our specific case and recall them in the Theorem 5.1.

**Theorem 5.1** *In the Skorohod space  $D[\tau_1, \tau_2]$  of cadlag functions, we have for all  $s = 1, \dots, S$  and  $k = 1, \dots, K$ , the weak convergence :*

$$\sqrt{M} \left( \hat{\Lambda}_{X_k^s}(\cdot) - \Lambda_{X_k^s}(\cdot) \right) \xrightarrow{\mathcal{D}} U_{X_k^s}(\cdot), \text{ when } M \rightarrow +\infty, \quad (5.10)$$

where, for all  $t > 0$ , the quantity  $U_{X_k^s}(t)$  is a mean zero Gaussian martingale with  $U_{X_k^s}(0) = 0$  and covariance function

$$\text{Cov} \left( U_{X_k^s}(u), U_{X_k^s}(t) \right) = \sigma_k^s(u \wedge t) = \left( \frac{1}{1 - F_{X_1}(u \wedge t)} \right)^{\exp(\beta^T Z_s + \gamma a(k, \theta))}.$$

### 5.5.2 Asymptotic behaviour of $\hat{\beta}$

It is easy to show that

$$\beta_0 = \frac{D^{-1}}{(\tau_2 - \tau_1)K} \sum_{r,s=1}^S \sum_{k=1}^K (Z_s - Z_r) \int_{\tau_1}^{\tau_2} \left( \log \frac{\Lambda_{X_k^s}(t)}{\Lambda_{X_k^r}(t)} \right) dt.$$

The following Theorem 5.2 gives the asymptotic distribution of  $\hat{\beta}$ .

**Theorem 5.2** *When  $M \rightarrow +\infty$ , we have*

$$\sqrt{M} (\hat{\beta} - \beta_0) \xrightarrow{\mathcal{D}} \mathcal{Z} = \frac{D^{-1}}{(\tau_2 - \tau_1)K} \sum_{r,s=1}^S \sum_{k=1}^K (Z_s - Z_r) \int_{\tau_1}^{\tau_2} \left( \frac{U_{X_k^s}(t)}{\Lambda_{X_k^s}(t)} - \frac{U_{X_k^r}(t)}{\Lambda_{X_k^r}(t)} \right) dt. \quad (5.11)$$

The limit process  $\mathcal{Z}$  is a mean zero Gaussian random variable.

Before proving this result, let us consider the following proposition which allows us to establish the asymptotic behavior of the estimators of the euclidean parameters.

**Proposition 5.3** *For all  $s = 1, \dots, S$  and for all  $k = 1, \dots, K$ , when  $M \rightarrow +\infty$ , we have the following results :*

1.

$$\sqrt{M} \left( \int_{\tau_1}^{\tau_2} \log \hat{\Lambda}_{X_k^s}(t) dt - \int_{\tau_1}^{\tau_2} \log \Lambda_{X_k^s}(t) dt \right) \xrightarrow{\mathcal{D}} \int_{\tau_1}^{\tau_2} \frac{U_{X_k^s}(t)}{\Lambda_{X_k^s}(t)} dt, \quad (5.12)$$

where the  $U_{X_k^s}$  is defined as in Theorem 5.1.

2.

$$\hat{I}_k^s \xrightarrow{\mathbb{P}} I_k^s, \quad (5.13)$$

where  $\hat{I}_k^s$  are defined as in Section 5.5.

**Proof** For all  $s = 1, \dots, S$  and  $k = 1, \dots, K$ , and when  $M \rightarrow +\infty$ , we know that (s.e.g page 192 of ANDERSEN, BORGAN et al. 1993)

$$\sup_{t \in [\tau_1, \tau_2]} |\hat{\Lambda}_{X_k^s}(t) - \Lambda_{X_k^s}(t)| \xrightarrow{as} 0.$$

Moreover, it is shown in DAUXOIS et al. 2018 that

$$\sqrt{M} \left( \log \hat{\Lambda}_{X_k^s}(\cdot) - \log \Lambda_{X_k^s}(\cdot) \right) \xrightarrow{\mathcal{D}} \frac{U_{X_k^s}(\cdot)}{\Lambda_{X_k^s}(\cdot)}, \text{ when } M \rightarrow +\infty. \quad (5.14)$$

Let  $f$  be the function defined on the Skorohod space  $D[\tau_1, \tau_2]$  by  $f(y) = \int_{\tau_1}^{\tau_2} y(t) dt$ . Since  $f$  is a linear function, its differential (in the sense of Fréchet) is equal to itself. Therefore, applying functional Delta-method (s.e.g Theorem 20.8 of der VAART 1998), for all  $s = 1, \dots, S$  and  $k = 1, \dots, K$ , we obtain :

$$\sqrt{M} \left( f \left( \log \hat{\Lambda}_{X_k^s}(\cdot) \right) - f \left( \log \Lambda_{X_k^s}(\cdot) \right) \right) \xrightarrow{\mathcal{D}} f \left( \frac{U_{X_k^s}(\cdot)}{\Lambda_{X_k^s}(\cdot)} \right), \text{ when } M \rightarrow +\infty.$$

In other words :

$$\sqrt{M} \left( \int_{\tau_1}^{\tau_2} \log \hat{\Lambda}_{X_k^s}(t) dt - \int_{\tau_1}^{\tau_2} \log \Lambda_{X_k^s}(t) dt \right) \xrightarrow{\mathcal{D}} \int_{\tau_1}^{\tau_2} \frac{U_{X_k^s}(t)}{\Lambda_{X_k^s}(t)} dt, \text{ when } M \rightarrow +\infty.$$

That proves the first result of the proposition (i.e relation (5.12)).

Since the random variables  $\int_{\tau_1}^{\tau_2} (\log \hat{\Lambda}_{X_{k+1}^s}(t) - \log \Lambda_{X_{k+1}^s}(t)) dt$  and  $\int_{\tau_1}^{\tau_2} (\log \hat{\Lambda}_{X_k^s}(t) - \log \Lambda_{X_k^s}(t)) dt$  are independent, we obtain

$$\sqrt{M} \left( \begin{array}{l} \int_{\tau_1}^{\tau_2} (\log \hat{\Lambda}_{X_k^s}(t) - \log \Lambda_{X_k^s}(t)) dt \\ \int_{\tau_1}^{\tau_2} (\log \hat{\Lambda}_{X_{k+1}^s}(t) - \log \Lambda_{X_{k+1}^s}(t)) dt \end{array} \right) \xrightarrow{\mathcal{D}} \left( \begin{array}{l} \int_{\tau_1}^{\tau_2} \frac{U_{X_k^s}(t)}{\Lambda_{X_k^s}(t)} dt \\ \int_{\tau_1}^{\tau_2} \frac{U_{X_{k+1}^s}(t)}{\Lambda_{X_{k+1}^s}(t)} dt \end{array} \right), \text{ when } M \rightarrow +\infty$$

Moreover, for all  $s = 1, \dots, S$  and  $k = 1, \dots, K - 1$ ,

$$\sqrt{M} (\hat{I}_k^s - I_k^s) = \sqrt{M} \left( \int_{\tau_1}^{\tau_2} (\log \hat{\Lambda}_{X_{k+1}^s}(t) - \log \Lambda_{X_{k+1}^s}(t)) dt - \int_{\tau_1}^{\tau_2} (\log \hat{\Lambda}_{X_k^s}(t) - \log \Lambda_{X_k^s}(t)) dt \right).$$

Using a continuous mapping transformation theorem, we obtain, when  $M \rightarrow +\infty$  and for all  $s = 1, \dots, S$  and  $k = 1, \dots, K - 1$  that

$$\sqrt{M} (\hat{I}_k^s - I_k^s) \xrightarrow{\mathcal{D}} \int_{\tau_1}^{\tau_2} \frac{U_{X_{k+1}^s}(t)}{\Lambda_{X_{k+1}^s}(t)} dt - \int_{\tau_1}^{\tau_2} \frac{U_{X_k^s}(t)}{\Lambda_{X_k^s}(t)} dt.$$

From this, we can deduce that

$$\sqrt{M} (\hat{I}_k^s - I_k^s) = O_p(1) \Rightarrow \hat{I}_k^s - I_k^s = o_p(1).$$

In other words, when  $M \rightarrow +\infty$ , we obtain the probability convergence for all  $s = 1, \dots, S$  and  $k = 1, \dots, K$  :

$$\hat{I}_k^s \xrightarrow{\mathbb{P}} I_k^s. \quad \square$$

Let us now prove Theorem 5.2.

**Proof** Referring to the expressions of  $\beta_0$  and  $\hat{\beta}$  provided in Section 5.5, we find that :

$$\begin{aligned}\hat{\beta} - \beta_0 &= \frac{D^{-1}}{(\tau_2 - \tau_1)K} \sum_{r,s=1}^S \sum_{k=1}^K (Z_s - Z_r) \int_{\tau_1}^{\tau_2} \log \left( \frac{\hat{\Lambda}_{X_k^s}(t)}{\hat{\Lambda}_{X_k^r}(t)} - \frac{\Lambda_{X_k^s}(t)}{\Lambda_{X_k^r}(t)} \right) dt \\ &= \frac{D^{-1}}{(\tau_2 - \tau_1)K} \sum_{r,s=1}^S \sum_{k=1}^K (Z_s - Z_r) \left( \int_{\tau_1}^{\tau_2} \left( \log \hat{\Lambda}_{X_k^s}(t) - \log \Lambda_{X_k^s}(t) \right) dt - \int_{\tau_1}^{\tau_2} \left( \log \hat{\Lambda}_{X_k^r}(t) - \log \Lambda_{X_k^r}(t) \right) dt \right)\end{aligned}$$

For all given  $Z = (Z_s)_{1 \leq s \leq S}$ , let the linear function  $g_Z$  defined on  $\mathcal{D}^{S \times K}[\tau_1, \tau_2]$  by

$$g_Z(y) = \frac{D^{-1}}{(\tau_2 - \tau_1)K} \sum_{r,s=1}^S \sum_{k=1}^K (Z_s - Z_r) \int_{\tau_1}^{\tau_2} (y_k^s(t) - y_k^r(t)) dt,$$

for all  $y = (y_1^1, \dots, y_K^1, \dots, y_1^S, \dots, y_K^S)$ .

Using the function  $g_Z(\cdot)$ , we can show that

$$\begin{aligned}\sqrt{M}(\hat{\beta} - \beta_0) &= \sqrt{M} \frac{D^{-1}}{(\tau_2 - \tau_1)K} \sum_{r,s=1}^S \sum_{k=1}^K (Z_s - Z_r) \int_{\tau_1}^{\tau_2} \log \left( \frac{\hat{\Lambda}_{X_k^s}(t)}{\hat{\Lambda}_{X_k^r}(t)} - \frac{\Lambda_{X_k^s}(t)}{\Lambda_{X_k^r}(t)} \right) dt \\ &= \sqrt{M} (g_Z(\hat{L}_X(\cdot)) - g_Z(L_X(\cdot))),\end{aligned}$$

where  $\hat{L}_X(\cdot) = (\log \hat{\Lambda}_{X_1^1}(\cdot), \dots, \log \hat{\Lambda}_{X_1^K}(\cdot), \dots, \log \hat{\Lambda}_{X_S^1}(\cdot), \dots, \log \hat{\Lambda}_{X_S^K}(\cdot))$  and

$L_X(\cdot) = (\log \Lambda_{X_1^1}(\cdot), \dots, \log \Lambda_{X_1^K}(\cdot), \dots, \log \Lambda_{X_S^1}(\cdot), \dots, \log \Lambda_{X_S^K}(\cdot))$ .

Since  $g_Z$  is linear and regarding the convergence given in Equation (5.14), the application of the functional delta-method leads to the following convergence when  $M \rightarrow +\infty$  :

$$\sqrt{M}(\hat{\beta} - \beta_0) \xrightarrow{\mathcal{D}} \mathcal{Z} = \frac{D^{-1}}{(\tau_2 - \tau_1)K} \sum_{r,s=1}^S \sum_{k=1}^K (Z_s - Z_r) \int_{\tau_1}^{\tau_2} \left( \frac{U_{X_k^s}(t)}{\Lambda_{X_k^s}(t)} - \frac{U_{X_k^r}(t)}{\Lambda_{X_k^r}(t)} \right) dt.$$

The limit variable  $\mathcal{Z}$  is a mean zero multivariate Gaussian variable.  $\square$

### 5.5.3 Asymptotic behaviour of $\hat{\zeta} = (\hat{\gamma}, \hat{\theta})$

Let  $\zeta_0 = (\gamma_0, \theta_0)$ . The weak convergence of our estimator, as defined in (5.9), depends on its consistency, which is ensured by the following theorem.

**Theorem 5.4** *Under assumptions  $\mathbf{H}_2$  and  $\mathbf{H}_3$ ,*

$$\hat{\zeta} \xrightarrow{\mathbb{P}} \zeta_0, \text{ when } M \rightarrow +\infty.$$

This result is obtained from the Lemma 5.10 of der VAART 1998 which is recalled below with some modifications (adapted to our case).

**Proposition 5.5** *(Adapted version of Lemma 5.10 of der VAART 1998) Let  $\mathbf{E}$  a subset of parameters. Let  $\varphi_n$  and  $\varphi$  two functions of  $\alpha \in \mathbf{E}$  such that  $\varphi_n$  is a random function and  $\varphi$  a fixed function. Assume that :*

- $\alpha \mapsto \varphi_n(\alpha)$  is continuous and has exactly one zero  $\hat{\alpha}_n$
- $\varphi_n(\alpha)$  tends to  $\varphi(\alpha)$  in probability for all  $\alpha \in \mathbf{E}$
- $\varphi$  has  $\alpha_0$  as a zero.

Then, as  $n \rightarrow +\infty$ ,

$$\hat{\alpha}_n \xrightarrow{\mathbb{P}} \alpha_0.$$

**Proof** Let us mention that the convergence given in the Equation (5.13) leads to some important results. Let denote

$\nabla J_I^{(1)}(\gamma, \theta) = \frac{\partial J_I(\gamma, \theta)}{\partial \gamma}$  and  $\nabla J_I^{(2)}(\gamma, \theta) = \frac{\partial J_I(\gamma, \theta)}{\partial \theta}$ , where  $J_I(\gamma, \theta)$  is defined as in Section 5.5. Then, we have

$$\begin{aligned} \nabla J_I^{(1)}(\gamma, \theta) &= -2 \sum_{s=1}^S \sum_{k=1}^{K-1} (a(k+1, \theta) - a(k, \theta)) \\ &\quad \times \int_{\tau_1}^{\tau_2} \left( \log \frac{\Lambda_{X_{k+1}^s}(t)}{\Lambda_{X_k^s}(t)} - \gamma (a(k+1, \theta) - a(k, \theta)) \right) dt \\ &= -2 \sum_{s=1}^S \sum_{k=1}^{K-1} (a(k+1, \theta) - a(k, \theta)) \\ &\quad \times (I_k^s - \gamma (a(k+1, \theta) - a(k, \theta))) (\tau_2 - \tau_1) \\ &= -2 \sum_{s=1}^S \sum_{k=1}^{K-1} d_k(\theta) (I_k^s - \gamma \tau d_k(\theta)), \end{aligned}$$

and

$$\begin{aligned} \nabla J_I^{(2)}(\gamma, \theta) &= -2\gamma \sum_{s=1}^S \sum_{k=1}^{K-1} \left( \frac{\partial a(k+1, \theta)}{\partial \theta}(\theta) - \frac{\partial a(k, \theta)}{\partial \theta}(\theta) \right) \\ &\quad \times \int_{\tau_1}^{\tau_2} \left( \log \frac{\Lambda_{X_{k+1}^s}(t)}{\Lambda_{X_k^s}(t)} - \gamma (a(k+1, \theta) - a(k, \theta)) \right) dt \\ &= -2\gamma \sum_{s=1}^S \sum_{k=1}^{K-1} \left( \frac{\partial a(k+1, \theta)}{\partial \theta}(\theta) - \frac{\partial a(k, \theta)}{\partial \theta}(\theta) \right) \\ &\quad \times (I_k^s - \gamma (a(k+1, \theta) - a(k, \theta))) (\tau_2 - \tau_1) \\ &= -2\gamma \sum_{s=1}^S \sum_{k=1}^{K-1} \dot{d}(k, \theta) (I_k^s - \gamma \tau d_k(\theta)), \end{aligned}$$

where  $\tau = \tau_2 - \tau_1$  and the functions  $d(\cdot)$  and  $\dot{d}(\cdot)$  are as defined in Section 5.5. The function defined on  $\mathbb{R}^{(K-1) \times S}$  by  $\forall x = (x_1^1, \dots, x_{K-1}^1, \dots, x_1^S, \dots, x_{K-1}^S) \mapsto -2 \sum_{s=1}^S \sum_{k=1}^{K-1} d_k(\theta) (x_k^s - \gamma \tau d_k(\theta))$  and  $\forall x = (x_1^1, \dots, x_{K-1}^1, \dots, x_1^S, \dots, x_{K-1}^S) \mapsto -2\gamma \sum_{s=1}^S \sum_{k=1}^{K-1} \dot{d}(k, \theta) (x_k^s - \gamma \tau d_k(\theta))$  are continuous. Considering the probability convergence provided by Equation (5.13) and applying the continuous mapping theorem (cf. e.g., Theorem 2.3 on page 7 of der VAART 1998) to the expressions for  $\nabla J_I^{(1)} \hat{I}(\gamma, \theta)$  and  $\nabla J_I^{(2)} \hat{I}(\gamma, \theta)$  yields results as  $M \rightarrow +\infty$ .

$$\nabla J_I^{(1)}(\gamma, \theta) \xrightarrow{\mathbb{P}} \nabla J_I^{(1)}(\gamma, \theta)$$

and

$$\nabla J_I^{(2)}(\gamma, \theta) \xrightarrow{\mathbb{P}} \nabla J_I^{(2)}(\gamma, \theta).$$

Thus we obtain we following convergence

$$\nabla J_I(\gamma, \theta) \xrightarrow{\mathbb{P}} \nabla J_I(\gamma, \theta). \quad (5.15)$$

Using a similar method, one can show that

$$H_I(\gamma, \theta) \xrightarrow{\mathbb{P}} H_I(\gamma, \theta), \text{ when } M \rightarrow +\infty. \quad (5.16)$$

In order to apply the result of the Proposition 5.5 to our case, let us introduce the following notations :

- $\varphi_n(\cdot) = \nabla J_I(\cdot)$ ,
- $\varphi(\cdot) = \nabla J_I(\cdot)$ ,
- $\alpha_n = \hat{\zeta}$ ,
- $\alpha_0 = \zeta_0$ ,
- $n = M$ .

□

The assumption  $\mathbf{H}_3$  implies that  $\hat{\zeta}$  is the unique root of the function  $\nabla J_I(\cdot)$  and the function  $\nabla J_I(\cdot)$  has  $\zeta_0$  as unique zero. Moreover the convergence (5.15) the assumption  $\nabla J_I(\zeta) \xrightarrow{\mathbb{P}} \nabla J_I(\zeta)$ . Thus, the result on the theorem is obtained. In other words, when  $M \rightarrow +\infty$ ,

$$\hat{\zeta} \xrightarrow{\mathbb{P}} \zeta_0.$$

Regarding asymptotic behaviour of the estimator  $\hat{\zeta}$ , its is established by the Theorem 5.6.

**Theorem 5.6** *Under assumptions  $\mathbf{H}_2$  and  $\mathbf{H}_3$  and when  $M \rightarrow +\infty$ , we have the weak convergence*

$$\sqrt{M} (\hat{\zeta} - \zeta_0) \xrightarrow{\mathcal{D}} H_I^{-1}(\zeta_0) \Psi,$$

where

$$\Psi = \begin{pmatrix} -2 \sum_{k=1}^S \sum_{k=1}^{K-1} d_k(\theta_0) \int_{\tau_1}^{\tau_2} \left( \frac{U_{X_{k+1}^S}(t)}{\Lambda_{X_{k+1}^S}(t)} - \frac{U_{X_k^S}(t)}{\Lambda_{X_k^S}(t)} \right) dt \\ -2\gamma_0 \sum_{k=1}^S \sum_{k=1}^{K-1} \dot{d}_k(\theta_0) \int_{\tau_1}^{\tau_2} \left( \frac{U_{X_{k+1}^S}(t)}{\Lambda_{X_{k+1}^S}(t)} - \frac{U_{X_k^S}(t)}{\Lambda_{X_k^S}(t)} \right) dt \end{pmatrix}$$

and  $H_I^{-1}(\zeta_0)$  is the inverse of the Hessian matrix  $H_I(\zeta_0) = H_I(\gamma_0, \theta_0)$ .

Theorem 2.11 of KOSOROK 2008 is required to prove our result. This latter, is recalled in the following proposition.

**Proposition 5.7** *Let  $\mathbf{E}$  be a normed space,  $\varphi_n$  and  $\varphi$  two functions of  $\zeta \in \mathbf{E}$  such that  $\varphi_n$  is a random function and  $\varphi$  a fixed function. It is assumed that :*

1.  $\{\hat{\zeta}_n\}$  is a random sequence from  $\mathbf{E}$  such that  $\sqrt{n} \varphi_n(\hat{\zeta}_n) \xrightarrow{\mathbb{P}} 0$ , when  $n \rightarrow +\infty$ ;
2.  $\varphi$  has  $\zeta_0$  as a root in the interior of  $\mathbf{E}$ ;
3.  $\|\hat{\zeta}_n - \zeta_0\| \xrightarrow{\mathbb{P}} 0$ , when  $n \rightarrow +\infty$ ;
4.  $\sqrt{n} (\varphi_n(\zeta_0) - \varphi(\zeta_0)) \xrightarrow{\mathcal{D}} \mathcal{L}$ , when  $n \rightarrow +\infty$ , where  $\mathcal{L}$  is an known random variable;
- 5.

$$\frac{\|\sqrt{n}[\varphi_n(\hat{\zeta}_n) - \varphi(\hat{\zeta}_n)] - \sqrt{n}[\varphi_n(\zeta_0) - \varphi(\zeta_0)]\|_2}{1 + \sqrt{n}\|\hat{\zeta}_n - \zeta_0\|} \xrightarrow{\mathbb{P}} 0, \text{ when } n \rightarrow +\infty.$$

If the function  $\zeta \rightarrow \varphi(\zeta)$  is Fréchet-differentiable such as the inverse of its differential at  $\zeta_0$  (denoted by  $\hat{\varphi}_{\zeta_0}$ ) exist, then when  $n \rightarrow +\infty$ ,

$$\|\sqrt{n}\hat{\varphi}_{\zeta_0}(\hat{\zeta}_n - \zeta_0) + \sqrt{n}(\varphi_n(\zeta_0) - \varphi(\zeta_0))\|_2 \xrightarrow{\mathbb{P}} 0$$

and  $\sqrt{n}(\hat{\zeta}_n - \zeta_0)$  converges in distribution to  $-\hat{\varphi}_{\zeta_0}^{-1}\mathcal{L}$ .

**Proof** As above, asymptotic analysis is conducted respect to  $M$ , the functions  $\varphi_M(\cdot) = \nabla J_{\hat{I}}(\cdot)$ ,  $\varphi(\cdot) = \nabla J_I(\cdot)$  and the estimator  $\hat{\zeta}_M$  is denoted  $\hat{\zeta}$ . As  $\hat{\zeta}$  is the root of the  $\nabla J_{\hat{I}}(\cdot)$ , for all  $M > 0$ , we have  $\sqrt{M}\nabla J_{\hat{I}}(\hat{\zeta}) = 0$  almost surely. That insures the first condition. Also, since  $\zeta_0$  is the root of  $\nabla J_I(\cdot)$ , from the assumption the  $\mathbf{H}_3$ , the second condition is satisfied. Moreover, an application of Theorem 5.4, gives  $\|\hat{\zeta} - \zeta_0\| \xrightarrow{\mathbb{P}} 0$ .

In addition, the assumptions  $\mathbf{H}_2$  and  $\mathbf{H}_3$  ensure that the function  $\nabla J_x(\cdot)$  is differentiable for all  $x$  and the inverse of its differential at  $\zeta_0$ , denoted by  $H_x(\zeta_0)$  exists.

Furthermore, for all  $k = 1, \dots, K$  and  $s = 1, \dots, S$ , and for all  $t \geq 0$ , we have

$$\nabla J_{\hat{I}}(\zeta_0) - \nabla J_I(\zeta_0) = -2 \sum_{s=1}^S \sum_{k=1}^{K-1} d_k(\theta_0) \int_{\tau_1}^{\tau_2} \left( \log \frac{\hat{\Lambda}_{X_{k+1}^s}(t)}{\hat{\Lambda}_{X_k^s}(t)} - \log \frac{\Lambda_{X_{k+1}^s}(t)}{\Lambda_{X_k^s}(t)} \right) dt.$$

Let the function  $f$  defined on  $\mathcal{D}^{K \times S}[\tau_1, \tau_2]$  by for all  $y = (y_1^1, \dots, y_K^1, \dots, y_1^S, \dots, y_K^S)$ ,

$$f(y) = -2 \sum_{s=1}^S \sum_{k=1}^{K-1} d_k(\theta_0) \int_{\tau_1}^{\tau_2} (y_{k+1}^s(t) - y_k^s(t)) dt.$$

Thus, we have

$$\begin{aligned} \sqrt{M} \left( \nabla J_{\hat{I}}^{(1)}(\zeta_0) - \nabla J_I^{(1)}(\zeta_0) \right) &= \sqrt{M} \left( -2 \sum_{s=1}^S \sum_{k=1}^{K-1} d_k(\theta_0) \int_{\tau_1}^{\tau_2} \left( \log \frac{\hat{\Lambda}_{X_{k+1}^s}(t)}{\hat{\Lambda}_{X_k^s}(t)} - \log \frac{\Lambda_{X_{k+1}^s}(t)}{\Lambda_{X_k^s}(t)} \right) dt \right) \\ &= \sqrt{M} (f(\hat{L}_X(\cdot)) - f(L_X(\cdot))). \end{aligned}$$

where  $L_X(\cdot)$  and  $\hat{L}_X(\cdot)$  are such as defined above.

Since  $f$  is a linear function, by considering the convergence of (5.14) and applying the functional delta-method, we obtain the following weak convergences when  $M \rightarrow +\infty$ :

$$\sqrt{M} \left( \nabla J_{\hat{I}}^{(1)}(\zeta_0) - \nabla J_I^{(1)}(\zeta_0) \right) \xrightarrow{\mathcal{D}} -2 \sum_{k=1}^S \sum_{k=1}^{K-1} d_k(\theta_0) \int_{\tau_1}^{\tau_2} \left( \frac{U_{X_{k+1}^s}(t)}{\Lambda_{X_{k+1}^s}(t)} - \frac{U_{X_k^s}(t)}{\Lambda_{X_k^s}(t)} \right) dt = \mathcal{L}_1.$$

With a similar line of reasoning, we can establish when  $M \rightarrow +\infty$  that

$$\sqrt{M} \left( \nabla J_{\hat{I}}^{(2)}(\zeta_0) - \nabla J_I^{(2)}(\zeta_0) \right) \xrightarrow{\mathcal{D}} -2\gamma_0 \sum_{k=1}^S \sum_{k=1}^{K-1} \dot{d}_k(\theta_0) \int_{\tau_1}^{\tau_2} \left( \frac{U_{X_{k+1}^s}(t)}{\Lambda_{X_{k+1}^s}(t)} - \frac{U_{X_k^s}(t)}{\Lambda_{X_k^s}(t)} \right) dt = \mathcal{L}_2.$$

That leads to

$$\sqrt{M} (\nabla J_{\hat{I}}(\zeta_0) - \nabla J_I(\zeta_0)) \xrightarrow{\mathcal{D}} \mathcal{L} = \begin{pmatrix} \mathcal{L}_1 \\ \mathcal{L}_2 \end{pmatrix},$$

and ensuring the fourth assumption of the proposition.

Moreover, we have

$$\begin{aligned}\sqrt{M} \left( \nabla J_{\hat{I}}^{(1)}(\hat{\zeta}) - \nabla J_I^{(1)}(\hat{\zeta}) \right) &= -2 \sum_{s=1}^S \sum_{k=1}^{K-1} (d_k(\hat{\theta})(\hat{I}_k^s - \hat{\gamma} \tau d_k(\hat{\theta})) - d_k(\hat{\theta})(I_k^s - \hat{\gamma} \tau d_k(\hat{\theta}))) \\ &= -2 \sum_{s=1}^S \sum_{k=1}^{K-1} d_k(\hat{\theta}) \sqrt{M} (\hat{I}_k^s - I_k^s)\end{aligned}$$

and

$$\begin{aligned}\sqrt{M} \left( \nabla J_{\hat{I}}^{(1)}(\zeta) - \nabla J_I^{(1)}(\zeta) \right) &= \sqrt{M} \left( \nabla J_{\hat{I}}^{(1)}((\gamma, \theta)) - \nabla J_I^{(1)}(\gamma, \theta) \right) \\ &= -2 \sum_{s=1}^S \sum_{k=1}^{K-1} (d_k(\theta)(\hat{I}_k^s - \gamma \tau d_k(\theta)) - d_k(\theta)(I_k^s - \gamma \tau d_k(\theta))) \\ &= -2 \sum_{s=1}^S \sum_{k=1}^{K-1} d_k(\theta) \sqrt{M} (\hat{I}_k^s - I_k^s).\end{aligned}$$

That leads to

$$\begin{aligned}& \left| \sqrt{M} \left( \nabla J_{\hat{I}}^{(1)}(\hat{\zeta}) - \nabla J_I^{(1)}(\hat{\zeta}) \right) - \sqrt{M} \left( \nabla J_{\hat{I}}^{(1)}(\zeta) - \nabla J_I^{(1)}(\zeta) \right) \right| \\ &= \left| -2 \sum_{s=1}^S \sum_{k=1}^{K-1} (d_k(\hat{\theta}) - d_k(\theta)) \sqrt{M} (\hat{I}_k^s - I_k^s) \right|\end{aligned}$$

From the convergence given by Theorem 5.4 and the assumption  $\mathbf{H}_4$ , an other application of the continuous mapping theorem gives  $d_k(\hat{\theta}) \xrightarrow{\mathbb{P}} d_k(\theta)$  and thus  $|d_k(\hat{\theta}) - d_k(\theta)| = o_p(1)$ . Moreover, from (5.13), we have  $|\sqrt{M} (\hat{I}_k^s - I_k^s)| = O_p(1)$ , for all  $k = 1, \dots, K-1$  and  $s = 1, \dots, S$ . Thus

$$|(d_k(\hat{\theta}) - d_k(\theta)) \sqrt{M} (\hat{I}_k^s - I_k^s)| = o_p(1), \text{ for all } k = 1, \dots, K-1 \text{ and } s = 1, \dots, S.$$

That leads to  $\left| \sqrt{M} \left( \nabla J_{\hat{I}}^{(1)}(\hat{\zeta}) - \nabla J_I^{(1)}(\hat{\zeta}) \right) - \sqrt{M} \left( \nabla J_{\hat{I}}^{(1)}(\zeta) - \nabla J_I^{(1)}(\zeta) \right) \right| = o_p(1)$ . Thus, we obtain

$$\frac{\left| \sqrt{M} \left( \nabla J_{\hat{I}}^{(1)}(\hat{\zeta}) - \nabla J_I^{(1)}(\hat{\zeta}) \right) - \sqrt{M} \left( \nabla J_{\hat{I}}^{(1)}(\zeta) - \nabla J_I^{(1)}(\zeta) \right) \right|}{1 + \sqrt{M} \|\hat{\zeta} - \zeta_0\|_2} \xrightarrow{\mathbb{P}} 0, \text{ when } M \rightarrow +\infty. \quad (5.17)$$

Moreover, we have

$$\sqrt{M} \left( \nabla J_{\hat{I}}^{(2)}(\zeta) - \nabla J_I^{(2)}(\zeta) \right) = \sqrt{M} \left( \nabla J_{\hat{I}}^{(2)}((\gamma, \theta)) - \nabla J_I^{(2)}(\gamma, \theta) \right) = -2\hat{\gamma} \sum_{s=1}^S \sum_{k=1}^{K-1} \hat{d}_k(\hat{\theta}) \sqrt{M} (\hat{I}_k^s - I_k^s)$$

and

$$\left| \sqrt{M} \left( \nabla J_{\hat{I}}^{(2)}(\hat{\zeta}) - \nabla J_I^{(2)}(\hat{\zeta}) \right) - \sqrt{M} \left( \nabla J_{\hat{I}}^{(2)}(\zeta) - \nabla J_I^{(2)}(\zeta) \right) \right| = o_p(1).$$

Using a similar line of reasoning as above, we can show that

$$\frac{\left| \sqrt{M} \left( \nabla J_I^{(2)}(\hat{\zeta}) - \nabla J_I^{(2)}(\zeta) \right) - \sqrt{M} \left( \nabla J_I^{(2)}(\hat{\zeta}) - \nabla J_I^{(2)}(\zeta) \right) \right|}{1 + \sqrt{M} \|\hat{\zeta} - \zeta_0\|_2} \xrightarrow{\mathbb{P}} 0, \text{ when } M \rightarrow +\infty. \quad (5.18)$$

The relation (5.17) and (5.18) lead to the following convergence that satisfies the last assumption of the proposition. In other words :

$$\frac{\|\sqrt{M} (\nabla J_I(\hat{\zeta}) - \nabla J_I(\zeta)) - \sqrt{M} (\nabla J_I(\hat{\zeta}) - \nabla J_I(\zeta))\|}{1 + \sqrt{M} \|\hat{\zeta} - \zeta_0\|_2} \xrightarrow{\mathbb{P}} 0, \text{ when } M \rightarrow +\infty.$$

Thus, applying the announced proposition (5.7), we obtain that when  $M \rightarrow +\infty$ ,

$$\sqrt{M} (\hat{\zeta} - \zeta_0) \xrightarrow{\mathcal{D}} H_I^{-1}(\zeta_0) \mathcal{L}, \quad (5.19)$$

where  $\mathcal{L}$  is the random variable defined by

$$\mathcal{L} = \begin{pmatrix} -2 \sum_{k=1}^S \sum_{k=1}^{K-1} d_k(\theta_0) \int_{\tau_1}^{\tau_2} \left( \frac{U_{X_{k+1}^s}(t)}{\Lambda_{X_{k+1}^s}(t)} - \frac{U_{X_k^s}(t)}{\Lambda_{X_k^s}(t)} \right) dt \\ -2\gamma_0 \sum_{k=1}^S \sum_{k=1}^{K-1} \dot{d}_k(\theta_0) \int_{\tau_1}^{\tau_2} \left( \frac{U_{X_{k+1}^s}(t)}{\Lambda_{X_{k+1}^s}(t)} - \frac{U_{X_k^s}(t)}{\Lambda_{X_k^s}(t)} \right) dt \end{pmatrix}.$$

This concludes the proof of the theorem.  $\square$

## 5.6 Simulation study

### 5.6.1 Simulation design

In order to assess the accuracy of the proposed estimators, we have conducted a Monte Carlo (MC) simulation study. For this purpose, we have simulated data from the extended GFRR model with covariates, using different combinations of parameter values representing various scenarios. Then, we have applied our estimators to the simulated data and compared the estimated values with the true parameter values used for simulation.

The interfailure times have been simulated according to an EGFR model with covariates defined by the Equation (5.5) with the following parameters :

- $\Lambda_{X^1}(t) = \left(\frac{t}{a}\right)^b$ , where  $a = 0.5$  and  $b = 3$ ;
- $Z_s = (Z_s^1, Z_s^2)^T$ , where  $Z_s^1 \sim Unif(0, 1)$  and  $Z_s^2 \sim \mathcal{N}(2, 1)$ , for all  $s = 1, \dots, S$ ;
- $a(k, \theta) = (k - 1)^\theta$ .

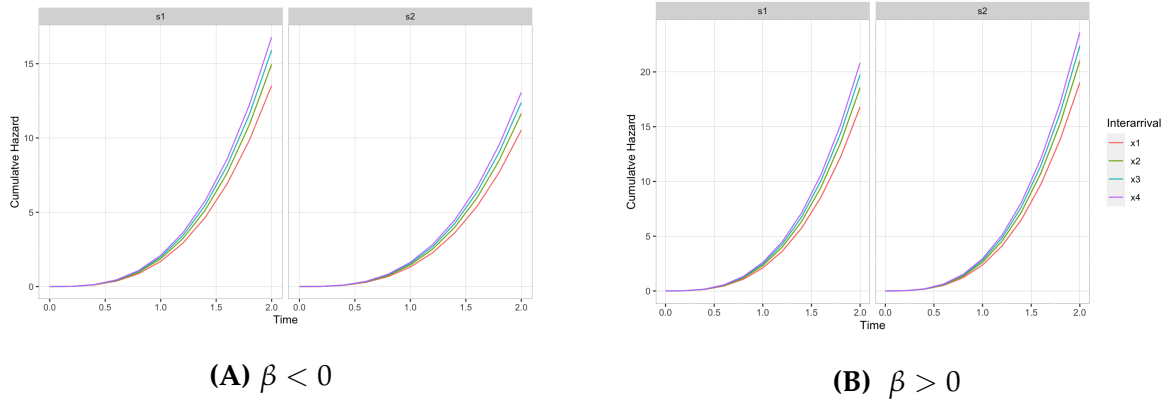
For the Monte Carlo simulation study of each estimator, we have considered three simulation designs. In the first design, the number of imperfect maintenance  $K$  and the number of systems  $S$  were fixed to  $K = 5$  and  $S = 5$ , respectively, while the number  $M$  of perfect repairs varied in  $\{50, 150, 300\}$ . In the second design,  $S = 5$ ,  $M = 100$  and  $K$  varied in  $\{5, 10, 15\}$ . In the last design,  $K = 5$ ,  $M = 100$  and  $S$  varied in  $\{5, 10, 15\}$ . For each design, the Mean Squared Error (MSE) values were computed based on 5000 replications of each estimator. More precisely, if we consider the  $\beta$  parameter, the MSE value for  $N$  replications of  $\hat{\beta}$  is given by

$$MSE(\hat{\beta}) = \frac{1}{N} \sum_{n=1}^N (\beta - \hat{\beta}_n)^2.$$

Before considering the results of our MC simulation study, let us analyze the effects of the Euclidean parameters on the cumulative hazard rate of the interfailures.

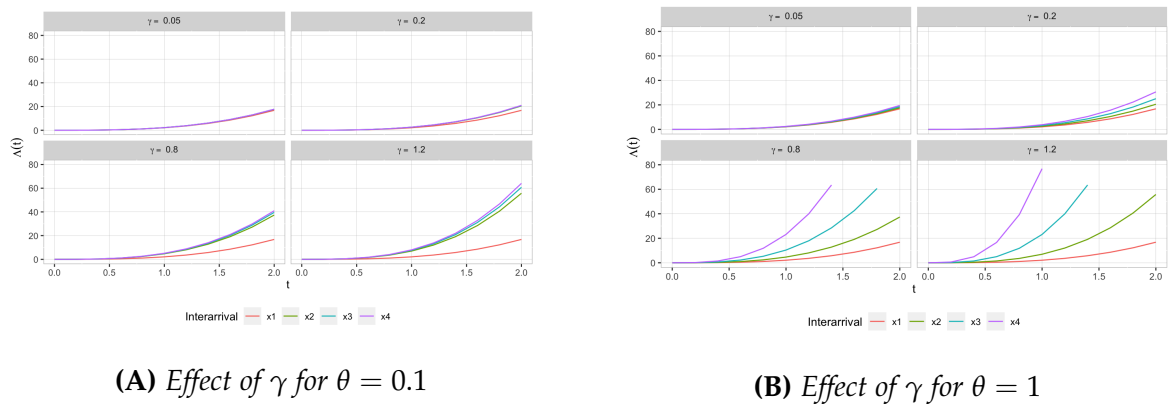


Firstly, we illustrate the effect of covariates in the simple case with only one covariate (i.e.  $p = 1$ ) and two systems ( $S = 2$ ) denoted by  $s_1$  and  $s_2$ , with covariate values  $Z_1 = 0.38$  and  $Z_2 = 0.78$ , respectively. We set the values of  $\gamma$  and  $\theta$  to be 0.1 and 0.7, respectively, and define  $a(k, \theta) = (k - 1)^\theta$ . Figure 5.2 shows the cumulative hazard curves of the first four interfailures of the two systems when  $\beta \in \{-0.25, 0.5\}$ . It can be seen that when  $\beta < 0$  (where  $\beta = -0.25$ ), the system with the smaller covariate value has higher risk (i.e. it exhibits higher values of the hazard function). Conversely, when  $\beta > 0$  (where  $\beta = 0.5$ ), the system with the larger covariate value shows a higher risk. This illustrates the influence of the covariates on the hazard rate and how different values of  $\beta$  affect the ordering of the systems based on their covariate values.



**Figure 5.2** – Cumulative hazard curves of the interarrival times  $X_1$  (red lines),  $X_2$  (green lines),  $X_3$  (blue lines) and  $X_4$  (purple lines) of the systems  $s_1$  and  $s_2$  for distincts values of the  $\beta$  parameter : on the left  $\beta = -0.25$  and on the right  $\beta = 0.5$ .

Next, we analyze the effect of repairs, which is modeled by  $\gamma$  (or  $\rho$ ). For this analysis, we consider the system  $s_1$  with covariate value  $Z_1 = 0.38$ . The value of  $\beta$  is fixed at  $-0.25$  and  $\theta$  is either equal to 0.1 or to 1. The cumulative hazard curves for  $\gamma$  taking values in  $\{0.1, 0.5, 0.8, 1.2\}$  are displayed in Figure 5.3. It is observed that the cumulative curves of the interfailure times are similar for small values of  $\gamma$  ( $\gamma = 0.1$  and  $\gamma = 0.5$ ). In contrast, when  $\gamma$  takes larger values (here  $\gamma = 1.2$ ), the cumulative hazard functions of the interfailure times exhibit considerable differences. Consequently, the scales of these interfailure times become significantly different, leading to an unrealistic description of the system behavior after repair.

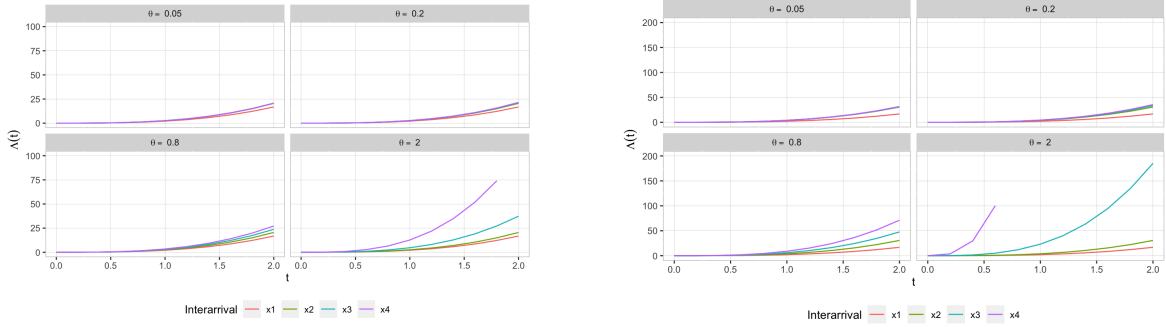


**Figure 5.3** – Cumulative hazard curves of the interarrival times  $X_1$  (red line),  $X_2$  (green lines),  $X_3$  (blue lines) and  $X_4$  (purple lines) with  $\gamma \in \{0.05, 0.2, 0.8, 1.2\}$  and  $a(k, \theta) = (k - 1)^\theta$ .

Regarding the parameter  $\theta$ , it is evident that it also affects the amplitude of the cumulative

hazard functions and these effect depends on the form of the function  $a(\cdot)$ . Figure 5.4 illustrates the effects of the different values of  $\theta$  on the cumulative hazard functions of system  $s_1$  when  $\gamma \in \{0.1, 0.5\}$  and  $a(k, \theta) = (k - 1)^\theta$ . It is observed that the failure process tends to "explode" for high values of  $\theta$ , regardless of the value of  $\gamma$  (with a faster explosion for larger values of  $\gamma$ ).

These observations underscore the importance of carefully selecting appropriate values for both  $\gamma$  and  $\theta$  to ensure a reliable and realistic model that accurately represents the behavior of the system over time.



(A) Effect of  $\theta$  for  $\gamma = 0.2$

(B) Effect of  $\theta$  for  $\gamma = 0.6$

**Figure 5.4** – Cumulative hazard curves of the interarrival times  $X_1$  (red lines),  $X_2$  (green lines),  $X_3$  (blue lines) and  $X_4$  (purple lines) for  $\theta \in \{0.05, 0.5, 0.8, 2\}$  when  $a(k, \theta) = (k - 1)^\theta$ .

### 5.6.2 MC simulation results for $\hat{\beta}$

To analyze the performance of our estimator  $\hat{\beta}$ , we use the values  $\gamma = 0.2$ ,  $\theta = 0.7$  and the two cases (for the vector  $\beta = (\beta_1, \beta_2)$ ):

- Case 1 :  $\beta_1 = -0.25$  and  $\beta_2 = 0.5$ ;
- Case 2 :  $\beta_1 = 0.5$  and  $\beta_2 = -0.25$ .

For each case, the simulation study is performed under the simulation designs described earlier. Figure 5.5 illustrates the behaviour of the MSE when one of the sample size of parameters increases.

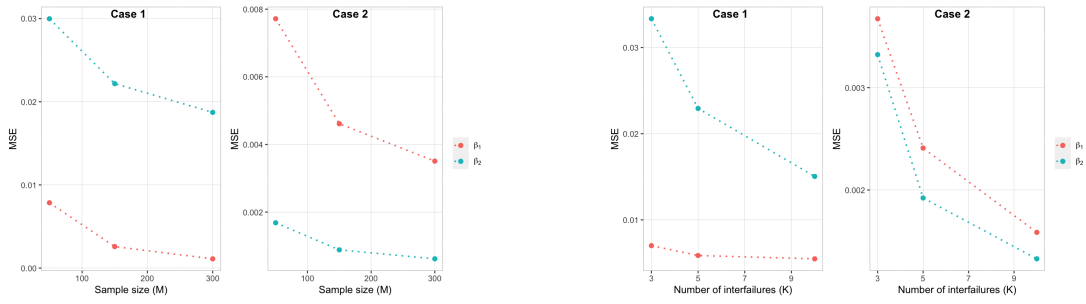
In all cases, one can see as depicted that the values of the Mean Squared Error (MSE) decreases as the sample size  $M$ , the number of minimal maintenance  $K$  and the number of systems  $S$  increases respectively. Furthermore, it can be observed that the estimated values are slightly better in case 2 than in case 1. This can be explained by the fact that in case 1, the value of  $\beta_2$  is larger, which introduces a greater variability in the quantity  $\beta^T Z$  compared to case 2.

### 5.6.3 MC simulation results for the estimators $(\hat{\gamma}, \hat{\theta})$

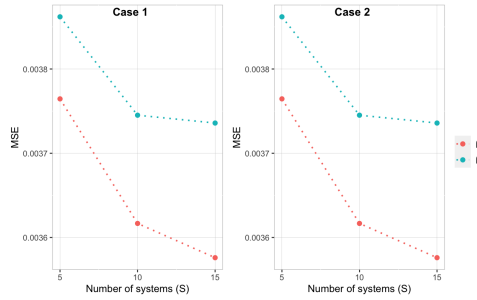
The study of the performance of the estimators  $(\hat{\gamma}, \hat{\theta})$  is made under the three simulation designs described above. For this,  $\beta = (\beta_1, \beta_2) = (0.25, 0.5)$  and we consider the following two cases :

- Case 1 :  $\gamma = 0.2$  and  $\theta = 2$ ;
- Case 2 :  $\gamma = 0.8$  and  $\theta = 0.5$ .

As shown in sub-figures A, B, and C of Figure 5.6, the Mean Squared Error (MSE) values decrease with the number of perfect repairs  $M$ , the number of failures  $K$ , and the number of systems  $S$ , respectively. However, we observed that the decreasing rate with the number of systems is relatively low compared to the other factors.

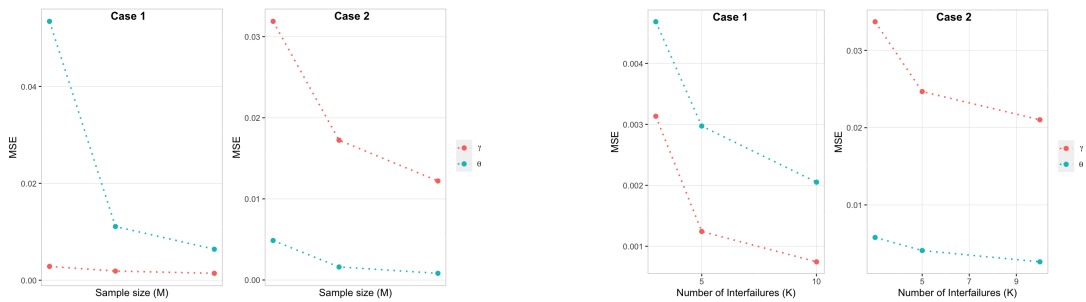


(A) MSE values of  $\hat{\beta}$  when  $K = 5, S = 5$  and the sample size  $M \in \{50, 150, 300\}$  (B) MSE values of  $\hat{\beta}$  when  $S = 5, M = 100$  and  $K \in \{3, 5, 10\}$

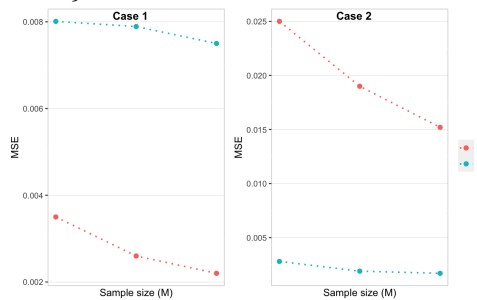


(C) MSE values of  $\hat{\beta}$  when  $K = 5, M = 100$  and  $S \in \{5, 10, 15\}$

Figure 5.5 – Simulation results for  $\hat{\beta}$  under the three simulation designs.



(A) MSE values of  $(\hat{\gamma}, \hat{\theta})$  when  $K = 5, S = 5$  and for  $M \in \{50, 150, 300\}$  (B) MSE values of  $(\hat{\gamma}, \hat{\theta})$  when  $S = 5, M = 100$  and  $K \in \{3, 5, 10\}$



(C) MSE values of  $(\hat{\gamma}, \hat{\theta})$  when  $K = 5, M = 100$  and  $S \in \{5, 10, 15\}$

Figure 5.6 – Simulation results for  $(\hat{\gamma}, \hat{\theta})$  under the three simulation designs.

## 5.7 Conclusion and Perspectives

We propose in this paper a model for imperfect maintenance that accounts effect of covariates. This EGFR model with covariates is introduced and a semiparametric inference for this latter is performed. In addition, we have established the asymptotic normality of the estimators and conducted a numerical study which proves good performance of our estimator on large simple size.

Furthermore, a persistent challenge in reliability studies lies in identifying significant covariates that affect system reliability. Therefore, an advancement in our research would involve proposing a model selection procedure. This could entail utilizing statistical tests, such as the Wald test, to evaluate the significance of the  $\beta$  parameters and identify the most influential covariates.

Additionally, a crucial consideration when employing a statistical model in practical applications is model adequacy. As a result, a promising avenue for further research would be to devise goodness-of-fit tests to evaluate the appropriateness of the model when applied to real-world data.

These approaches could empower engineers to apply our model in practical contexts and gain a deeper understanding of the factors that influence system reliability.

---

## Conclusion générale

---

Les travaux de thèse présentés dans ce manuscrit, qui s'articule en cinq chapitres, abordent deux problématiques fondamentales dans le domaine de la fiabilité, en particulier dans le contexte du Génie Civil. Ces deux problématiques sont au cœur de notre recherche : la modélisation de l'évolution inévitable de la dégradation des systèmes opérationnels et l'optimisation de leur maintenance et de leur fiabilité.

Dans un premier temps, nous avons réalisé une étude statistique approfondie de la carbonatation, l'une des principales causes de la corrosion des armatures des structures en béton armé. Cette étude a débouché sur le développement de modèles statistiques accessibles, à la fois en termes d'application et d'interprétation. Ces modèles permettent de mieux comprendre et de prédire la fiabilité des structures en béton armé exposées à la carbonatation naturelle.

Dans un second temps, nous avons cherché à améliorer les performances des modèles de maintenance imparfaite de systèmes réparables. Nous avons introduit l'effet des covariables dans les modèles la réduction géométrique du taux de défaillance ou modèles GFRR (Geometric Failure Rate Reduction), améliorant ainsi leur capacité à prédire l'état réel du système après réparation.

Le premier chapitre a établi une base solide en réalisant une revue bibliographique complète, abordant de manière approfondie les connaissances essentielles concernant les bétons armés, le processus de corrosion induit par la carbonatation, ainsi que les méthodes statistiques appliquées dans ce contexte. Nous avons pu constater que la carbonatation, influencée par de nombreux facteurs, réduit la fiabilité et la sécurité des structures en béton armé en initiant le processus de corrosion des armatures en acier. Dans ce contexte, différents modèles prédictifs de la carbonatation ont été présentés, chacun utilisant des méthodes de modélisation spécifiques. Les modèles analytiques basés sur des approches statistiques se sont avérés efficaces et accessibles aux ingénieurs en génie civil. Cependant, la plupart de ces modèles contiennent un grand nombre de paramètres d'entrée ou sont calibrés sur des données issues de tests de carbonatation accélérée. Cela souligne la nécessité de disposer d'un modèle statistique simple et efficace qui permette de prédire la fiabilité d'une structure en béton armé exposée à la carbonatation naturelle. Cette analyse bibliographique s'est donc avérée être une étape cruciale pour poser les bases conceptuelles et théoriques du travail à venir, fournissant ainsi un cadre solide pour le développement des modèles ultérieurs.

Le deuxième chapitre, bien que plus concis, joue un rôle clé en exposant en détail les bases de données pertinentes utilisées pour la modélisation statistique de la carbonatation. La qualité et la fiabilité de ces données sont fondamentales pour la robustesse et la validité des modèles qui en découleront.

Les trois chapitres suivants constituent le cœur de ce mémoire, présentant des articles soit déjà publiés, soit en cours de soumission à des revues scientifiques. Le chapitre 3 se penche

sur la modélisation de la profondeur de carbonatation. Deux modèles statistiques, calibrés sur les données issues de carbonatation naturelle sont proposés : le premier est basé sur les GLMs (Generalized Linear Models) et le second sur les processus de dégradation Gamma. Ce chapitre propose également une comparaison éclairante entre ces modèles et le modèle de référence du *fib-Model Code 2010*. Il en ressort que les modèles proposés présentent de meilleures performances en termes d'erreur quadratique moyenne (Mean Squared Error), et que le modèle de dégradation Gamma est le plus performant des deux. Cette analyse comparative renforce ainsi la pertinence des modèles développés.

Les modèles de profondeur de carbonatation développés dans le chapitre 3 ne prennent pas en compte la variabilité temporelle des caractéristiques environnementales. Pour remédier à cette limitation, dans le chapitre 4, une analyse de survie de la carbonatation a été effectuée. Ce chapitre détaille la procédure de modélisation du temps de carbonatation, c'est-à-dire le temps mis par le processus de carbonatation pour atteindre les aciers d'armatures, en utilisant des modèles de régression en survie. Un modèle à risque proportionnel du temps de carbonatation, qui inclut les principaux facteurs de carbonatation, est proposé. Ce modèle trouve une application concrète en évaluant la fiabilité d'un matériau en béton armé exposé à différentes conditions climatiques.

Les régressions linéaires généralisées et les processus de dégradation Gamma demeurent des outils statistiques puissants pour évaluer l'usure des structures en béton armé en prenant en considération divers facteurs de détérioration. Cependant, lorsque la dégradation dépend de paramètres évoluant dans le temps, les modèles de régression linéaire généralisée montrent leurs limites. Les modèles de dégradation Gamma peuvent prendre en compte ces paramètres dépendant du temps à condition d'avoir des données longitudinales. Pour l'étude spécifique de la carbonatation, cela suppose de disposer de mesures à différents moments de vie de la structure. Malheureusement, la littérature scientifique ne fournit pas de telles données, principalement en raison de la lenteur du processus de carbonatation et de la nature destructive des mesures. Ainsi, nous avons focalisé le chapitre 4 sur la modélisation du temps de carbonatation en utilisant des modèles de régression en survie. Ces modèles ont permis de prendre en compte l'aspect temporel des facteurs environnementaux dans la modélisation du temps de carbonatation en utilisant des données de type *current status data* provenant de la bibliographie. En résumé, les modèles que nous avons élaborés pour la profondeur de carbonatation et le temps de carbonatation peuvent être facilement utilisés pour prédire et évaluer le niveau de ce processus dans les structures en béton armé, ce qui facilitera la planification des actions de maintenance.

Une des actions des mesures susceptibles d'améliorer l'efficacité de l'ensemble des modèles proposés est l'enrichissement de la base de données. En effet, une base de données plus complète pourrait améliorer la précision des modèles, leur permettant ainsi de prédire plus précisément le degré de carbonatation réel des structures en béton armé exposées à la carbonatation naturelle.

Enfin, le chapitre 5 constitue un pas en avant dans l'amélioration des performances des modèles de maintenance imparfaite en présentant un modèle EGFR à covariable. L'usure intrinsèque du système et l'efficacité des maintenances, déjà prises en compte par les modèles EGFR, peuvent être en partie expliquées par des covariables (fixes dans le temps). Nous avons mis en place un cadre mathématique permettant de manipuler ces modèles. L'estimation sémi-paramétrique des paramètres du modèle et l'étude des propriétés asymptotiques des estimateurs sont présentées. Une étude numérique menée pour vérifier et valider la convergence asymptotique des estimateurs est également présentée dans ce chapitre. En outre, diverses perspectives de recherche future ont été évoquées à la fin de ce chapitre.

Ce travail représente une contribution à la compréhension et à la modélisation de la fiabilité des systèmes réparables.

---

## Bibliographie

---

- 13670, N. E. (2013). NF EN 13670, Annex F. *Execution of concrete structures*.
- 206/CN, N. E. (2014). BETON-SPÉCIFICATION, PERFORMANCE, PRODUCTION et CONFORMITE, Complément national à la norme NF EN 206.
- A, I., M, A., A, A.-P., J, B., A, B., A, B., J, D., R, E., H, E., J, F., V, H., L, J., Z, K., S, M., M, P., V-H, P., M, R., S, R., M, S., & SUTTIE. (2019). CAMS global reanalysis (EAC4). Copernicus Atmosphere Monitoring Service (CAMS) Atmosphere Data Store (ADS), <https://ads.atmosphere.copernicus.eu/cdsapp#!/dataset/cams-global-ghg-reanalysis-egg4-monthly?tab=overview>.
- AKAIKE, H. (1998). Information Theory and an Extension of the Maximum Likelihood Principle. In E. PARZEN, K. TANABE & G. KITAGAWA (Éd.), *Selected Papers of Hirotugu Akaike* (p. 199-213). Springer New York. [https://doi.org/10.1007/978-1-4612-1694-0\\_15](https://doi.org/10.1007/978-1-4612-1694-0_15)
- ALEXANDER, M., MACKECHNIE, J., & AM, W. (2007). Carbonation of concrete bridge structures in three South African localities. *Elsevier Science*.
- ALEXANDER, M., MACKECHNIE, J., & YAM, W. (2007). Carbonation of concrete bridge structures in three South African localities. *Cement Concrete Composites*, 29, 750-759.
- ANDERSEN, P. K., BORGAN, O., GILL, R. D., & KEIDING, N. (1993). *Statistical models based on counting processes*. Springer Science & Business Media.
- ANDERSEN, P. K., & KEIDING, N. (2002). Multi-state models for event history analysis [PMID : 12040698]. *Statistical Methods in Medical Research*, 11(2), 91-115. <https://doi.org/10.1191/0962280202SM276ra>
- ATIŞ, C. D. (2004). Carbonation-Porosity-Strength Model for Fly Ash Concrete. *Journal of Materials in Civil Engineering*, Vol. 16, Issue 1, February 2004, American Society of Civil Engineers.
- BAGDONAVICIUS, V., & NIKULIN, M. (2000a). Estimation in degradation models with explanatory variables. *life time and data analysis*, 85-103.
- BAGDONAVICIUS, V., & NIKULIN, M. (2000b). Estimation in degradation models with explanatory variables. *life time and data analysis*.
- BAGDONAVIČIUS, V., KRUPIS, J., & S.NIKULIN, M. (2011). Non-parametric Tests for Complete Data. *ISTE Ltd and John Wiley Sons, Inc*.
- BAKKER, R. (1994). Predictions of service life of reinforcement in concrete under different climatic conditions at given cover. *International Conference on Corrosion and corrosion protection of steel in concrete*, Sheffield, UK.
- BALAYSSAC, J. P., DÉTRICHÉ, C. H., & GRANDET, J. (1994). Effects of curing upon carbonation of concrete. *ELSEVIER, Construction and building Materials*, Volume 9.
- BALEN, K. V., & GEMERT, D. V. (1994). Modelling lime mortar carbonation. *Materials and Structures*, Vol. 27, 393-398.
- BAMFORTH, P. (2004). Enhancing reinforced concrete durability : Guidance on selecting measure for minimizing the risk of corrosion of reinforcement in concrete. *T.C. Society Ed*, 61.

- BARRY, B., & SELLIER, A. (1990). Coupled moisture-carbon dioxide-calcium transfer model for carbonation of concrete. *Cement and Concrete Research*, 20(6), 875-883.
- BARY, B., & SEILLIER, A. (2003). Coupled moisture-carbon dioxide-calcium transfer for carbonation of concretes. *Proceeding of EUROCC-2003, Computational modelling of the Concrete Structures*.
- B.G., S., H., B., & M.G., A. (2015). Oxygen permeability of concrete and its relation to carbonation. *Construction and Building Materials*, 85, 30-37.
- BORDES, L., & MERCIER, S. (1983). Imperfect Repair. *Journal of Applied Probability*, 20(4), 851-859. <https://doi.org/https://doi.org/10.2307/3213596>
- BORDES, L., & MERCIER, S. (2013). Extended geometric processes : semiparametric estimation and application to reliability. *Journal of The Iranian Statistical Society*, 12(1), 1-34.
- BRENIÈRE, L. (2021). *Effets conjoints du vieillissement, de la maintenance et de l'hétérogénéité sur des systèmes réparables : modélisation, inférence et prise de décision* (publication N° 2021GRALM008) [Theses]. Université Grenoble Alpes [2020-....] <https://theses.hal.science/tel-03295882>
- CEB-FIP. (2010). Model Code for Service Life Design, bulletin 34, section B1.2.7.
- COX, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society, Series B*, 34(2).
- DAUXOIS, J.-Y., GASMI, S., & GAUDOIN, O. (2018). Semiparametric inference for an extended geometric failure rate reduction model. *Journal of Statistical Planning and Inference*.
- der VAART, A. W. V. (1998). *Asymptotic Statistics*. Cambridge University Press.
- DOYEN, L., & GAUDOIN, O. (2004). Classes of imperfect repair models based on reduction of failure intensity or virtual age [Selected papers from ESREL 2002]. *Reliability Engineering System Safety*, 84(1), 45-56. [https://doi.org/https://doi.org/10.1016/S0951-8320\(03\)00173-X](https://doi.org/https://doi.org/10.1016/S0951-8320(03)00173-X)
- DULLIEN, F. (1979). Porous media fluid transport and pore structure. *New York Academic Press*.
- EKOLU, S. O. (2020). Model for natural carbonation prediction (NCP) : Practical application worldwide to real life functioning concrete structures. *Engineering Structures, Volume 224*.
- EKOLU, S. O., & van EDE, W. (2014). Condition assessment of a Johannesburg skyscraper. *International Conference on Construction Materials and Structures (ICCMATS), Johannesburg, South Africa*, 1052-1059.
- FABRICE, D. (2008). *Approche probabiliste de la durabilité des bétons en environnement marin* [thèse de doct.].
- FIGUEIREDO, C., & NEPOMUCENO, A. (2005). the Carbonation of reinforced concrete building in BRAZIL.
- FINKELSTEIN, M. (2008). *Failure rate modelling for reliability and risk*. Springer Science & Business Media.
- FUSHIKI, T. (2009). Estimation of prediction error by using K-fold cross-validation. *Springer Science, Business Media*.
- G.C., B. (2004). Carbonation induced corrosion of reinforcement. *Challenge of Concrete Construction, Repair, Rejuvenation and Enhancement of Concrete*, 3.
- GREVE-DIERFELD, S., & GEHLEN, C. (2014). Performance based deemed to satisfy rules. *The Fourth International fib congress, Mumbai*.
- GUIGLIA, M., & TALIANO, M. (2013). Comparison of carbonation depths measured on in-field exposed existing r.c. structures with predictions made using fib-Model Code 2010. *ELSEVIER, Cement and Concrete Research Volume 38*.
- HUANG, J., & WELLNER, J. A. (1997). Interval Censored Survival Data : A Review of Recent Progress. *The First Seattle Symposium in Biostatistics : Survival Analysis*, 92, 123-169.
- HYVERT, N. (2009). *Application de l'approche probabiliste à la durabilité des produits préfabriqués en béton* [thèse de doct.].



- IPCC. (2007). *Climate Change 2007 : The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (rapp. tech.).
- ISGOR, O. B., & RAZAQPUR, A. (2004). Finite element modeling of coupled heat transfer, moisture transport and carbonation processes in concrete structures. *Cement and Concrete Composites*, 26(1), 57-73.
- J, C., & GLASSER, F. P. (1991-1992). The reaction between cement and natural waters containing dissolved carbon dioxide. *Advances in Cement Research*, 4, 119-134.
- JIAN, H. (1996). Efficient Estimation for the Proportional Hazards Model with Interval Censoring. *The Annals of Statistics*, 24(2), 540-568.
- JOHANNESON, B. (2003). A theoretical describing diffusion of a mixture of different types of ions in pore solution of concrete coupled to moisture transport. *Cement and Concrete Research*, 33, 481-488.
- K., S., & L., F. (2007). Carbonation rates of concretes containing high volume of pozzolanic materials. *Construction and Building Materials*, 37, 1647-1653.
- KAHLE, W., MERCIER, S., & PAROISSIN, C. (2016). *Degradation Processes in Reliability* (T. 3). Wiley-ISTE.
- KARI, O. P. (2011). Deterioration Models with Interaction. *DuraInt Report-Task 4, Aalto University*.
- KART, A., & THOMPSON, A. (1986). Quantitative prediction of permeability in porous rocks. *Physical Review B*, 34, 8179-8181.
- KIJIMA, M. (1989). Some results for repairable systems with general repair. *Journal of Applied probability*, 26(1), 89-102.
- KIM, J.-K., & LEE, C.-S. (1999). Moisture diffusion of concrete considering self-desiccation at early ages. *Elsevier, Cement and Concrete Research, Volume 29, Issue 12, pp 1921-1927*.
- KLEIN, J. P., & MOESCHBERGER, M. L. (2003). *Survival Analysis : Techniques for Censored and Truncated Data*. Springer New York, NY. <https://doi.org/https://doi.org/10.1007/b97377>
- KLOPFER, H. (1978). The carbonation of external concrete and the control of it. *Bautenschutz Bausanier*.
- KOLMOGOROV, A. (1933). Über die Grenzwertsätze der Wahrscheinlichkeitsrechnung. *Bulletin de l'Académie des Sciences de l'URSS. Classe des sciences mathématiques et na*, (3), 363-372.
- KONG, J., ABABNEH, A., FRANGOPOL, D., & XI, Y. (2002). Reliability analysis of chloride penetration in saturated concrete. *Probabilistic Engineering Mechanics*, 17, 305-315.
- KOSOROK, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer New York, NY. <https://doi.org/10.1007/978-0-387-74978-5>
- KOUASSI, F. A., DAUXOIS, J.-Y., DUPRAT, F., LARRARD, T. D., & DEBY, F. (2022). Engineering statistical models for carbonation depth. *European Journal of Environmental and Civil Engineering*.
- L., J., B., L., & Y., C. (2000). A model for predicting carbonation of high volume fly ash concrete. *Cement and Concrete Research*, 30, 699-702.
- LAM, Y. (1992). Nonparametric inference for geometric processes. *Communications in statistics-theory and methods*, 21(7), 2083-2105.
- LAM, Y. (1998). A note on the optimal replacement problem. *Advanced Applied Probability*, 2(20), 479-482.
- LAWLESS, J., & M.CROWDER. (2004a). Covariate and random effects in a Gamma process model with application to degradation and failure. *life time and data analysis*, 213-227.
- LAWLESS, J., & M.CROWDER. (2004b). Covariate and random effects in a Gamma process model with application to degradation and failure. *life time and data analysis*.
- LAY, S. (2003). Service life design of reinforced concrete structures exposed to de-icing salts – A case study. *fib WP 5.3-1, TG2.6 The Joint Committee on Structural Safety Workshop on Probabilistic Modelling of Deterioration Processes in Concrete Structures – Proceedings*, 107.

- LEEMANN, A., NYGAARD, P., KAUFMANN, J., & LOSER, R. (2015). Relation between carbonation resistance, mix design and exposure of mortar and concrete. *Elsvier, Cement and Concrete Composites, Volume 62*, pp 33-43.
- LI, S., SUN, Y., HUANG, C., FOLLMANN, D., & KRAUSE, R. (2016). Recurrent event data analysis with intermittently observed time-varying covariates. *Stat Med*, 18(35), 3049-65.
- LIANG, M.-T., HUANG, R., & FANG, S.-A. (2013). Carbonation Service Life Prediction of Existing Concrete Viaduct/ bridges using time-dependent reliability analysis. *Journal of Marine Science and Technology, Vol. 21*.
- LIISMA, E., SEIN, S., & JÄRVPÖLD, M. (2017). The influence of carbonation process on concrete bridges and durability in Estonian practice. *IOP Conf. Ser. : Mater. Sci. Eng. 251 012072*.
- LIU, T., ZHAO, S., & YI, C. (2008). The Forecast of Carbonation Depth of Concrete Based on RBF Neural Network. *Second International Symposium on Intelligent Information Technology Application*.
- MAI-NHU, J. (2013). Corrosion des armatures du béton : couplage carbonatation/chlorures en présence de cycles hydriques. *Thèse de doctorat de l'Université de Toulouse III*.
- MCCULLAGH, P., & NELDER, J. (1989). *Generalized linear Models, second edition*. Routledge.
- MCCULLOCH, C., & SEARLE, S. R. (2001a). *Generalized, Linear, and Mixed Models* (C. HALL/CRC, Éd.).
- MCCULLOCH, C., & SEARLE, S. R. (2001b). *Generalized, Linear, and Mixed Models* (C. HALL/CRC, Éd.).
- MEEKER, W., & ESCOBAR, L. A. (1998). *Statistical Methods for Reliability Data*. Wiley, New York.
- MEIJERS, S., BIJEN, J., BORST, R. D., & FRAAIJ, A. (2005). Computational results of a model for chloride ingress in concrete including convection, drying-wetting cycles and carbonation. *Mater Structures*, 38, 145-154.
- MONTEIRO, I., BRANCO, F., de BRITO, J., & NEVES, R. (2012). Statistical analysis of the carbonation coefficient in open air concrete structures. *Construction and Building Materials*, 29, 263-269.
- MORINAGA, S. (1990). Prediction of service lives reinforced concrete buildings based on the corrosion rate of reinforcing steel. *Durability of building Materials and Component*.
- NELDER, J. A., & WEDDERBURN, R. W. M. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society. Series A (General)*, 135(3), 370-384.
- NEVES, R., BRANCO, F., & DE BRITO, J. (2013). Field assessment of the relationship between natural and accelerated concrete carbonation resistance. *Cement and Concrete Composites*, 41, 9-15. <https://doi.org/https://doi.org/10.1016/j.cemconcomp.2013.04.006>
- NING, J., CAI, C., CHEN, Y., HUANG, X., & WANG, M.-C. (2020). Semiparametric modelling and estimation of covariate-adjusted dependence between bivariate recurrent events. *Biometrics*, 4(76), 1229-1239.
- NISCHER, P. (2013). Effect of environment and concrete quality on carbonation. *Betonwerk + Fertigteil-Technik*, (11), 752-757.
- NOVÁK, S. (2015). Regression Models for Repairable Systems. *Methodology and Computing in Applied Probability*, 17(4), 963-972. <https://doi.org/https://doi.org/10.1007/s11009-014-9419-2>
- OLLIVIER, J.-P., & VICHOT, A. (2008a). *La durabilité des bétons* (P. de l'école nationale des ponts et CHAUSSÉES, Éd.).
- OLLIVIER, J.-P., & VICHOT, A. (2008b). *La durabilité des bétons* (P. de l'école nationale des ponts et CHAUSSÉES, Éd.).
- OLLIVIER, J.-P., & VICHOT, A. (2008c). *La durabilité des bétons* (P. de l'école nationale des ponts et CHAUSSÉES, Éd.).
- OLLIVIER, J.-P., & VICHOT, A. (2008d). *La durabilité des bétons* (P. de l'école nationale des ponts et CHAUSSÉES, Éd.).

- P., S. (1976). Zur Frage de zulässigen rissbeite und der erforder lichen betondeckung im stahlbetonbauunter besonderer berücksichtigung der karbonatisierung des betongs. *Technical University of Munich*.
- PAPADAKIS, V. G., VAYENAS, C. G., & FARDIS, M. N. (1991). Fundamental Modeling and Experimental Investigation of Concrete Carbonation. *ACI Materials Journal*. <https://doi.org/10.14359/1863>
- PAPADAKIS, V., & DEMIS, S. (2011). Estimation and Validation of Concrete Strength and Service Life Using Software Packages based on Predictive Models. *International Conferences on Building Materials and Componets, Porto-Portugal*.
- PARK, C., & PADGETT, W. (2005). Accelerated degradation models for failure based on geometric Brownian motion and Gamma process. *Lifetime Data analysis, vol. 11*.
- PARROT, L. J. (2009). A study of carbonation induced corrosion. *Magazine of Concrete Research, 46*, 23-28.
- PETRE-LAZAR, I. (2000). Evaluation du comportement en service des ouvrages en béton armé soumis à la corrosion des aciers. *EDF, France*.
- P.HILL, T., GORDON, F., H.FLORIN, N., & S.FENNELLA, P. (2015). Statistical analysis of the carbonation rate of concrete. *ELSEVIER, Cement and Concrete Research Volume 72*.
- REVIRON, N., & URRUTIA, A. (2016). Corrosion investigation programme on 28 Frenchcooling tower shells. *Proceeding of the international conference on industrial chimneys cooling towers, ICCT2016*, 379-86.
- RIZVI, S. S. (2017). Carbonation induced deterioration of concrete structures. *The Indian Concrete Journal*.
- SAETTA, A., SCHREFLER, B., & VITALIANI, R. (1993). The carbonation of concrete and the mechanism of moisture, heat and carbon dioxide flow through porous materials. *Cement and Concrete Research, 23(4)*, 761-772.
- SCHMITT, L. (2019). *Durabilité des ouvrages en béton soumis à la corrosion : optimisation par une approche probabiliste* [thèse de doct.].
- SILVA, A., NEVES, R., & BRITO, J. D. (2014). Statistical modelling of carbonation in reinforced concrete. *Cement and Concrete Composites, 50*, 73-81.
- SPARLING, Y. H., YOUNES, N., & LACHIN, J. M. (2006). Parametric survival models for interval-censored data with time-dependent covariates. *Biostatistics 7, 4*, pp. 599–614.
- STEFFENS, A., DINKLER, D., & AHRENS, H. (2002). Modeling carbonation for corrosion risk prediction of concretes structures. *Cement and Concrete Research, 32*, 935-941.
- SUN, J. (1978). *The Statistical Analysis of Interval-censored Failure Time Data*. Springer New York, NY. <https://doi.org/https://doi.org/10.1007/0-387-37119-2>
- TAFFESE, W. Z., SISTONEN, E. [E.], & PUTTONEN, J. (2015). CaPrM : Carbonation prediction model for reinforced concrete using machine learning methods. *ELSEVIER, Construction and Building Materials, Volume 100*.
- TAFFESE, W. Z., SISTONEN, E. [Esko], & PUTTONEN, J. (2015). SCaPrM : Carbonation prediction model for reinforced concrete using machine learning methods. *ELSEVIER, Cement and Concrete Research Volume 100*.
- TAHLAITI, M. (2010). *Étude de la pénétration des chlorures et de l'amorçage de la corrosion en zone saturée et en zone de marnage* [thèse de doct.].
- TANG, L., & NILSSON, L.-O. (1996). A Numerical Method for Prediction of Chloride Penetration into Concrete Structures. *Jennings, H., Kropp, J., Scrivener, K. (eds) The Modelling of Microstructure and its Potential for Studying Transport Properties and Durability. NATO ASI Series, vol 304. Springer, Dordrecht*.
- TERUZZI, T., CADONI, E., & FRIGERI, G. (2003). Evaluation of the residual service life of existing concrete structures : a valuable tool for maintenance planning, Paris, France. *Proc. 2nd International Rilem Workshop on "Life Prediction and Aging Management of Concrete Structures"*, 207-216.

- THIERY, M. (2006). Modélisation de la carbonatation atmosphérique des matériaux cimentaires. *Etudes et Recherches des LPC*.
- TUUTTI, K. (1982). Corrosion of steel in concrete. *Swedish Cement and Concrete Research Institute*.
- TUUTTI, K. (1992a). Corrosion of steel in concrete, *Swed. Cement-Concrete Researches Institute, Stockholm*.
- TUUTTI, K. (1992b). Corrosion of steel in concrete, *Swed. Cement-Concrete Researches Institute, Stockholm*.
- VENUAT, M., & ALEXANDRE, J. (1968). De la carbonatation du béton. *Publication du CERILH*, (638).
- V.G., P., C.G., V., & N, F. M. (1991). Experimental investigation and mathematical modeling of the concrete carbonation problem. *Chem. Eng. Sci.*, 1333-1338.
- V.L TA, S. B., KIESSÉ, T. S., & VENTURA, A. (2016). A new meta-model to calculate carbonation front depth within concrete structures. *Construction and Building Materials*, 172-181.
- WANG, X. (2009). Nonparametric estimation of the shape function in a Gamma process for degradation data. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, 37(1), 102-118.
- WELLNER, J. A., & ZHAN, Y. (1997). A hybrid algorithm for computation of the nonparametric maximum likelihood estimator from censored data. *Journal of the American Statistical Association*, 92(439), 945-959.
- WU, S., & SCARE, P. (2015). Decline and repair, and covariate effects. *European Journal of Operational Research*, 244(1), 219-226. <https://doi.org/https://doi.org/10.1016/j.ejor.2015.01.041>
- ZHANG, L., & GLASSER, F. (2005). Investigation of the microstructure and carbonation of CSA-based concretes removed from services. *Cement and Concrete Research*.