

Développement d'un outil d'interpolation inter-régimes d'une turbine hydroélectrique

par

QUANG HUNG PHAM

THÈSE PAR ARTICLES PRÉSENTÉE EN COTUTELLE À L'ÉCOLE DE
TECHNOLOGIE SUPÉRIEURE ET À INSA DE LYON
COMME EXIGENCE PARTIELLE À L'OBTENTION
DU DOCTORAT EN GÉNIE
Ph. D.

MONTREAL, LE 10 JANVIER 2022

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC



Quang Hung Pham, 2021



Cette licence Creative Commons signifie qu'il est permis de diffuser, d'imprimer ou de sauvegarder sur un autre support une partie ou la totalité de cette œuvre à condition de mentionner l'auteur, que ces utilisations soient faites à des fins non commerciales et que le contenu de l'œuvre n'ait pas été modifié.

PRÉSENTATION DU JURY

CETTE THÈSE A ÉTÉ ÉVALUÉE

PAR UN JURY COMPOSÉ DE :

M. Antoine S. Tahan, Directeur de thèse
Département de génie mécanique
École de technologie supérieure

M. Jérôme Antoni, Directeur de thèse
Laboratoire Vibrations Acoustique
INSA de Lyon

M. Thomas Dupont, président du jury
Département de génie mécanique
École de technologie supérieure

M. Eric Chatelet, membre de jury
Laboratoire de Mécanique des Contacts et des Structures
INSA de Lyon

Mme Nadine Martin, examinatrice externe indépendante
Laboratoire GIPSA-lab
CNRS Grenoble

Mme Mitra Fouladirad, examinatrice externe indépendante
Laboratoire M2P2 AMU,
École Centrale Marseille

M. Imad Tawfiq, examinateur externe indépendant
Laboratoire Quartz
Institut Supérieur de Mécanique Paris

IV

M. Xavier Chimentin, examinateur externe indépendant
Institut de Thermique, Mécanique, et Matériaux,
Université de Reims Champagne-Ardenne

M. Martin Gagnon, examinateur invité
Institut de recherche d'Hydro-Québec (IREQ)

Mme Christine Monette, examinatrice invitée
Andritz Hydro Canada Inc

ELLE A FAIT L'OBJET D'UNE SOUTENANCE DEVANT JURY ET PUBLIC

LE 14 DÉCEMBRE 2021

À L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

REMERCIEMENTS

Je souhaite particulièrement remercier mes deux directeurs de thèse, les professeurs Antoine Tahan (professeur à ÉTS Montréal), et Jérôme Antoni (professeur à INSA de Lyon) pour leur confiance et pour le temps qu'ils ont consacré pour le projet. Leurs expertises ainsi que leurs encouragements ont joué un rôle important pour mener à bien ma thèse. De plus, étant une thèse en cotutelle, mon temps de travail au Canada ou en France s'est confortablement passé grâce à leur accueil.

Je tiens aussi à remercier chaleureusement monsieur Gagnon Martin (chercheur à IREQ) et madame Christine Monette (Leader technique du groupe analyses mécaniques à Andritz Hydro Canada). Leurs expériences industrielles et leurs expertises techniques dans le domaine ont grandement contribué à la réussite de mon projet.

Je remercie les organismes MITACS, Hydro-Québec et Andritz Hydro Canada pour leurs soutiens financiers et leurs confiances.

Enfin, je souhaite adresser mes remerciements et ma reconnaissance à mes collègues et organismes (ÉTS et INSA) pour leurs soutiens.

Développement d'un outil d'interpolation inter-régimes d'une turbine hydroélectrique

Quang Hung PHAM

RÉSUMÉ

L'introduction récente d'autres sources d'énergie dans le réseau électrique (éolienne, solaire, etc.) a récemment fait en sorte que les centrales hydroélectriques sont amenées à opérer la turbine hydraulique dans des régimes de fonctionnement qui s'écartent du meilleur point de rendement afin de régulariser et stabiliser le réseau. Cette expansion de la plage d'opération augmente le risque de fatigue de la roue principalement à cause des grandes fluctuations des contraintes. Pour estimer le dommage, il est impératif de connaître les niveaux de contraintes pour tous les différents régimes de fonctionnement. Or, cette information n'est pas disponible, d'où la nécessité d'avoir un modèle d'interpolation qui permettra d'estimer, avec une précision suffisante, les niveaux de contraintes cycliques pour un régime de fonctionnement donné. Pour l'évaluation de la fatigue de la roue, les mesures expérimentales de contraintes par jauges sont les informations essentielles. Cependant, à cause du coût élevé et des exigences temporelles et logistiques, ces mesures expérimentales ne peuvent couvrir que quelques points sur l'étendue de la plage possible d'opérations de la turbine. Dans cette thèse, on s'intéresse au développement d'outils de prédiction de contraintes pouvant couvrir tous les régimes de fonctionnement. Un outil mathématique basé sur des méthodes de krigeage est développé, ce qui permet d'interpoler et de reconstruire le signal de contraintes dans des régimes non mesurés à partir de ceux mesurés. Les signaux expérimentaux sont prétraités et transformés vers des formes différentes (par exemple, des processus stochastiques ou des composantes cachées dans le signal...) pour les adapter par la suite à chaque stratégie d'interpolation. Les outils développés sont validés sur des études de cas tirés d'applications industrielles. Cette thèse contribue à répondre à un besoin actuel du concepteur et de l'exploitant des turbines hydrauliques. Le but, est d'améliorer l'estimation de la fiabilité en fatigue de la roue, de réduire les coûts d'une campagne de mesures expérimentales ou de simulations numériques, et de mieux estimer la durée de vie résiduelle.

Mots-clés : Turbine Hydroélectrique, Contraintes/Déformations, Fatigue de la roue, Interpolation, Krigeage, Cyclostationnarités, Comptage Rainflow, Simulation Stochastique.

Development of an interpolation tool over operating conditions of the hydroelectric turbine

Quang Hung PHAM

ABSTRACT

Due to the change of the energy sector (wind, solar...) in recent years, the hydroelectric power plant must expand the operating range which is far from the Best Efficiency Point condition in order to regularize and stabilize the electrical grid. These expanded operating conditions increase the fatigue damage on the runner due to higher dynamic strain/stress fluctuation. To estimate the damage, it is imperative to have the stress levels over all the different operating conditions. However, this information is not available, hence an interpolation model, which allows estimations with sufficient accuracy of cyclic stress levels at a given operating condition, is needed. In analyzing of runner fatigue, strain measurements by gauges plays an essential role. However, due to the high cost and time constraint, these experimental measurements can only cover certain points, with finite number, on the possible range of turbine operating conditions. Therefore, this thesis concentrates on the development of prediction tools for the strain/stress covering every possible operating condition. A mathematical process based on the kriging methods is developed, which allows the interpolations and then the reconstructions of stress/strain signals at non-measured operating conditions using the measured ones. The experimental signal is pre-processed and transformed into different forms (e.g., stochastic process or components involved in the signal...) for adapting to each interpolation strategy. The developed interpolations are validated by cases studies from the industrial application. This thesis contributes to response to an actual industry need of the hydraulic turbine designer and operator for the reduction of experimental measurement or numerical simulation costs, and for a better estimation of residual lifetime.

Keywords: Hydroelectric Turbine, Stress/Strain, Runner Fatigue, Interpolation, Kriging, Cyclostationarities, Rainflow-Counting, Stochastic Simulation.

TABLE DES MATIÈRES

	Page
INTRODUCTION	1
CHAPITRE 1 REVUE DE LA LITTÉRATURE.....	7
1.1 Fatigue de la roue dans les régimes de fonctionnement	7
1.2 Exploitation des connaissances des contraintes/déformations de la roue	14
1.2.1 Mesures expérimentales dans les aubes.....	15
1.2.2 Simulation numérique par CFD et EF	17
1.2.3 Traitement du signal de contrainte/déformation sur l'aube	18
1.3 Processus d'interpolation	21
1.3.1 Prétraitement du signal	22
1.3.1.1 Modélisation par densité de probabilité.....	22
1.3.1.2 Modélisation dans le domaine fréquentiel	28
1.3.1.3 Comptage Rainflow	32
1.3.1.4 Cyclostationnarité	33
1.3.1.5 Conclusion sur les approches de modélisation	37
1.3.2 Méthode d'interpolation.....	38
1.3.2.1 Méthodes déterministes	39
1.3.2.2 Krigeage.....	42
1.3.2.3 Conclusion	47
1.3.3 Méthodes de validation	48
1.3.3.1 Mesures de l'erreur de prédiction	48
1.3.3.2 Validation croisée	50
1.3.3.3 Métrique de surface (Area Metric)	51
1.3.3.4 Conclusion	53
CHAPITRE 2 OBJECTIFS, PLAN ET CONTRIBUTIONS.....	55
2.1 Bilan de l'originalité	55
2.2 Objectifs de recherche.....	56
2.3 Plan de la thèse.....	56
2.4 Contributions scientifiques et industrielles	58
CHAPITRE 3 RAINFLOW-COUNTING MATRIX INTERPOLATION OVER DIFFERENT OPERATING CONDITIONS FOR HYDROELECTRIC TURBINE FATIGUE ASSESSMENT.....	61
3.1 Contribution à la thèse	61
3.2 Résumé de l'article en français	62
3.3 Abstract.....	63
3.4 Introduction.....	63
3.5 Proposed Methodology	67
3.5.1 Asymmetric Rainflow Counting Matrices	67
3.5.2 Kriging Interpolation Method	68

3.5.3	Simulation Reconstruction.....	73
3.6	Case Study	75
3.7	Results and Discussions.....	79
3.8	Conclusions.....	81

CHAPITRE 4 PREDICTION OF HYDROELECTRIC TURBINE RUNNER STRAIN SINGAL VIA CYCLOSTATIONARY DECOMPOSITION AND KRIGING INTERPOLATION.....

4.1	Contribution à la thèse	85
4.2	Résumé de l'article en français	86
4.3	Abstract.....	87
4.4	Introduction.....	87
4.5	Interpolation Methodology	91
4.5.1	Kriging Interpolation method	91
4.5.2	Periodic Component.....	92
4.5.3	Vortex Rope Component	95
4.5.4	Residue Component.....	97
4.5.5	Complete Interpolation Methodology.....	99
4.6	Case Study	99
4.7	Results and Discussion	103
4.7.1	Interpolation of Periodic Component.....	103
4.7.2	Interpolation of Vortex Rope Component	105
4.7.3	Interpolation of Residue.....	106
4.7.4	Complete Interpolated Signal Verification	106
4.8	Conclusion	108

CHAPITRE 5 SIMULATION OF NON-GAUSSIAN STOCHASTIC PROCESSES WITH PRESCRIBED RAINFLOW CYCLE COUNT USING SHORT-TIME FOURIER TRANSFORM.....

5.1	Contribution à la thèse	111
5.2	Résumé de l'article en français	113
5.3	Abstract.....	113
5.4	Introduction.....	114
5.5	Simulation Methodology	117
5.5.1	Signal Model.....	117
5.5.1.1	STFT decomposition.....	117
5.5.1.2	Intrinsic correlation.....	119
5.5.1.3	Nonlinear unit	121
5.5.2	Model identification.....	123
5.5.2.1	Identification of the correlation structure	123
5.5.2.2	Identification of the nonlinear unit	124
5.5.3	Weighted Area Metric for Rainflow Counting.....	126
5.6	Interpolation Methodology	129
5.7	Case Study	130
5.7.1	Simulation Results and Discussion.....	131

5.7.2	Interpolation results and Discussion	135
5.8	Conclusion	137
CONCLUSION		139
RECOMMANDATIONS ET PERSPECTIVES		143
ANNEXE I THE CROSS-VALIDATION PERFORMANCE FOR SEVERAL NEIGHBOURHOOD SETS		147
LISTE DE RÉFÉRENCES BIBLIOGRAPHIQUES.....		149

LISTE DES TABLEAUX

	Page
Tableau 1.1	Comparaison la capacité des approches de modélisation pour l'interpolation.....38
Tableau 1.2	Comparaison entre des méthodes d'interpolation présentées dans la revue.....47
Tableau 1.3	Avantages des différentes méthodes de validation53
Tableau 3.1	Parameters of the proposed interpolation process adapted for this case study78
Tableau 4.1	Input and Output sets of the interpolation process.....99
Tableau 4.2	Estimation error values of each component and complete signal104
Tableau 5.1	Set of parameters in the case study132
Tableau 5.2	Distance metrics calculated from three simulation sets for the case study133

LISTE DES FIGURES

	Page
Figure 0.1	Diagramme Kitagawa et Takahashi Tirée et adapté de Gagnon, Tahan, Bocher, et Thibault (2014).....3
Figure 0.2	Limitations liées aux mesures expérimentales sur l’aube et description de la problématique de recherche4
Figure 0.3	Outil d’interpolation inter-régimes5
Figure 1.1	(Gauche) Modèle simple des chargements de la turbine; (Droite) Combinaison HCF+LCF ; Tirée et adapté de (Gagnon et al., 2012)9
Figure 1.2	Zone de fonctionnements de la turbine Francis9
Figure 1.3	Exemple d’un signal de déformation de la roue dans la zone stable10
Figure 1.4	Exemple d’observation de torche dans l’aspirateur Tirée de Seidel et al. (2014)12
Figure 1.5	Exemple d’observation de torche dans l’aspirateur Tirée de (Ciocan & Iliescu, 2007).....13
Figure 1.6	Exemple de signal de déformation de la roue dans la zone de torche (gauche) et observation des fréquences de torche (droite).....13
Figure 1.7	Signal de déformation de la roue dans la zone ‘Stochastique’14
Figure 1.8	Un exemple de localisations critiques des jauges de déformation installées sur l’aube.....16
Figure 1.9	Un exemple de simulation numérique au régime de MAV Tirées de Morissette et al. (2016)18
Figure 1.10	Comparaison des valeurs de contraintes entre les interpolations par quelques méthodes de krigeage avec la référence ÉF (les cercles noirs représentent les points d’entrée de l’interpolation) Tirée de Ben Salah (2014).....19
Figure 1.11	Illustration du processus d’extrapolation utilisant les décompositions cyclostationnaires Tirée de Poirier et al. (2017)20

Figure 1.12	(Gauche) La technique des valeurs supérieures (VSS) avec des valeurs extrêmes extraites dans le rectangle rouge ; (Droite) La technique de bloc maxima (BM)	25
Figure 1.13	Une méthode pour choisir le seuil Tirée de Poirier (2013).....	26
Figure 1.14	Estimation par noyau pour lisser un histogramme.....	27
Figure 1.15	La PSD (sous la forme de spectre Waterfall) aide à visualiser et identifier les phénomènes physiques du signal dans le domaine fréquentiel	30
Figure 1.16	Un exemple de la matrice du comptage Rainflow qui est créée par l'algorithme de "Four-Points"	33
Figure 1.17	Décomposition du signal acoustique d'un moteur diesel	34
Figure 1.18	Processus d'extraction de CS2 qui est potentiellement utilisé pour séparer le phénomène de torche Tirée de Poirier et al. (2017)	37
Figure 1.19	Un exemple de semivariogramme expérimental et d'un modèle paramétrique de semivariogramme Tirée de Pham, Gagnon, Antoni, Tahan, et Monette (2019).....	45
Figure 1.20	Un exemple de différents modèles de semivariogramme spatio-temporel Tirée de Pebesma et Heuvelink (2016).....	46
Figure 1.21	Un exemple de l'algorithme de la technique de validation croisée	50
Figure 1.22	Exemples graphiques de la métrique de surface avec différents types de CDF	52
Figure 2.1	Structure du plan de la thèse	58
Figure 3.1	Measurement limits and interpolation/extrapolation solutions for the fatigue assessment of hydroelectric turbine runner	66
Figure 3.2	Four Point Counting algorithm applied for strain signal	68
Figure 3.3	Interpolation objective: new ARM histogram (at a given turbine operating condition) is interpolated from all experimental ARM histograms	69
Figure 3.4	Experimental semivariogram and example of semivariogram model - Sum metric (Semivariogram values in this figure are normalized)	72

Figure 3.5	Schema of the Kriging process in this paper applied for the studied variables	72
Figure 3.6	Percentage of the sample values outside the bin dimension is approximately equal to $y\%$	74
Figure 3.7	Flowchart displays the interpolation scheme	74
Figure 3.8	ARMs over different operating conditions for this case study and example of ARM histogram (the white bin corresponds to the null value)	77
Figure 3.9	Cross validation performance of sum-metric model (its parameters are presented in Tableau 3.1)	77
Figure 3.10	Comparison between experimental and interpolated ARM histogram at 55% OV	79
Figure 3.11	Experimental and generated cycles (one simulation) at 55 %OV	79
Figure 3.12	Comparison between experimental signal and one reconstructed signal at 55%OV; (Above) 10 seconds of signals; (Below) Histogram of whole signal	80
Figure 3.13	Rainflow strain spectrum between experimental values and simulations	82
Figure 4.1	Waterfall spectrum with respect to operating conditions displaying the different physical phenomena present in a runner strain signal	89
Figure 4.2	Proposed strategy to predict a missing strain signal	90
Figure 4.3	Flowchart of the proposed interpolations strategy	100
Figure 4.4	Example of component extractions displayed in time (left column) and frequency (right column) domains	101
Figure 4.5	Spectra of the vortex rope phenomenon (normalized) in a runner strain (left) and in a shaft bending (right)	102
Figure 4.6	Extracted VRC of runner strain and shaft bending at the same operating condition	102
Figure 4.7	Interpolation results for periodic components at 20Vo, 40Vo, 90 Vo	104
Figure 4.8	Interpolation results for the VRC (including semivariograms)	105

Figure 4.9	Kriging interpolation (middle) of the RFC (left) and the coefficient of variations (right)107
Figure 4.10	Simulation of residue at 20Vo, 40Vo, 90Vo (left column: experimental RFCs, middle column: simulated RFCs, right column: simulated rainflow spectra).....108
Figure 4.11	Comparison of interpolation result of complete strain signal displayed in the time (left column) and frequency (right column) domains109
Figure 4.12	Rainflow Strain Spectra of 500 simulations and experimental value for full signals at 20Vo, 40Vo, 90Vo110
Figure 5.1	Weighted area metric applied to two Rainflow Counting Counts127
Figure 5.2	Flowchart of the proposed simulation process.....128
Figure 5.3	(Left) Experimental STFT and (right) STFT of a simulated signal.....132
Figure 5.4	Comparisons of the generation performance of the spectral-based simulation (left column), the iterative simulation (middle column) and the proposed simulation (right column) according to the rainflow cycle counts (a, b, c), one of generated signal (d, e, f), their PDF (g, h, i) and their PSD (j, k, l).....134
Figure 5.5	Interpolation sets of PSD at 20Vo (a), 45Vo (b), 70Vo (c), and interpolation of α_0 (d)135
Figure 5.6	RCC and relative error for maximum range of simulations set using interpolated parameters at 20Vo (a, b), 45Vo (c, d) and 70Vo (e, f)136
Figure-A I-1	Cross-validation performances for several neighbourhood sets147

LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES

AR	<i>AutoRegressive</i>
ARM	<i>Asymmetric Rainflow Matrix</i>
BEP	<i>Best Efficient Point</i>
BGD	<i>Bivariate Gaussian Distribution</i>
BM	Technique de Bloc Maxima
CDF	Fonction de répartition (<i>Cumulative Distribution Function</i>)
CFD	Mécaniques des Fluides Numériques (<i>Computational Fluid Dynamics</i>)
CK	CoKrigage
CL	Coefficients de Lissage
CS	Cyclostationnarité ou Cyclostatinnaire
CSS	<i>Conditional Sequential Simulation</i>
DC	<i>Decreasing Cycle</i>
DPL	<i>Deep Part-Load</i>
EAM (MAE)	Erreur Absolue Moyenne (<i>Mean Absolute Error</i>)
EPN	Estimation Par Noyau
ER	Erreur Relative
ES	Semivariogramme expérimental (<i>Experimental Semivariogram</i>)
FEA	<i>Finite Element Analysis</i>
FPC	<i>Four-Points Counting</i>
GE	Groupe d'Entraînement
GEV	Distribution d'extremum généralisée
GPD	Distribution de Pareto généralisée

GV	Groupe de Validation
HCF	High Cycle Fatigue
IC	<i>Increasing Cycle</i>
i.i.d	Indépendant et identiquement distribué
IPV	Interpolation par Voisins
IREQ	Institut de Recherche d'Hydro-Québec
IRS	Interaction entre Rotor et Stator
IS	Itakura-Saito distance
IVN	Interpolation par Voisins Naturels
JS	Jensen-Shannon divergence
KI	Krigeage d'Indicatrice
KO	Krigeage Ordinaire
KS	Krigeage Simple
KST (STK)	Krigeage Spatio-Temporel (<i>Spatio-Temporal Kriging</i>)
KU	Krigeage Universel
LCF	Low Cycle Fatigue
LMCR	Modèle de co-régionalisation linéaire (<i>Linear Model CoRegionalization</i>)
MAA	Méthodes d'Apprentissage Automatique
MAV (SNL)	Marche à Vide (<i>Speed No-Load</i>)
MD	Méthodes Déterministes
MG	Méthodes Géostatistiques
MISE	Erreur quadratique moyenne intégrée (Mean Integrated Squared Error)
MM	Modèle de Mélange

MMG	Modèle de Mélange Gaussien
MNP	Modèle Non Paramétrique
MP	Modèle Paramétrique
MRF	Matrice de Rainflow
MS (AM)	Métrique de Surface (Area Metric)
EQM (MSE)	Erreur quadratique moyenne (Mean Squared Error)
OV	Ouverture de Valves de directrices
PID	Pondération inverse a la distance
PSD	Densité Spectrale de Puissance (<i>Power Spectral Density</i>)
RCC	<i>Rainflow-Counting Count</i>
REQM (RMSE)	Racine de l'Erreur Quadratique Moyenne (<i>Root Mean Square Error</i>)
RF	Régime de Fonctionnement
RFC	<i>Rainflow Cycle Counting or Rainflow Cycle Count</i>
RP	Régime Permanent
SAS	<i>Scale Adapted Simulation</i>
SPM	Spline en Plaque Mince
SV	SemiVariogramme (<i>Semivariogram</i>)
TF	Transformée de Fourier
TFCT (STFT)	Transformée de Fourier à Court Terme (<i>Short Time Fourier Transform</i>)
TFD	Transformée de Fourier Discrète
TVE	Théorie des Valeurs Extrêmes
VC	Validation Croisée

VE	Valeurs Extrêmes
VRC	Composante de phénomène de torche (<i>Vortex Rope Component</i>)
VSS	Technique des Valeurs Supérieurs à un Seuil

LISTE DES SYMBOLES

K	Fonction de noyau
H	Matrice des coefficients de lissage
B	Distribution des cycles dans la matrice de comptage Rainflow
p_i	Positions de 'bin' i dans la matrice de comptage Rainflow
o_j	Un régime de fonctionnement de la turbine
i_a	Indice de l'angle de rotation a (<i>Angular Index</i>)
γ	Semivariogramme
$v(.)$	Ensemble des paires de points séparés par une incrément donnée
h	Incrément spatial dans la matrice de Rainflow
u	Incrément entre deux niveaux d'ouverture de vanne
v	Incrément temporel/angularaire
$N(.)$	Nombre de paires de points dans $\vartheta(.)$
$L(.)$	Fonction de Lagrange
n_*	Nombre des régimes de fonctionnement dans l'entrée de l'interpolation
μ	Coefficient de Lagrange
κ	Correction de l'anisotropie
f_o	Fréquence synchrone de rotation de la roue dans les régimes permanents
ϑ_i	Ensemble de positions au voisin de position p_i dans la matrice de Rainflow
λ_*	Pondération de régression du Krigage
α_*	Pondération de régression pour la covariable du CoKrigage
V_o	Niveau d'ouverture de vanne
σ_e^2	Variance de l'erreur du Krigage

$P(.)$	Composantes périodiques extraites
$V(.)$	Phénomène de torche extraites
R_s	Fonction d'autocorrélation du signal s
S_{ik}	TFCT du signal avec indice temporel i et indice fréquentiel k
E_{ik}	Distribution d'énergie dans le temps
w	Fonction de la fenêtre
N_w	Longue de la fenêtre w
P_s	Densité spectrale de puissance du signal s
a_p	Coefficients du modèle d'autorégressive AR
R_{cc}	Fonction de comptage Rainflow
$\mathbb{E}\{*\}$	Opérateur de l'espérance
λ	Coefficient de la transformation Box-Cox

INTRODUCTION

Au Québec, l'hydroélectricité occupe une partie importante dans l'économie en général et est prépondérante dans le secteur de l'électricité en particulier. En utilisant l'eau qui est une source d'énergie propre et renouvelable, l'hydroélectricité possède un rendement énergétique très performant, et elle est, de loin, plus écologique que d'autres formes d'énergies (fossile ou nucléaire). Le développement de l'hydroélectricité est directement lié à la turbine hydraulique. Cette turbine est une machine rotative qui possède, structurellement parlant, des comportements dynamiques très complexes. En effet, la structure de la turbine hydroélectrique et son mode de fonctionnement créent un très grand nombre de cycles des chargements avec de grandes amplitudes de sollicitations qui entraînent des risques de défaillance par fatigue (surtout sur la roue). En effet, la corrosion par cavitation et la fissuration par fatigue sont les deux principaux modes de défaillance de ce type d'équipement. X. Liu, Luo, et Wang (2016) ont montré que le dommage par fatigue peut causer plusieurs conséquences graves (pertes des énergies/économies, danger pour des opérateurs à la centrale hydroélectrique). Par conséquent, il est admis que l'étude de la fatigue est un aspect essentiel durant la conception (dimensionnement) et durant le fonctionnement de la turbine tout le long de son cycle de vie qui peut s'étendre sur quelques décennies.

Les deux partenaires industriels de ce sujet de thèse, Andritz Hydro Canada Inc et L'Institut de recherche d'Hydro-Québec (IREQ) sont respectivement des leaders mondiaux dans la conception/fabrication et l'exploitation des turbines hydrauliques. Leur intérêt commun est l'amélioration de la fiabilité en fatigue. Les deux partenaires collaborent depuis environ une quinzaine d'années à travers des projets de recherche, spécialement depuis que les opérations de la turbine ont été élargies à cause des changements dans le mode de gestion des réseaux électriques et l'ouverture du marché de l'énergie. Ces changements ont eu pour conséquences que la turbine doit fonctionner dans des conditions qui s'écartent considérablement du meilleur point de rendement (BEP), ce qui entraîne un risque plus élevé de fissuration de

fatigue. Monette et al. (2016) ont mis en évidence la réduction de la durée de vie en fatigue de la roue et les coûts de réparation élevés causés par cette plage de fonctionnement élargie.

Problématique de recherche

Comme mentionnés précédemment, deux principaux modes de défaillances sont présents dans la turbine hydraulique : l'érosion par cavitation et la fissuration par fatigue. Notre projet s'attarde à la problématique de l'évaluation de la fatigue, par conséquent à l'estimation des amplitudes des contraintes mécaniques pour chaque point d'opération de la turbine. D'un point de vue économique et dans le but de maximiser les revenus générés, les turbines fonctionnent généralement à leurs limites de spécification. Cela entraîne les apparitions des comportements anormaux ou des défaillances non désirées. Chaque défaillance par fatigue qui ne peut pas être prédite, la plus précisément et le plus tôt possible, peut donc causer des dommages importants (coût de réparation, temps de travail, perte d'énergie, etc.). Par conséquent, la connaissance de la fiabilité résiduelle en fatigue de la roue d'une turbine hydraulique est essentielle pour planifier les périodes d'inspection, réparation et réfection.

Un modèle de fiabilité de la roue a été développé par Gagnon, Tahan, Bocher, et Thibault (2013). Ils ont utilisé le diagramme de Kitagawa et Takahashi (avec la correction de El Haddad pour la petite fissure, Figure 0.1) afin de délimiter une zone de défaillance en fatigue dans le cas de la roue d'une turbine hydraulique. Ce modèle a montré que la fiabilité sur la fatigue dépend de plusieurs paramètres comme les propriétés de matériau de la roue, la taille des fissures (défauts) à un instant quelconque et les niveaux des sollicitations sur la roue (les contraintes maximales, les fluctuations des contraintes...). Du côté du domaine de l'endommagement, la taille initiale (à l'instant de mesure) de défaut et les cycles de chargements influent de façon importants sur la vitesse de propagation des fissures. Dans le cas de la turbine, en raison de la conséquence catastrophique, nous souhaitons au mieux éviter l'augmentation de la vitesse de propagation (même quand les fissures sont petites) par les contraintes cycliques d'un régime stationnaire. Cela exige des connaissances fiables et précises sur les cycles des chargements appliqués sur la roue afin de prédire l'espérance de vie de la roue. Pourtant, à cause des comportements complexes dans différentes conditions

d'opération, la récupération (mesure, estimation, prédiction ...) de ces connaissances des sollicitations sur la roue d'une turbine hydroélectrique n'est pas une tâche facile. Elle présente un défi pour les chercheurs.

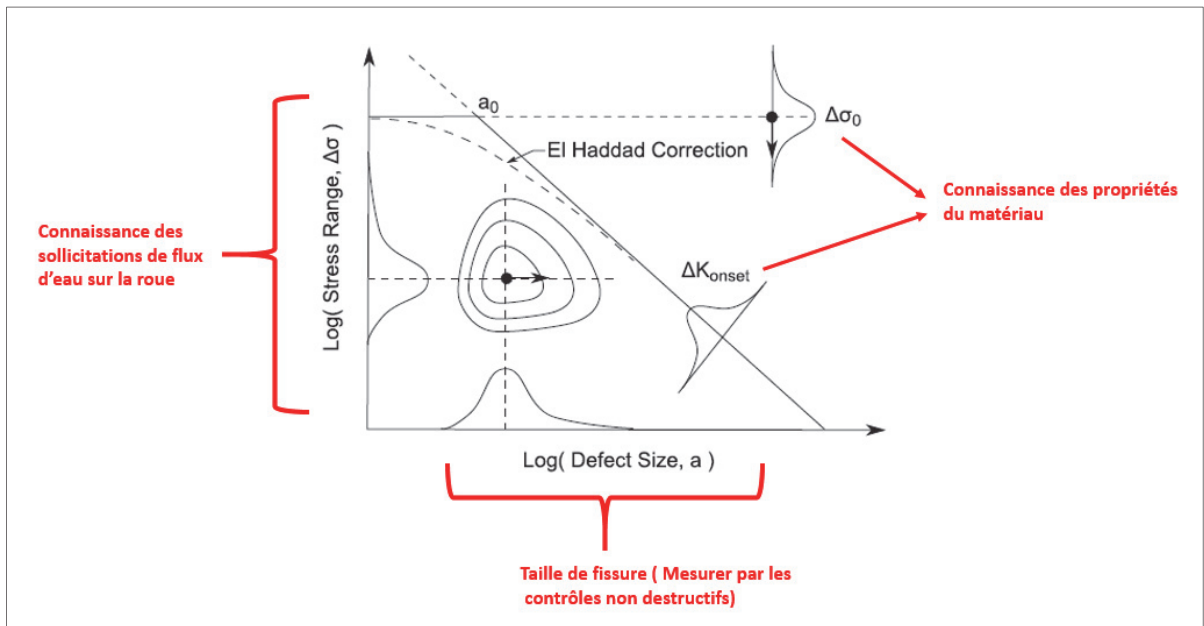


Figure 0.1 Diagramme Kitagawa et Takahashi
Tirée et adapté de Gagnon, Tahan, Bocher, et Thibault (2014)

Une des informations qui demeurent fiables pour connaître les niveaux de contraintes dans la turbine est fournies par les données collectées expérimentalement *in situ* (à partir de capteurs installés sur les aubes de la roue). Les données expérimentales sont toujours un bon choix pour comprendre et analyser les comportements mécaniques. Elles sont par contre très limitées dans le cas de la turbine hydroélectrique. Aussi, le coût économique pour les mesures expérimentales à la centrale est important. La Figure 0.2 montre trois principales limitations (les détails se trouvent dans la revue de littérature) avec les mesures expérimentales de contraintes par jauges de déformation fixée sur les aubes :

- les mesures sont seulement disponibles pendant un temps court (quelques secondes à quelques minutes);
- les mesures sont seulement disponibles sur quelques positions critiques de la roue;
- les mesures ne peuvent pas couvrir tous les régimes d'opérations.

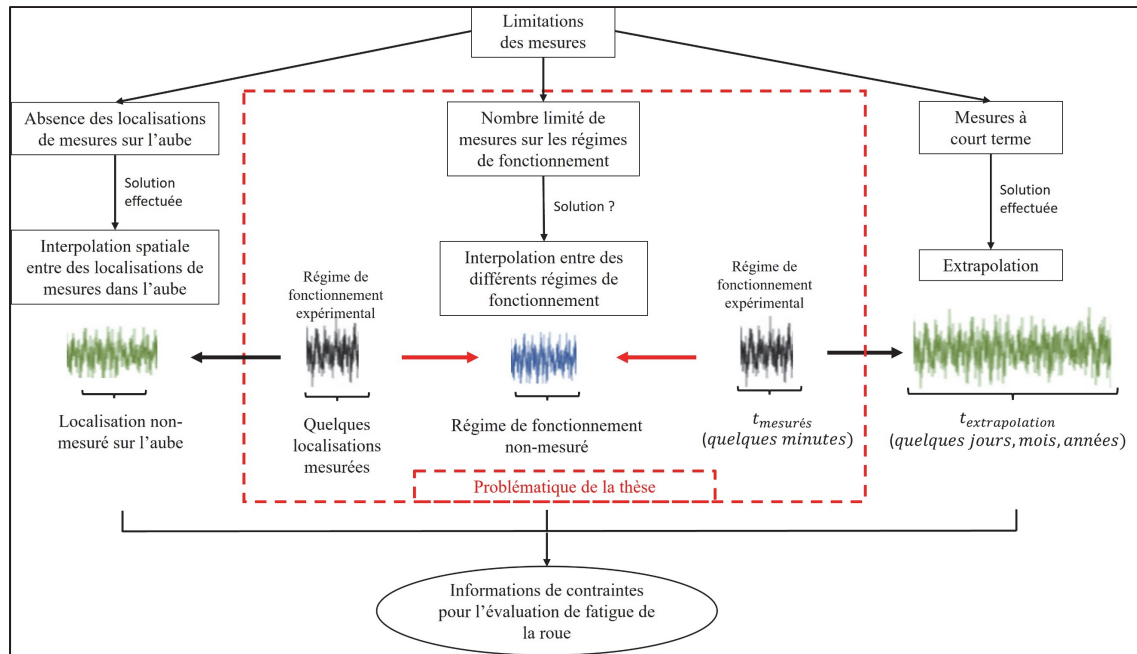


Figure 0.2 Limitations liées aux mesures expérimentales sur l'aube et description de la problématique de recherche

Certains chercheurs ont proposé des solutions pour les deux premières limitations (respectivement par l'interpolation spatiale (Ben Salah, 2014) et l'extrapolation temporelle (Poirier, Gagnon, Tahan, Coutu, & Chamberland-lauzon, 2017))

Cette thèse a pour but de proposer une solution pour résoudre le troisième problème : nombre réduit de mesures sur la plage de fonctionnement de la turbine. Des années récentes, à cause du changement sur la marché électrique par l'introduction des sources intermittentes (énergie éolienne et solaire), les centrales d'hydroélectriques ont élargi la plage d'opération de la turbine (Monette et al., 2016) pour régulariser le réseau d'électricité. Cet élargissement entraîne que la turbine doit parfois fonctionner dans des régimes non optimaux par rapport la conception de la turbine. En effet, un choix judicieux des conditions d'opération peut avoir une grande influence sur les amplitudes des contraintes dynamiques en régime stationnaire. L'identification des contraintes maximales de chaque régime d'opération ne peut être aboutie que si nous possédons une approximation fiable du comportement dynamique de l'ensemble du système pour tous les régimes de fonctionnement. Cependant, la difficulté imposée est

que chacun des différents régimes de fonctionnement (par exemple à différents niveaux d'ouverture, en charge ou à vide) possède un comportement typique et spécifique. Les profils de chargements sur la roue peuvent varier significativement entre les différents régimes à cause de la présence des fréquences naturelles, les interactions complexes entre l'écoulement fluide et la structure de la roue, ou encore par les phénomènes typiques à la turbine hydroélectrique comme les interactions entre rotor et stator (IRS), le phénomène de torche (Dörfler, Sick, & Coutu, 2012). Par conséquent, la récupération de ces informations de la roue devient un point essentiel et challengé dans l'industrie d'hydroélectricité. La Figure 0.3 montre l'objectif principal de cette thèse qui développe un outil d'interpolation inter-régimes pour recapturer les profils de contraintes/déformations dans toute la plage de fonctionnement d'une turbine hydroélectrique.

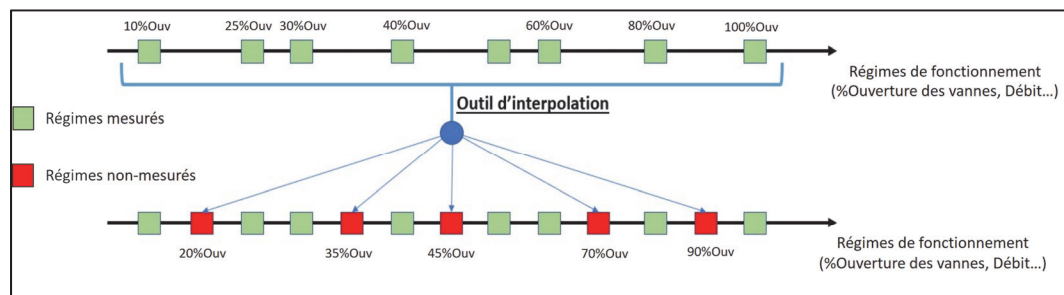


Figure 0.3 Outil d'interpolation inter-régimes

La question de recherche posée dans ce cas est : comment peut-on récupérer de la manière plus économique et plus précise les informations du comportement dynamique des différents régimes de fonctionnement d'une turbine hydroélectrique quand les mesures expérimentales ne sont pas toujours disponibles (ou sont limitées) ? C'est la question de base à laquelle notre présent projet tentera de répondre. Un modèle d'interpolation sur toute la plage de fonctionnement est notre idée pour maximiser notre connaissance du comportement dynamique de la turbine, et par conséquent, pour contribuer à l'amélioration de fiabilité en fatigue de la roue à l'avenir. Pour mieux clarifier l'originalité, les objectifs et les contributions de notre actuel projet, une revue de littérature est présentée dans le chapitre suivant.

Ce projet est une thèse en cotutelle entre l'École de technologie supérieure (QC, Canada) et l'INSA de Lyon, en collaborant avec Andritz Hydro Canada Inc et Institut de Recherche d'Hydro-Québec (IREQ). Le financement pour le doctorant pendant la thèse est fourni par le programme MITACs Accélération. Ce mémoire est présenté sous une forme de thèse par articles.

CHAPITRE 1

REVUE DE LA LITTÉRATURE

À notre connaissance, l'idée de développer une méthodologie permettant l'interpolation des signaux mesurés sur toute la plage de fonctionnement de la turbine est originale. Il n'y a pas encore, ou très peu, de publications scientifiques spécifiques liées directement à cette problématique. Dans ce chapitre, on se concentre sur une revue de littérature des domaines contigus à notre problématique. En particulier, la revue comprend trois thématiques :

- la fatigue de la roue dans les différents régimes de fonctionnement : dans cette section, on présente un état de l'art sur l'influence des comportements dynamiques du signal de contrainte/déformation sur la fatigue pour chaque zone des conditions de fonctionnement de la turbine. Cette revue permet de mettre en relief les difficultés et les défis scientifiques pour la prédiction précise du comportement pour une plage arbitraire de fonctionnement ;
- les méthodes pour l'exploitation des connaissances des contraintes/déformation : cette section présente les méthodes pour améliorer l'exploitation de comportement dynamique de contrainte, qui contribuent à la fiabilité en fatigue de la roue ;
- les méthodes d'interpolation : cette section présente une revue succincte des méthodes potentielles pour construire des principales étapes d'une interpolation adaptée à notre problématique.

1.1 Fatigue de la roue dans les régimes de fonctionnement

Il y a plusieurs types de turbines hydroélectriques (ex, Kaplan, Pelton, Francis, etc.), mais en raison de la configuration territoriale au Québec, notre recherche se concentre sur la turbine type Francis. C'est le type de turbine le plus utilisé au Québec. La turbine Francis est du type de turbine à *réaction* dans laquelle la roue hydraulique est tournée par l'impact de débit d'eau entrant dans la roue depuis la bêche spirale. La forme et le dimensionnement des aubes de la roue sont conçus et modélisés pour maximiser la transformation de l'énergie (le

rendement hydraulique peut atteindre 93%-95%). Dans une turbine Francis, le débit d'eau entrant dans la roue est contrôlé par les directrices. Ces directrices sont des aubes mobiles qui pourraient changer le niveau de l'ouverture, donc le débit. L'ouverture maximale correspond approximativement au régime de pleine charge.

Gagnon et al. (2013) ont défini le domaine de défaillance en fatigue en décomposant (simplifiant) les comportements des chargements de la turbine. En effet, le fonctionnement de la turbine peut être séparé en deux types : les régimes permanents (ce sont les régimes d'opération où la vitesse de rotation, le chargement... sont quasiment constants dans le temps) et les régimes transitoires (ex., le démarrage, l'arrêt). Le modèle le plus simple pour le chargement contient donc au moins une partie dite HCF « *High Cycle Fatigue* » qui représente le régime permanent, et une partie dite LCF « *Low Cycle Fatigue* » qui modélise les régimes transitoires. Les LCFs sont engendrées quand il y a des changements des régimes, ils contribuent à la propagation des fissures, ainsi à l'initiation des fissures. Gagnon et al. ont proposé que la limite de défaillance soit définie par le passage dans un diagramme de Kitagawa et Takahashi, vers une zone où les HCFs commencent à contribuer à la propagation. Ce qui est considéré comme une défaillance. Les simulations de propagation de fissures indiquent que la fissure se propage lentement si on passe un temps limité au régime permanent de pleine charge (Gagnon, Tahan, Bocher, & Thibault, 2012). C'est-à-dire, on considère qu'un chargement contient seulement la partie « LCF », car la partie HCF créée par le régime permanent est petite. Malheureusement, en réalité, le temps de fonctionnement de la turbine sous les régimes permanents n'est pas à négliger, surtout pour des régimes de fonctionnement où la turbine n'a pas été conçue pour l'origine. Par conséquent, quand la taille de fissure atteint à un point appelé « *HCF Onset* » (Gagnon et al., 2012; Gagnon et al., 2014), ou en considérant une taille critique de fissure, la combinaison de HCF et LCF va créer une propagation brutale et catastrophique (Figure 1.1). Par rapport à ces recherches, on peut affirmer l'importance de l'exploitation du comportement dynamique dans les régimes permanents qui contiennent les parties HCFs.

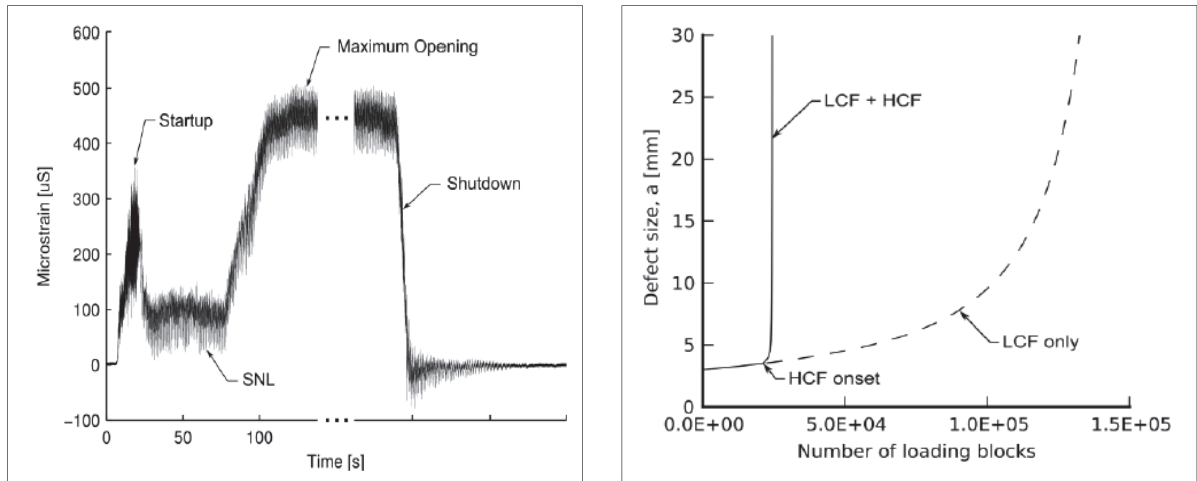


Figure 1.1 (Gauche) Modèle simple des chargements de la turbine; (Droite) Combinaison HCF+LCF ; Tirée et adapté de (Gagnon et al., 2012)

Dans ce projet, les Régimes Permanents (RP) sont donc les régimes ciblés pour l'interpolation. Les régimes transitoires sont uniques pour chaque cas. Par conséquent, l'interpolation de ces régimes n'est pas réaliste. Il y a deux indices importants qui peuvent être utilisés pour définir et localiser les régimes permanents : la hauteur de la chute [m] et le niveau de l'ouverture des aubes directrices [%] qui contrôle le débit. En réalité, les mesures expérimentales dans différentes chutes sont difficiles et sont onéreuses. De plus, d'après la disponibilité des données fournies et le besoin des partenaires industriels, les régimes permanents dans ce projet de recherche sont identifiés par différentes ouvertures des directrices pour une chute unique considéré comme constante tout le long de la période d'essais. Par conséquent, les parties suivantes de cette revue de littérature détailleront détailler plus clairement les comportements dynamiques de la contrainte/déformation de la roue pour chaque régime permanent et leur influence sur la fatigue.

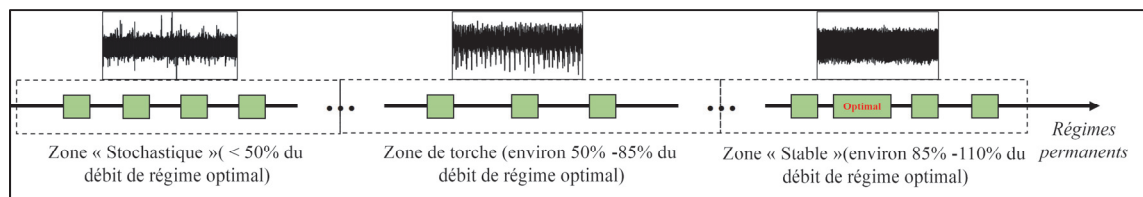


Figure 1.2 Zone de fonctionnements de la turbine Francis

Seidel, Mende, Hübner, Weber, et Otto (2014) ont publié une recherche sur les comportements dynamiques de contrainte de la roue dans les différents régimes de la turbine et leur influence sur la fatigue. De plus, deux revues détaillées de X. Liu et al. (2016) et Presas, Luo, Wang, et Guo (2019) ont aussi mentionné ces influences. D'après les comportements typiques, on pourrait séparer les régimes permanents par trois zones d'opération : zone 'stable', zone de 'torche' et zone 'stochastique' (Figure 1.2).

Zone 'Stable'

Cette zone contient des régimes autour du meilleur point d'efficacité de fonctionnement de la turbine (BEP en Anglais - *Best Efficiency Point*), et des régimes de grand chargement (ou proche pleine charge). Dans BEP, le débit d'eau entrant vers la roue est cohérent avec l'angle de l'aube et les lignes de débit suivent la géométrie des contours de la roue (Seidel et al., 2014). La localisation du BEP est unique pour chaque turbine. Le voisinage du BEP est normalement localisé où le débit est approximativement entre 85%-110% du débit nominal de BEP. Dans cette zone, la périodicité domine dans le comportement dynamique de la contrainte de la roue.

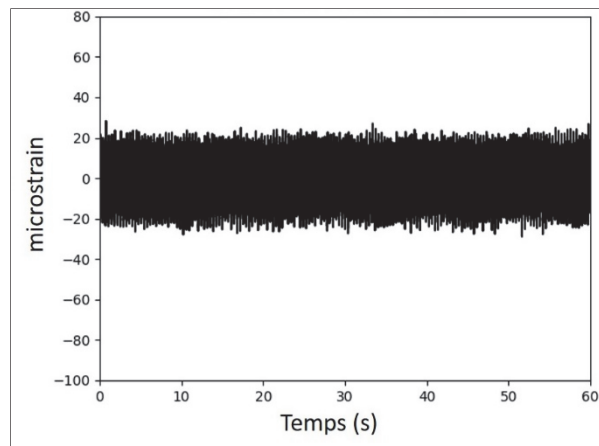


Figure 1.3 Exemple d'un signal de déformation de la roue dans la zone stable

En général, l'endommagement par fatigue créé dans cette zone est théoriquement plus petit que dans d'autres zones des régimes. La Figure 1.3 montre un exemple des déformations mesurées dans cette zone. L'amplitude du signal peut être grande, mais la fluctuation est souvent plus stable et réduite que dans d'autres zones.

Zone de torche

Pour le cas de la turbine à aubes fixes comme la turbine Francis, et dans certains régimes de chargement partiel, l'instabilité du tourbillon de débit d'eau à la sortie de la roue conduit à une torche hélicoïdale dans l'aspirateur (Figure 1.4 et Figure 1.5). Cette torche génère de grandes pulsations de pression dans l'aspirateur, et des vibrations dans d'autres composants (la roue, l'arbre) de la turbine. Les recherches sur ces phénomènes et sur les influences liées à la torche ont été développées depuis deux décennies via l'observation expérimentale ou les modèles mathématiques/numériques (Aschenbrenner, Otto, & Moser, 2006; Ciocan & Iliescu, 2007; Fanelli, 1989; Kuibin, Okulov, Susan-Resiga, & Muntean, 2010). La torche apparaît dans des conditions dites non optimales dans la plage de fonctionnement de la turbine. Cependant, comme mentionnée précédemment, l'expansion des régimes de fonctionnement dans les centrales nous oblige à étudier de plus en plus les influences du phénomène de torche. Ce dernier influe fortement sur les endommagements dans la turbine par fatigue (Seidel et al., 2014) et par la cavitation érosive (Aschenbrenner et al., 2006; Favrel et al., 2018).

Dans certains régimes de chargement partiel, la torche est hélicoïdalement formée et crée des fluctuations de pressions dans la turbine à basse fréquence¹. La Figure 1.6 montre un exemple du spectre d'un signal de déformation de la roue dans cette zone. Dans cet exemple, à côté de la fréquence synchrone f_0 de la roue², on constate la présence de deux autres fréquences dominantes qui se localisent à environ $0.3f_0$ et $0.7f_0$. Une fréquence correspond à

¹ La fréquence de torche est souvent inférieure à la fréquence synchrone.

² La fréquence synchrone correspondante à la rotation synchrone (avec une vitesse quasiment constante) de la roue dans un régime permanent.

la composante axiale de la pression observée dans l'aspirateur, et l'autre correspond à la composante radiale liée au décalage entre la rotation de la roue et la rotation de vortex dans l'aspirateur³ ($0.7f_0 = f_0 - 0.3f_0$). Le phénomène de torche crée une grande fluctuation représentée par de grands pics. Ces pics peuvent augmenter le risque de l'endommagement par fatigue. Ce phénomène apparaît souvent dans quelques conditions spécifiques et disparaît brutalement dans les conditions voisines. Généralement, pour les turbines du type Francis, on peut observer ce phénomène quand le débit est entre 50% - 85% du débit de BEP (Dörfler et al., 2012). En raison de ce comportement d'apparition et disparition, l'estimation par interpolation de ce type de régimes représente un défi de taille.



Figure 1.4 Exemple d'observation de
torche dans l'aspirateur
Tirée de Seidel et al. (2014)

³ Le signe moins indique dans ce cas que les deux rotations mentionnées sont dans des sens opposés.

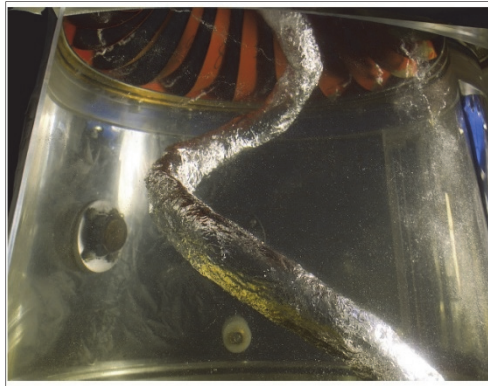


Figure 1.5 Exemple d'observation de torche dans l'aspirateur
Tirée de (Ciocan & Iliescu, 2007)

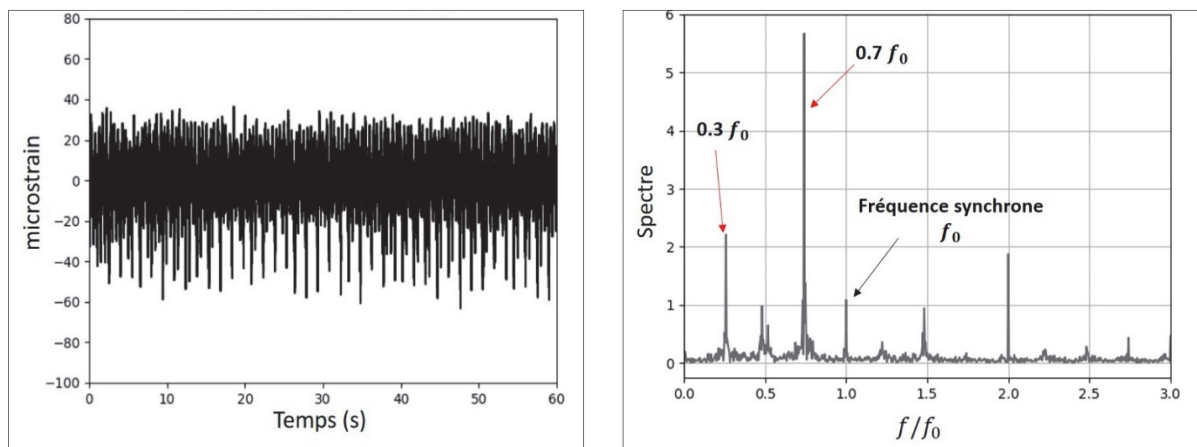


Figure 1.6 Exemple de signal de déformation de la roue dans la zone de torche (gauche) et observation des fréquences de torche (droite)

Zone 'Stochastique'

Cette zone contient le régime de Marche à Vide (MAV) et les régimes de charge partielle dans lesquels le débit est inférieur à 50% de débit à BEP. Le régime MAV est un régime où la turbine fonctionne à une vitesse synchrone sans production d'électricité (Morissette et al., 2016; Nennemann et al., 2014). Dans ce régime, il y a toujours un écoulement d'eau entrant à la roue tandis qu'aucune production d'électricité ne se produit. Cela entraîne donc qu'une quantité importante d'énergie doit être dissipée dans la roue de turbine (Morissette et al., 2016). Le mot 'stochastique' veut dire que dans cette zone, les composantes stochastiques

dominant dans le signal de déformation de la roue, et elles possèdent un comportement complexe. Le débit d'eau entrant à la roue dans ce régime a souvent des états instables et peuvent causer des grandes fluctuations dynamiques sur les aubes de la roue. Ces phénomènes ont des impacts importants sur la défaillance par fatigue de la roue.

À cause des changements dans le marché de l'électricité, les opérateurs doivent parfois opérer la turbine dans des régimes dans cette zone (Monette et al., 2016). Les flux d'eau instables et « agressifs » causent des limites (sur la position et le temps de mesure) dans les mesures expérimentales par capteurs. De plus, la grande partie aléatoire du signal de ce régime cache parfois la partie périodique du signal (c'est la partie typique du signal dans les machines tournantes comme la turbine). Cela rend la prédiction des contraintes plus difficile. Toutes les limitations mentionnées peuvent causer des difficultés pendant le processus de prédiction par interpolation.

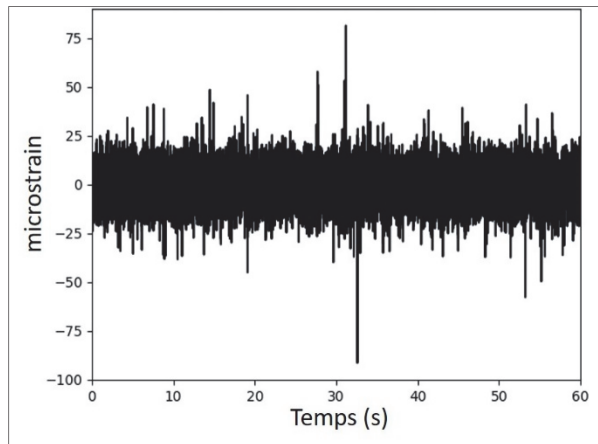


Figure 1.7 Signal de déformation de la roue dans la zone 'Stochastique'

1.2 Exploitation des connaissances des contraintes/déformations de la roue

Cette section permet de dresser un portrait de l'état de l'art sur la modélisation des contraintes mécaniques dans une roue d'une turbine. Cette revue nous aidera à sélectionner

les pistes potentielles et à mettre en évidence la contribution de notre projet pour apporter des solutions à la problématique soulevée.

1.2.1 Mesures expérimentales dans les aubes

L'idée est simple, les déformations sont récupérées par des jauges de déformation qui sont installées sur les aubes de la roue. La loi de Hook permet de lier cette information à l'état des contraintes mécaniques. Ces mesures sont une information importante à laquelle notre partenaire IREQ s'intéresse. Les données dans le cas d'étude de ce projet sont fournies par Hydro-Québec. Ce sont des mesures réalisées sur une centrale hydroélectrique à Québec.

Le nombre de ces mesures expérimentales n'est pas exhaustif à cause de la difficulté d'installation, et surtout à cause des coûts inhérents à de tels montages. Chaque campagne de mesure exige de grandes préparations des équipements, des ressources humaines, des stratégies de mesure afin de maximiser l'efficacité des mesures, car pour la turbine hydraulique, chaque arrêt (pour installer des capteurs, vérifier les mesures...) est incontestablement une perte de production et aussi cause des endommagements indésirables sur les aubes.

Les points des mesures sur l'aube sont localisés seulement dans quelques endroits considérés critiques, où les contraintes les plus grandes sont prévues. Un exemple de ces localisations est montré dans la Figure 1.8, correspondant aux données dans ce cas d'étude. Ces mesures sont souvent de courte durée à cause des conditions hydrauliques extrêmes (Poirier et al., 2017). Ces mesures courtes sont théoriquement insuffisantes pour l'évaluation de fatigue à long terme, qui a normalement besoin d'information durant quelques jours, semaines ou mois.

Actuellement, la plage des Régimes de Fonctionnement (RF) actifs est élargie en raison du changement de marché d'électricité, mais le pas entre des RFs mesurés reste relativement

grand⁴. Dans une centrale hydroélectrique, le plan des mesures expérimentales est optimisé pour pouvoir passer des RF nécessaires avec un coût minimum. Par exemple, dans les données disponibles, si 100% d'Ouverture de Valves de directrices (noté 100OV), le pas de RF mesuré est varié de 50V jusqu'à 100V. C'est-à-dire qu'il y a seulement environ une dizaine de régimes mesurés.

Il est clair que les limitations des mesures sur l'aube causent des difficultés pour comprendre le comportement dynamique de contrainte, et aussi pour évaluer la fatigue de la roue. Dans les sous-sections suivantes, des méthodes numériques pour remplacer ou améliorer ces mesures expérimentales sont discutées.

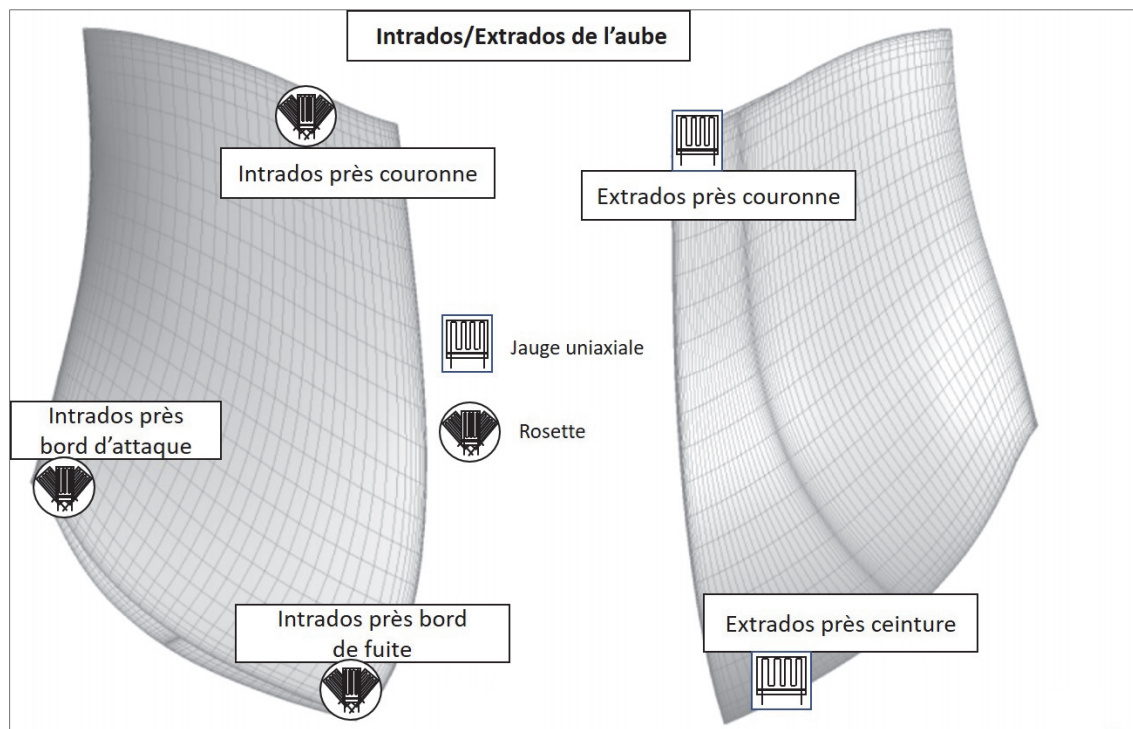


Figure 1.8 Un exemple de localisations critiques des jauges de déformation installées sur l'aube

⁴ Un pas plus petit peut considérablement augmenter le coût des mesures.

1.2.2 Simulation numérique par CFD et EF

Considérant que les informations expérimentales sont limitées, l'utilisation de simulations numériques par CFD et EF permet d'obtenir plusieurs informations avec un coût plus abordable que les mesures expérimentales. Ce type de simulation est souvent utilisé aujourd'hui lors de la phase de conception et d'exploitation de la turbine hydroélectrique.

Les simulations numériques peuvent donc fournir les informations sur le comportement de contraintes dans toutes les conditions de fonctionnement possibles. Plusieurs recherches ont proposé des approches pour estimer le niveau des contraintes sur l'aube en utilisant des simulations numériques développées (Bonnardot, 2004; Nennemann et al., 2014; Trivedi & Cervantes, 2017). Chaque année, des recherches qui utilisent les simulations CFD sont publiées dans les conférences annuelles du domaine. La recherche de Presas et al. (2019) a donné une revue pertinente sur ces simulations numériques. Pour les valeurs statiques de contrainte sur l'aube, le modèle de simulation semble être assez mature, il peut fournir des estimations très proches aux valeurs expérimentales (Presas et al., 2019; Seidel et al., 2014). Cependant, même si les modèles numériques sont de plus en plus performants, l'estimation de façon précise du niveau dynamique est encore problématique, spécialement dans les zones de fonctionnement qui subissent de grandes fluctuations, comme la zone stochastique ou la zone de torche. Dans ce cas, les simulations numériques ne peuvent pas correctement capturer la complexité du comportement dynamique. Un des modèles qui peuvent fournir de bonnes estimations est le modèle de turbulence SAS (*Scale Adapted Simulation*) (Mende, Weber, & Seidel, 2016; Morissette et al., 2016; Nennemann et al., 2014). Ce modèle permet de bien exprimer la nature stochastique et de montrer les tourbillons qui sont le principal responsable de la dynamique (Nennemann et al., 2014). La Figure 1.9 montre des exemples de simulations numériques dans la condition de MAV. En général, comparés au cas réel, les modèles numériques sont encore plus simples, par contre, ils limitent parfois la précision des résultats des simulations (ex, comme aux basses fréquences, voir Figure 1.9).

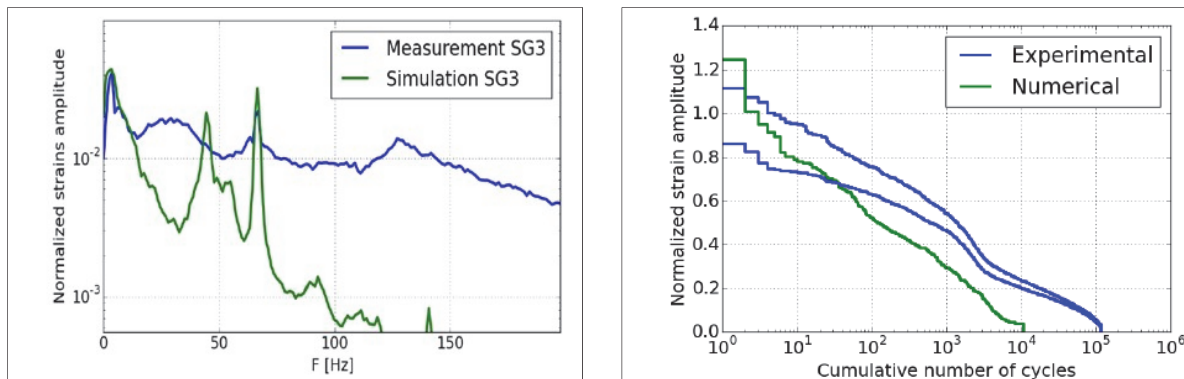


Figure 1.9 Un exemple de simulation numérique au régime de MAV
Tirées de Morissette et al. (2016)

Le temps nécessaire pour les simulations est court comparé au temps exigé pour une campagne de mesures expérimentales, mais c'est quand même un temps considérable de calcul. Le temps de simulation dépend de la complexité du modèle numérique (le type et la taille des maillages, les hypothèses imposées...), une simulation des valeurs dynamiques de contraintes à une condition de fonctionnement donnée peut durer 3 à 5 jours. De plus, un autre problème posé est celui de la calibration des modèles numériques. En d'autres termes, comment peut-on vérifier et valider la qualité des résultats obtenus par simulation ? Les valeurs expérimentales sont incontournables pour cette tâche essentielle.

Notre projet n'aborde pas le développement des simulations par CFD et EFs. Il emploie les techniques et les outils propres au traitement des signaux. Cependant, cette revue met en évidence le besoin de prédictions précises des contraintes sur l'aube, ce qui peut aider à réduire le temps des simulations et aussi peut contribuer à l'évaluation précise de la fatigue pendant la phase de conception de la turbine.

1.2.3 Traitement du signal de contrainte/déformation sur l'aube

À cause du manque de mesures, l'utilisation des méthodes de traitement de signal (TdS) est nécessaire pour améliorer la connaissance du comportement dynamique de la contrainte sur la roue. De plus, si ces méthodes de TdS sont suffisamment fiables, elles pourront aider à réduire les coûts des mesures expérimentales (et des simulations numériques) et ainsi à

contribuer à améliorer l'évaluation de la fatigue. Dans cette sous-revue, nous nous concentrons aux recherches qui ont résolu les problématiques des limitations de mesures et qui sont davantage appliquées pour estimer des comportements de la roue. De plus, par cette sous-revue, nous pourrions situer et positionner notre projet.

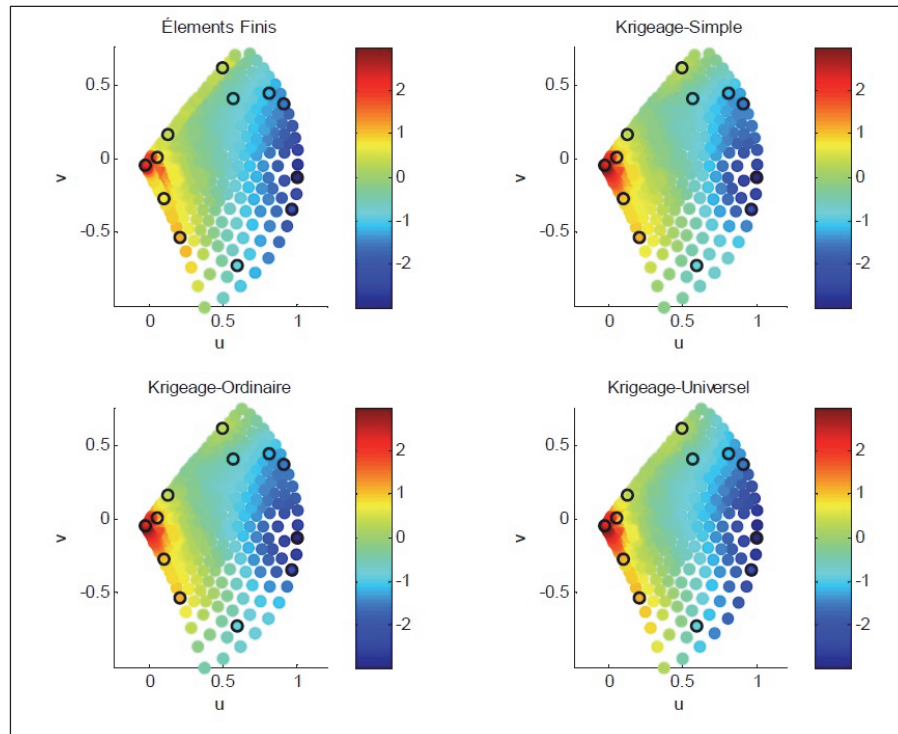


Figure 1.10 Comparaison des valeurs de contraintes entre les interpolations par quelques méthodes de krigage avec la référence ÉF
(les cercles noirs représentent les points d'entrée de l'interpolation)
Tirée de Ben Salah (2014)

Ben Salah (2014) a appliqué la méthode Krigage pour réaliser l'interpolation spatiale sur l'aube (Figure 1.10) à partir d'un nombre limité de mesures expérimentales. Ses résultats ont aussi contribué à estimer la propagation des incertitudes des mesures expérimentales sur l'aube. Il a aussi utilisé les simulations numériques par CFD et EF pour consolider la qualité de l'interpolation. Les simulations numériques n'étant pas capables de bien représenter le cas réel. Une hypothèse a été émise : elles sont suffisamment précises pour créer une carte de covariance spatiale. Ce processus de l'interpolation peut aider à résoudre la limitation des

points de mesure. Par ailleurs, d'après (Presas et al., 2019), l'utilisation de la fibre optique (*Optical Fiber*) pourrait être une technique future pour mesurer les contraintes sur l'aube. Comparé aux jauges de déformations, l'avantage de cette technique est qu'on pourrait mesurer plus de points en considérant une fibre comme un système de plusieurs capteurs. Si on pouvait appliquer au futur l'interpolation spatiale par krigeage à partir des mesures par fibre optique, ce serait une nette amélioration de la prédiction des contraintes avec un coût plus raisonnable que les mesures expérimentales par jauges classiques.

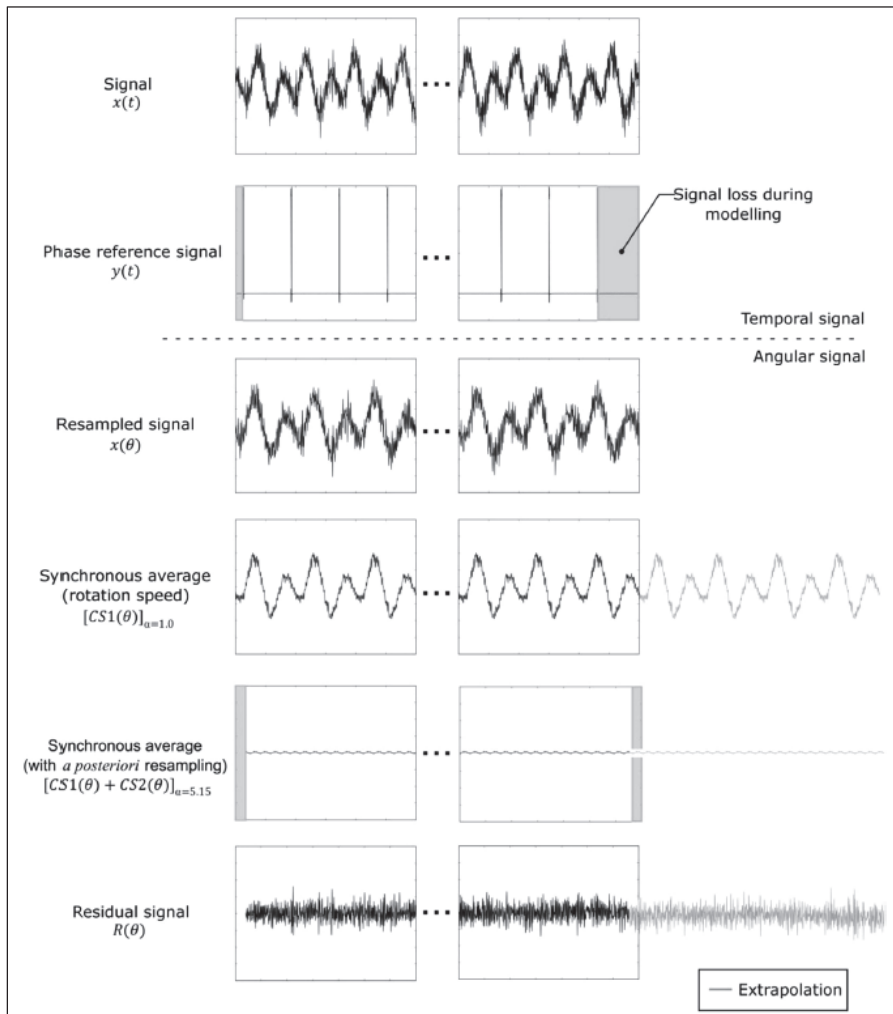


Figure 1.11 Illustration du processus d'extrapolation utilisant les décompositions cyclostationnaires
Tirée de Poirier et al. (2017)

Pour faire face au problème de limitation du temps de mesure, Poirier et al. (2017) ont proposé un processus d'extrapolation pour estimer des contraintes à long terme. Ils ont utilisé des décompositions cyclostationnaires pour séparer une mesure de période courte en plusieurs parties, et ensuite les ont extrapolées dans le temps. Plusieurs méthodes d'extrapolation pourraient être utiles pour prolonger une information à court terme. Par exemple, Pär Johannesson (2006) a proposé une technique utilisant le théorème des valeurs extrêmes pour extrapoler seulement des amplitudes excédant un seuil donné. Une extrapolation de la matrice de Rainflow (Pär Johannesson & Thomas, 2001; D. Socie, 2001) est aussi utile pour accéder à l'évaluation de la fatigue à long terme. Cependant, pour traiter un type de signaux complexes comme la mesure de contrainte sur l'aube hydroélectrique, l'utilisation de la cyclostationnarité est un point clé, comme démontré par Poirier (Figure 1.11).

Les recherches de Ben Salah (2014) et Poirier et al. (2017) sont détaillées, car elles ont directement proposé une solution aux deux premières limitations des mesures expérimentales mentionnées dans l'Introduction. De plus, elles ont inspiré notre projet. Cependant, à côté de ces recherches, d'autres méthodes de prédictions ont été proposées dans la littérature. On peut mentionner les recherches sur les modèles probabilistes de la fatigue de la roue (Gagnon et al., 2012; Gagnon et al., 2013), la méthode indirecte qui a utilisé une autre mesure (dite par mesure indirecte) dans une zone fixe de la turbine pour estimer le comportement dynamique de la roue (Diagne, Gagnon, & Tahan, 2016). Ou plus récemment, Presas et al. (2021) qui ont appliqué les réseaux de neurones pour estimer les contraintes dynamiques d'une turbine Francis.

1.3 Processus d'interpolation

Trois étapes sont nécessaires dans un processus d'interpolation. Par conséquent, la revue de littérature est présentée en trois sous-thématiques : le prétraitement d'un signal, la méthode d'interpolation et la technique de validation à employer pour évaluer la qualité du modèle de l'interpolation et l'erreur de l'estimation.

1.3.1 Prétraitement du signal

Un prétraitement du signal avant interpolation est une étape importante pour réussir une bonne interpolation. Cette étape consiste généralement en une "modélisation" du signal étudié, qui nous permet d'identifier la nature du modèle ou le processus entrée (*Input*) de l'interpolation. Le but est de simplifier l'interpolation en n'estimant seulement que des caractéristiques (temporelles, fréquentielles, cyclostationnaires...) intéressantes pour la modélisation des phénomènes rencontrés au lieu d'interpoler un signal complexe.

Basées sur la compréhension des phénomènes physiques qui composent le signal étudié et illustré dans les sections précédentes, quelques approches d'interpolation nous semblent potentielles : l'interpolation en se basant sur la conservation du spectre fréquentiel, l'interpolation de l'histogramme de comptage Rainflow, et l'interpolation en employant les outils de la cyclostationnarité.

1.3.1.1 Modélisation par densité de probabilité

La modélisation avec des fonctions de densité de probabilité est une technique populaire dans plusieurs domaines scientifiques. Cette technique permet de représenter (donc, interpoler) un modèle par une fonction de probabilité et par la suite, elle permet de régénérer le signal par un processus de l'échantillonnage. D'un point de vue statistique, on peut séparer ces approches en deux types: le modèle paramétrique (MP) qui utilise les fonctions de densité connues comme loi gaussienne, loi bêta, loi chi-deux... (Walpole, Myers, Myers, & Ye, 1993) et le modèle non paramétrique (MNP) qui ne relie pas à une distribution spécifique paramétrée. Un exemple de MNP est la méthode de l'estimation par noyau (Wand & Jones, 1994), ou plus simple, à partir d'un histogramme.

En pratique, un modèle de densité simple facilite l'interpolation, mais il ne peut pas bien décrire toutes les données dans un cas réel. Par exemple, l'utilisation d'une distribution paramétrique unique est simple pour représenter, approximativement, un signal mesuré sur la roue hydroélectrique, mais il ne traduira pas toute la richesse du contenu. D'autre part, l'utilisation d'un histogramme empirique (ou d'une fonction de répartition empirique), qui est

réalisée depuis une mesure unique, est réductrice et très risquée pour une estimation à long terme. Par conséquent, dans le cas spécifique de notre projet, que ce soit MP ou MNP, il faut trouver des approches qui soient suffisamment riches pour décrire le comportement réel (complexe). Dans cette revue, on mentionnera quelques approches qui sont potentielles pour modéliser le comportement des contraintes dans une roue et qui se basent aussi sur les fonctions de densité, comme : le modèle de mélange, les distributions des valeurs extrêmes et l'estimation par noyau.

Modèle de mélange

Un modèle de mélange (MM) permet de modéliser différentes sous-densités pour obtenir une densité globale du modèle. En général, un MM peut être écrit sous la forme d'une somme pondérée :

$$f(x_i) = \sum_{i=1}^I \pi_i f_i(x_i) \quad (1.1)$$

où π_i sont les coefficients de pondération avec $\sum_{i=1}^I \pi_i = 1$, f_i sont les composantes de sous-densités. Un des modèles populaires est le modèle de mélange gaussien (MMG) (Reynolds, 2009). Comparé à autres MM, le MMG est plus pratique en raison de sa simplicité et la facilité à estimer ses paramètres (par exemple, par la méthode de maximisation de la vraisemblance). Par ailleurs, dans certains cas spécifiques, quand la queue de la distribution est plus longue (à noter que c'est notre cas de signal, voir Figure 1.6 et Figure 1.7), un modèle de mélange des distributions de type Student semble être plus adéquat, le coefficient d'aplatissement (Kurtosis) étant plus élevé (Peel & McLachlan, 2000). Par contre, ceci se traduit aussi par l'ajout de paramètres à identifier (le nombre de degrés de liberté).

Du point de vue de l'interpolation, les paramètres d'un MM, les poids de pondération et les paramètres des sous-distributions sont les valeurs d'entrée de l'interpolation. Par conséquent, le choix du nombre de sous-distributions et l'estimation des paramètres clés sont les deux

étapes importantes. Par exemple, un trop petit nombre de sous-distributions ne peut pas bien décrire le signal, tandis qu'un grand nombre génère des difficultés de robustesse pour l'interpolation. Pour aider à choisir ce nombre et à estimer les paramètres des sous-distributions d'un MM, certaines méthodes bien connues comme le maximum de vraisemblance (Reynolds, 2009), et son implémentation par l'algorithme d'espérance-maximisation (Biernacki, Celeux, & Govaert, 2003; Foulley, 2002) peuvent être employées.

Distributions des valeurs extrêmes

Le signal de contrainte mesuré sur la roue hydroélectrique contient un grand nombre de cycles de grande amplitude responsable de l'augmentation du risque d'endommagement par fatigue. Par conséquent, au lieu d'analyser la densité de probabilité du signal complet, on analyse seulement la portion des grands pics (ou les valeurs extrêmes qui dépassent un seuil établi *a priori*). Les analyses des amplitudes maximales/minimales locales nous conduisent à la théorie des valeurs extrêmes (TVE) (Beirlant, Goegebeur, Segers, & Teugels, 2006; Cetin, Naess, & Härkegård, 2013; Pär Johannesson, 2006). La TVE est un sous-domaine important de la statistique, ses applications sont employées dans plusieurs domaines scientifiques et industriels, par exemple les modèles de l'intensité des vagues (*Ocean Wave Modeling*), la modélisation du vent (éolienne), la modélisation des contraintes mécaniques d'un écoulement (turbine hydroélectrique), ou l'analyse des sollicitations pour un tremblement de terre... (Coles, Bawa, Trenner, & Dorazio, 2001).

En général, la TVE permet de modéliser les maxima (ou minima) locaux d'un signal par une (ou des) distribution(s) comme la distribution de Pareto généralisée (GPD) ou des distributions d'extremum généralisées (GEV)⁵. Cependant, pour identifier les paramètres des distributions GPD ou GEV, il faut isoler d'abord les valeurs extrêmes. Deux techniques ont été proposées dans la littérature : la technique de bloc maxima (BM) et la technique des valeurs supérieures à un seuil (VSS, ou POT - *Peaks Over Threshold*). La Figure 1.12 montre

⁵ GEV est une famille de distributions qui comprend plusieurs lois comme Weibull, Gumbel...

ces deux techniques. La technique BM extrait les valeurs extrêmes (VE) en localisant le maximum (minimum) de chaque bloc de signal. Tandis que la technique VSS extrait les VE qui sont supérieures (ou inférieures pour le minimum) à un seuil prédéterminé. En général, l'approche par la technique BM peut parfois causer des manques des valeurs observés (Ferreira & de Haan, 2013). Cet inconvénient peut être corrigé en augmentant les nombres de blocs observés. Mais dans le cas où la mesure est limitée (notre cas), la technique BM pourrait involontairement manquer certaines valeurs. Cela entraîne une mauvaise identification de la distribution. Au contraire, la technique VSS semble plus adaptée à notre projet. De plus, la VSS est aussi appliquée pour l'évaluation de la fatigue par une extrapolation temporelle (Pär Johannesson, 2006; Poirier, 2013), ce qui laisse un bon espoir d'interpoler les paramètres de TVE seulement au lieu du signal complet.

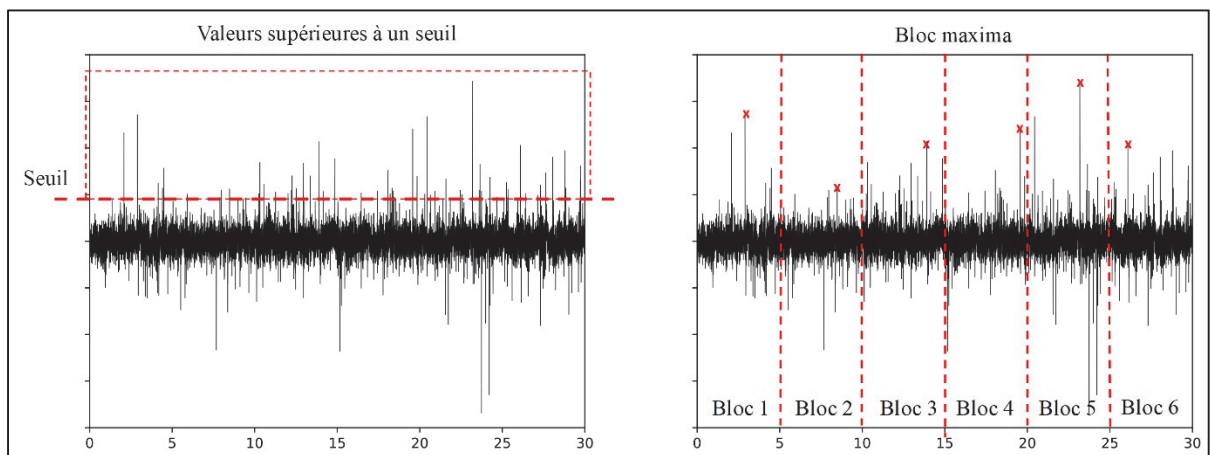


Figure 1.12 (Gauche) La technique des valeurs supérieures (VSS) avec des valeurs extrêmes extraites dans le rectangle rouge ; (Droite) La technique de bloc maxima (BM)

Si on utilise la TVE avec la technique de VSS pour le prétraitement avant l'interpolation, nous pouvons interpoler les seuils et les paramètres de distribution. Donc, si le seuil n'est pas bien déterminé, la qualité de l'interpolation sera imprécise. Ce seuil doit être bien choisi pour avoir un nombre suffisant de valeurs extrêmes qui peuvent être représentées par une distribution d'extrêmes (Pär Johannesson, 2006). Il y a des méthodes graphiques comme la moyenne des excès (*Mean Excess Plot*), ou des méthodes qui utilisent les tests statistiques comme test de Cramer-Von Mises, test de Anderson-Darling, etc. Poirier (2013) a présenté

une méthode qui combine certaines techniques (graphique et analytique) afin d’avoir un bon choix de seuil (voir Figure 1.13).

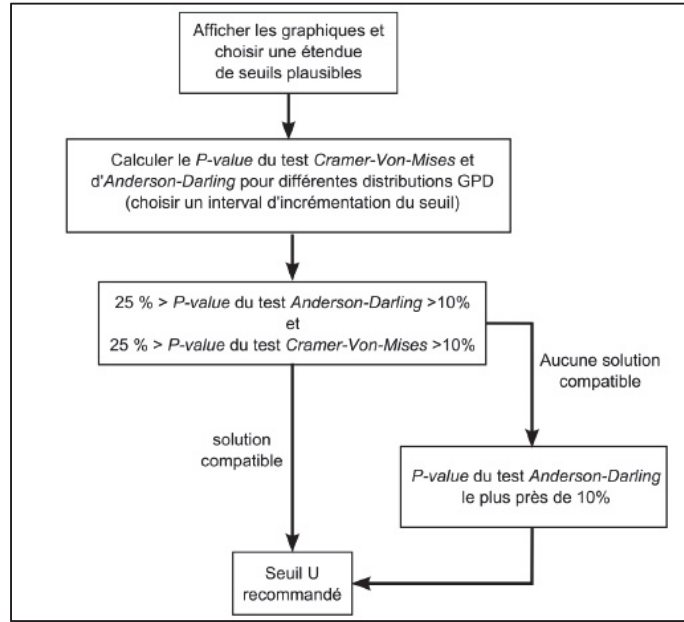


Figure 1.13 Une méthode pour choisir le seuil
Tirée de Poirier (2013)

Estimation par noyau

L'estimation par noyau (EPN) (*Kernel Density Estimation*) est une technique non-paramétrique employée pour estimer la densité de probabilité d'un ensemble de données (Wand & Jones, 1994). Le principe est d'estimer la densité de probabilité dans un point ciblé en déterminant une combinaison de différentes valeurs de noyaux d'observations approchées de ces points. L'EPN est souvent utilisée comme une méthode de lissage de l'histogramme (Figure 1.14), spécialement en cas de données limité. Une densité univariée estimée par EPN peut être présentée comme :

$$F_{EPN}^{uni}(x; h) = \frac{1}{n} \sum_{i=1}^n h^{-1} K\{(x - X_i)/h\} \quad (1.2)$$

où K est la fonction de noyau qui est souvent supposée symétrique, h est un coefficient de lissage (*Smoothing Bandwidth*) qui est positif et qui est lié au facteur de diffusion du noyau, $[X_1, X_2, \dots, X_n]$ sont les observations disponibles dans l'ensemble de données.

L'EPN est aussi présentée pour le cas de densité multivariée :

$$F_{EPN}^{multi}(x; H) = \frac{1}{n} \sum_{i=1}^n |H|^{-1/2} K\{H^{-1/2}(x - X_i)\} \quad (1.3)$$

où H est une matrice symétrique et définie positive qui contient des coefficients de lissage.

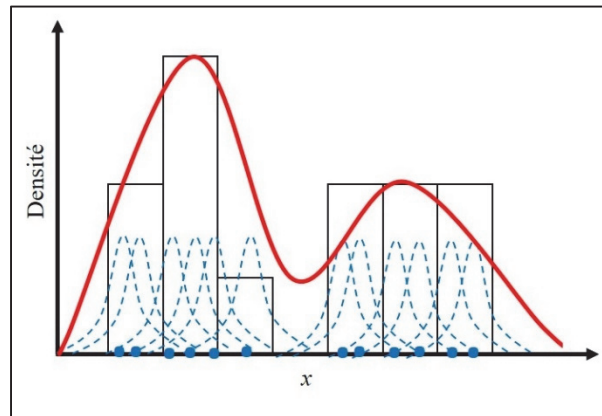


Figure 1.14 Estimation par noyau pour lisser un histogramme

Il y a plusieurs types de fonctions noyau qui sont utilisées (uniforme, gaussien, Epanechnikov, circulaire, etc). Cependant, pour l'EPN, le choix des coefficients de lissage (CL) joue un rôle plus important que le choix de la fonction de noyau (Wand & Jones, 1994). Les CLs peuvent être constants qui sont optimisés en utilisant certains calculs d'erreur comme l'erreur quadratique moyenne (MSE en anglais) ou l'erreur quadratique moyenne intégrée (MISE) (Heidenreich, Schindler, & Sperlich, 2013). Cependant, une valeur unique de CL pour toutes les données peut parfois causer de mauvaises estimations si les distributions des valeurs sont différentes dans différentes zones de données. En conséquence,

le CL peut être variés selon les positions des points ou des zones de distribution. La sélection de ces coefficients de lissage variables est réalisée par plusieurs méthodes comme la validation croisée (Duong & Hazelton, 2005), la méthode de Sheather–Jones et ses versions améliorées (Botev, Grotowski, & Kroese, 2010; Sheather & Jones, 1991), l'estimation bayésienne (Zougab, Adjabi, & Kokonendji, 2014), ou d'autres méthodes mentionnées dans la revue de Heidenreich et al. (2013).

En particulier, l'EPN est aussi appliquée dans le domaine de l'analyse de fiabilité des structures (Eckert-Gallup & Martin, 2016; Fu, Wang, & Zhao, 2016; D. Socie, 2001; D. F. Socie & Pompetzki, 2004). Donc, dans notre projet, l'utilisation de l'EPN présente un potentiel d'application pour l'estimation du niveau de fatigue après l'interpolation en générant les valeurs depuis la densité estimée. Cependant, si nous considérons la densité estimée par EPN comme le terme d'entrée de l'interpolation, les paramètres à interpoler seront les coefficients de lissage. Cela entraînera à un grand nombre d'interpolations à faire si nous utilisons des CL non constants.

1.3.1.2 Modélisation dans le domaine fréquentiel

Pour les signaux complexes, l'analyse spectrale permet d'identifier ou de modéliser des phénomènes physiques qui ne peuvent pas être observés en temporel. La transformation du domaine temporel au domaine fréquentiel est réalisée par la transformée de Fourier (TF) (Layer & Tomczyk, 2015). En particulier, pour un nombre de données limité comme les mesures expérimentales de contraintes sur la roue hydroélectrique, nous utiliserons la TF discrète (TFD).

Dans notre projet, une conservation du contenu fréquentiel pendant l'interpolation est essentielle. Il y a plusieurs approches pour visualiser et modéliser le spectre, mais toutes ne donnent pas une bonne visualisation dans le cas d'un signal complexe, et toutes ne sont pas compatibles avec l'interpolation. Par conséquent, on présentera certaines approches (et les méthodes de modélisations concernées) qui sont communément utilisées et qui sont

potentielles pour le projet de thèse : la densité spectrale de puissance et la transformée de Fourier à court terme.

Densité spectrale de puissance

La densité spectrale de puissance (PSD) est reliée à la distribution de puissance du signal dans le domaine fréquentiel. La PSD peut être mathématiquement définie comme la TF de la fonction d'autocorrélation (Heinzel, Rüdiger, & Schilling, 2002; Petrucci & Zuccarello, 1999). Pour un signal discret $x[n]$ on a :

$$P_x[k] = \sum_{n=1}^{N-1} R_x[\eta] e^{j2\pi k n / N} \quad (1.4)$$

où $P_x[k]$ est la PSD, N est le nombre de points dans le domaine temporel, et $R_x[\eta] = E\{x[n + \eta]x^*[n]\}$ est la fonction de l'autocorrélation.

Avec la capacité d'identifier les phénomènes typiques du signal en fréquences, la PSD est souvent utilisée dans la conception ou l'exploitation des structures (surtout des structures de vibration avec les chargements complexes. En particulière, la PSD est aussi appliquée pour l'analyse de fatigue des structures (Bishop & Sherratt, 1989; Larsen & Irvine, 2015; Petrucci & Zuccarello, 1999). Dans le domaine des turbines hydroélectriques, une des applications de la PSD est de visualiser et identifier dans différents régimes de fonctionnement les différents phénomènes physiques (ex, les phénomènes périodiques, le phénomène de torche, etc.) (voir les diagrammes de « *Waterfall* » dans la Figure 1.15). Aussi, à partir de la PSD, on peut créer un modèle stochastique qui permet la régénération du signal par simulation. Elle permet donc d'interpoler un signal stochastique. Une méthode connue de simulation stochastique, basée sur PSD, a été présentée en premier par (Shinozuka & Jan, 1972), puis dans ses autres travaux (Shinozuka & Deodatis, 1991; Yamazaki & Shinozuka, 1988). La formulation est :

$$x(t) = 2 \sum_{k=1}^N \sqrt{P_x(k\Delta\omega)\Delta\omega} \cdot e^{j\phi_k} e^{jk\Delta\omega t} \quad (1.5)$$

où P_x est la PSD de x , $\Delta\omega$ est l'intervalle de fréquence et, ϕ_k est une phase aléatoire sur $[0, 2\pi]$. Ce modèle stochastique est largement appliqué pour des simulations stochastiques dans différents domaines (l'hydroélectricité (Poirier et al., 2017), les ponts à longue portée (Cao, Xiang, & Zhou, 2000)). Dans l'état de l'art, ce modèle a été amélioré pour augmenter la qualité de la simulation (Bocchini & Deodatis, 2008; Masters & Gurley, 2003). En conclusion, un grand nombre de recherches ont été publiées sur les simulations stochastiques à partir de la PSD, ce qui présente un potentiel d'application et de développement pour notre cas.

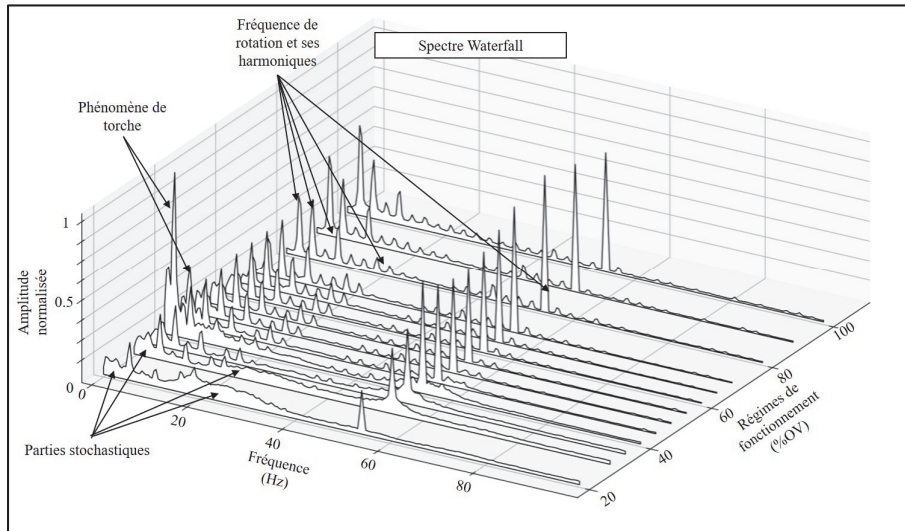


Figure 1.15 La PSD (sous la forme de spectre Waterfall) aide à visualiser et identifier les phénomènes physiques du signal dans le domaine fréquentiel

Modélisation en utilisant la transformée de Fourier à court terme

La Transformée de Fourier à Court Terme (TFCT) est une transformation qui permet de déterminer les informations fréquentielles (l'amplitude et la phase) pour chaque bloc temporel local d'un signal. La "coupe" de ces blocs temporels est réalisée avec l'aide d'une

fonction de fenêtre. La résolution de la TFCT dépend de la forme de fenêtre. Pour un signal discret $s[n]$, la TFCT peut être écrite :

$$S_{ik} = \sum_{i=-\infty}^{\infty} x[n]w[n-i] e^{-2j\pi nk/N_w} \quad (1.6)$$

où w est la fonction fenêtre de longue N_w . Le choix de la fenêtre dépend du signal étudié. Pour le cas de signal de contraintes, les fenêtres Hann ou Hamming sont recommandées. On trouvera plusieurs types de fenêtres dans (Heinzel et al., 2002; Layer & Tomczyk, 2015).

La TFCT est souvent utilisée pour l'analyse temps-fréquence du signal. Dans certaines recherches dans le domaine hydroélectrique, la TFCT est appliquée pour la représentation temps-fréquence du signal vibratoire sous la forme d'un spectrogramme (Q Dollon, Antoni, Tahan, Gagnon, & Monette, 2021; Quentin Dollon, Tahan, Antoni, Gagnon, & Monette, 2019; Gagnon & Léonard, 2013). De plus, la TFCT est aussi appliquée comme outil de reconstruction d'un signal. Par exemple, la génération du signal en utilisant l'amplitude de la TFCT a été étudiée depuis longtemps (Nawab, Quatieri, & Lim, 1983; Sturmél & Daudet, 2011). Une autre recherche a reconstruit le signal en utilisant la crête de la TFCT (*STFT Ridge* en anglais) (Zhu, Zhang, Gao, & Li, 2019).

La TFCT peut procurer un potentiel d'application et de développement pour modéliser avant et après l'interpolation. De plus, en particulier, non seulement la TFCT originale est utilisée pour régénérer le signal, mais aussi le signal peut être reconstruit à partir d'une TFCT modifiée qui est générée par la minimisation de l'erreur avec le signal original (Griffin & Lim, 1984). Cela élargit la stratégie de notre interpolation. Au lieu d'interpoler les coefficients de la TFCT originale, on pourrait interpoler les paramètres d'un processus de simulation qui permet de générer une TFCT modifiée. Plus précisément, nous visons une interpolation des processus stochastiques en utilisant la TFCT. Dans l'état de l'art, il n'y a pas encore des recherches qui utilisent la TFCT pour simuler un processus stochastique.

1.3.1.3 Comptage Rainflow

Le comptage Rainflow est une technique connue pour l'analyse de la fatigue induite par un signal vibratoire. Il a été introduit pour la première fois par Matsuishi et Endo (1968), et est devenu une technique très populaire. En particulière, le comptage rainflow est utilisé dans le domaine hydroélectrique pour l'évaluation de la fatigue de la roue ou pour vérifier les estimations numériques. Cette technique permet de déterminer les cycles de fatigue en attribuant les amplitudes extrêmes dans des cycles de chargement fermés. Ce comptage est souvent représenté par une matrice (noté comme MRF-Matrice de RainFlow) qui contient les cycles de fatigue (voir la Figure 1.16). Cette MRF nous permet de voir la distribution d'apparition des différents cycles du signal. Il y a plusieurs algorithmes pour le comptage rainflow comme l'algorithme "off-line" proposé par Rychlik (1987), ou les algorithmes "on-line" du comptage de "Four-Point" (FPC) (Dreßler, Hack, & Krüger, 1997), ou un algorithme récursif qui augmente l'efficacité du comptage Rainflow dans le cas des données limitées en réel (Musallam & Johnson, 2012).

Le comptage Rainflow nous permet d'approcher directement l'évaluation de la fatigue. Une des modélisations typiques est l'extrapolation de la MRF qui a été développée dans certaines recherches (Pär Johansson & Thomas, 2001; D. Socie, 2001; D. F. Socie & Pompetzki, 2004). L'objectif de ces extrapolations est de générer des cycles de fatigue en utilisant un processus probabiliste (simulation) qui est basé sur les informations de la MRF originale. En effet, vu que les signaux vibratoires mesurés en réel sont souvent disponibles avec une période très courte, l'idée de génération des cycles dans la MRF ouvre des perspectives pour la capacité d'évaluation du niveau de fatigue à long terme. De plus, il est aussi possible de reconstruire le signal temporel à partir de la MRF modifiée/générée par une méthode inverse (D. Socie, 2001). L'inconvénient est que le signal reconstruit est totalement aléatoire. Cependant, dans notre point de vue, si on se concentre sur l'objectif de l'évaluation de fatigue, la modélisation du signal dans le domaine Rainflow sera visée comme une méthode potentielle de traitement du signal pour la création du processus d'interpolation.

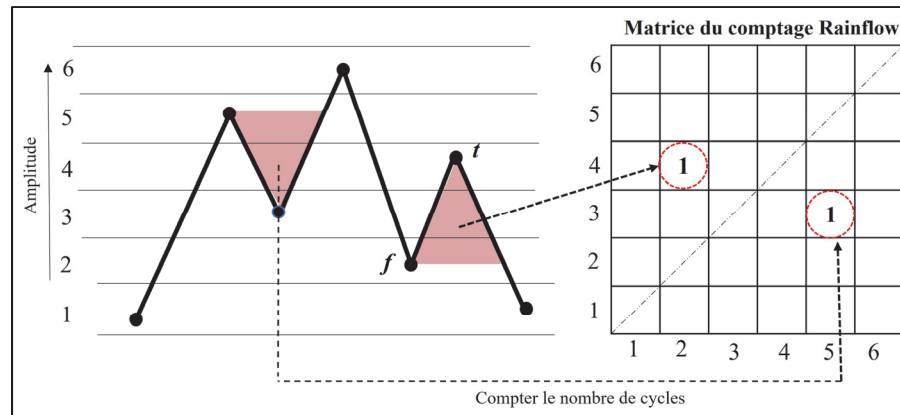


Figure 1.16 Un exemple de la matrice du comptage Rainflow qui est créée par l'algorithme de "Four-Points"

1.3.1.4 Cyclostationnarité

Dans les systèmes mécaniques tournants comme les turbines hydroélectriques, les composantes périodiques existent toujours dans les signaux qui sont mesurés sur la partie tournante comme la roue. Cependant, la détection de ces périodicités n'est pas évidente, spécialement dans le cas de la turbine hydraulique où les signaux mesurés sont parfois dominés par les composantes stochastiques. On a besoin donc d'une méthode qui peut isoler ces parties périodiques. Ceci conduit au cadre CycloStationnaire (CS) qui permet d'extraire les périodicités cachées dans un signal. Autrement dit, la CS permet de séparer les composantes périodiques et stochastiques d'un signal en régime stationnaire. Par conséquent, avec une capacité de couvrir deux types typiques du signal (signal périodique et signal aléatoire stationnaire), un modèle basé sur la CS est une bonne solution afin de diminuer les pertes d'information lors de la création du modèle d'interpolation. La CS devient de plus en plus une technique importante dans le traitement du signal. La CS est un domaine de recherche mature, les revues de Gardner, Napolitano, et Paura (2006); Serpedin, Panduru, Sari, et Giannakis (2005) sont deux bonnes présentations de l'état de l'art. En particulière, pour un tutoriel de la CS spécifique dans le domaine mécanique, nous référons à Antoni (2009). Par ailleurs, les modélisations du signal de machine tournante (Antoni, Bonnardot, Raad, & El Badaoui, 2004) et spécifiquement de la turbine hydroélectrique (Pennacchi,

Borghesani, & Chatterton, 2015; Poirier et al., 2017) dans le cadre CS sont aussi dignes d'intérêt.

Dans le cadre cyclostationnaire, un signal $x(t)$ peut être décomposé comme :

$$x(t) = \mathcal{P}[x(t)] + \mathcal{R}[x(t)] \quad (1.7)$$

où $\mathcal{P}[x(t)]$ est la partie périodique du signal $x(t)$ qui comprend seulement la CS de premier ordre, $\mathcal{R}[x(t)]$ est la partie résiduelle qui présente la CS d'ordre 2 et/ou des ordres supérieurs. Un exemple de cette décomposition est montré dans la Figure 1.17.

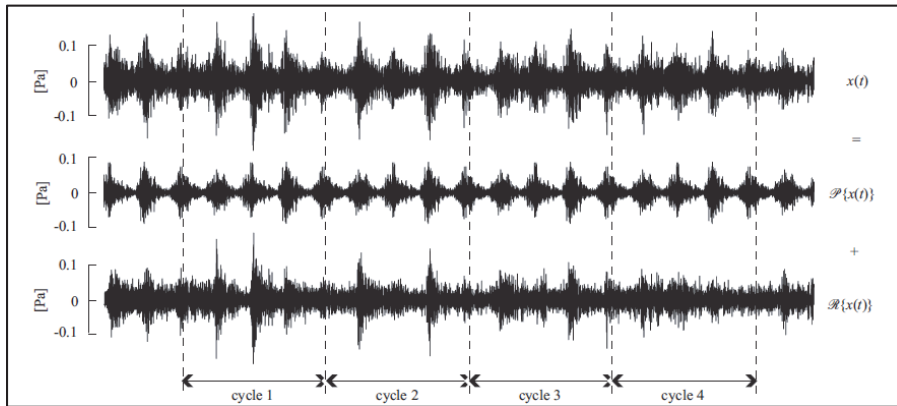


Figure 1.17 Décomposition du signal acoustique d'un moteur diesel
Tirée de Antoni (2009)

Les ordres de cyclostationnarité permettent de distinguer différents phénomènes physiques dans un signal. La CS de premier ordre caractérise un comportement physique lié souvent au comportement périodique (par exemple, la rotation synchrone de la roue hydroélectrique). Par ailleurs, le deuxième ordre de CS exprime souvent les comportements stochastiques qui sont difficiles à prédire, et qui correspondent aux microphénomènes typiques dans chaque type de système (Antoni, 2009). Le phénomène de torche causé par le flux d'eau sortie de la roue d'une turbine hydroélectrique (Poirier, 2013), ou le signal mesuré du palier (*Bearing*) dans une boîte de vitesse (*Gearbox*) (Antoni & Randall, 2002) sont des exemples typiques pour la CS d'ordre 2.

Il n'est pas facile de déterminer précisément les natures physiques de cyclostationnarité des ordres supérieurs (>2) (la réalisation d'un outil d'extraction de ces ordres supérieurs est difficile). Le choix du nombre d'ordres à exprimer et l'outil d'extraction des composantes cyclostationnaires sont donc importants. Une réalisation plus simple de la décomposition ne nous permet peut-être pas d'observer toutes les natures du signal dont on a besoin. En pratique, seulement deux premiers ordres sont souvent utilisés afin de montrer les caractéristiques de cyclostationnarité dans un signal (Antoni, 2009). Par conséquent, on pourrait écrire :

$$x(t) \approx CS1(t) + CS2(t) + R(t) \quad (1.8)$$

où $CS1(t)$ est la composante du premier ordre de cyclostationnarité, $CS2(t)$ du deuxième ordre de cyclostationnarité, et $R(t)$ est la partie résiduelle. En particulier, Poirier et al. (2017) ont utilisé seulement deux premiers ordres de CS pour définir un signal de contrainte de la roue hydroélectrique dans le cadre cyclostationnaire. Vu que l'interpolation d'un signal complet est toujours un défi, ce modèle pourrait donc donner un grand potentiel de prétraitement du signal avant interpolation.

Il faut noter que le modèle de cyclostationnarité est basé sur l'étude de la variable angulaire plutôt que le temps (Antoni, 2009; Antoni et al., 2004; Bonnardot, El Badaoui, Randall, Danieri, & Guillet, 2005). En particulier, pour la turbine hydroélectrique, les parties périodiques cachées sont liées aux angles de rotation de l'arbre. Par conséquent, l'application d'un rééchantillonnage angulaire sur le signal temporel étudié permet de mieux préserver les caractérisations cyclostationnaires. C'est-à-dire que la variable temporelle t dans l'équation (1.8) doit être remplacée par la variable angulaire θ . Plusieurs méthodes ont été développées pour l'échantillonnage angulaire avec une bonne performance. Par exemple, la méthode directe qui utilise un encodeur angulaire (Bonnardot, 2004) et les méthodes de rééchantillonnage angulaire a posteriori ou '*Order Tracking*' (Combet & Zimroz, 2009; Fyfe & Munck, 1997; McFadden, 1989). Dans un rééchantillonnage a posteriori, l'étape de l'estimation de la position de l'arbre demande souvent un capteur de position (encodeur), mais ce capteur n'est pas toujours disponible. Par conséquent, le rééchantillonnage a

posteriori sans capteur a aussi été développé. Bonnardot, El Badaoui, et al. (2005) ont proposé une méthode sans capteur en utilisant le signal de l'accélération.

Vu que les natures de chaque ordre de cyclostationnarité sont différentes, les outils d'extraction des composantes pour chaque ordre sont différents. Comparé aux autres ordres supérieurs, le premier ordre de CS est le plus facile à extraire. Pour l'extraction des composantes périodiques, plusieurs méthodes ont été proposées pour extraire la CS1 comme la moyenne synchrone qui permet d'extraire la moyenne dans chaque localisation angulaire du signal (Antoni et al., 2004; Boustany & Antoni, 2005), ou les analyses par la transformée de Fourier (Antoni, 2005), etc. Par ailleurs, en raison de la nature stochastique, l'extraction des composantes cyclostationnaires de second ordre (CS2) est plus difficile. Une méthode est d'utiliser les outils liés au cumulants d'ordre deux du signal, comme la fonction d'autocovariance synchrone et les autres quantités reliées (Antoni et al., 2004). Par ailleurs, dans le domaine de SAS - séparation aveugle de source (*Blind Source Separation*), certaines méthodes d'extraction CS2 ont été réalisées. Par exemple, une méthode qui utilise une décomposition en sous-espaces en se basant sur les fréquences cycliques a été proposé par Boustany et Antoni (2005). Une autre, qui est aussi basée sur les fréquences cycliques, utilise une extension optimisée du filtre Wiener a été proposé par (Bonnardot, Randall, & Guillet, 2005). De plus, en profitant des avantages de deux méthodes mentionnées précédentes, Boustany et Antoni (2008) ont proposé une extraction en utilisant une régression cyclique de rang réduit (*Reduced-Rank Cyclic Regression*). Poirier et al. (2017) a proposé une extraction de la CS2 en utilisant les phases estimées depuis une bande de fréquences d'intérêt, cela permet d'extraire les fluctuations d'une fréquence et ses harmoniques (Figure 1.18). Cette méthode nous semble être la plus pertinente pour notre projet car elle a été appliquée avec succès dans le cas des signaux de la roue hydroélectrique. De plus, il est possible de réaliser une extraction des composantes du phénomène de torche caché dans le signal de déformation.

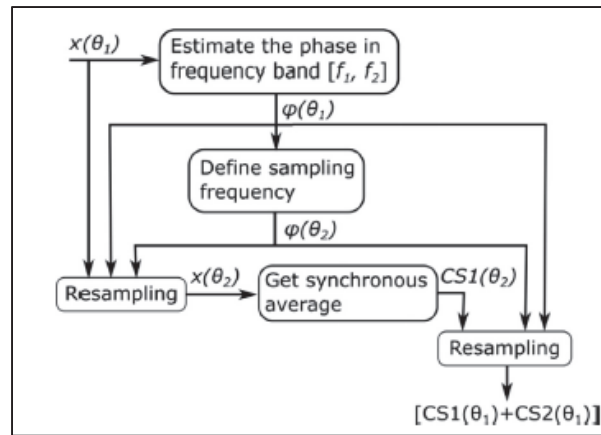


Figure 1.18 Processus d'extraction de CS2
qui est potentiellement utilisé
pour séparer le phénomène de torche
Tirée de Poirier et al. (2017)

En conclusion, on limite la séparation du signal d'intérêt à seulement les deux premiers ordres de CS. En général, l'extraction de la CS1 est bien établie. Le choix de l'outil pour la CS2 dépend de la nature physique qu'on veut extraire. Dans le cas des signaux de la roue hydroélectrique, le terme de CS2 peut-être lié au phénomène de torche ou encore lié à quelques fluctuations d'amplitude autour d'une fréquence synchrone.

1.3.1.5 Conclusion sur les approches de modélisation

Le choix des approches de modélisations avant l'interpolation est important. Chaque approche s'avère potentiellement adapté à une situation. Dans cette revue, nous avons présenté généralement les procédures de différentes approches. Pour conclure, le Tableau 1.1 regroupe les principaux éléments qui affectent la capacité de l'interpolation.

Tableau 1.1 Comparaison la capacité des approches de modélisation pour l'interpolation

Approche de modélisation	Avantages	Difficultés
Modèle de mélange	Possibilité d'interpoler les paramètres pour le cas du modèle de mélange gaussien	Difficile d'interpoler pour le cas du modèle non Gaussien
Distributions des valeurs extrêmes	Interpolation rapide adaptée aux valeurs extrêmes	Le choix du seuil est très sensible pour l'interpolation
EPN	Interpolation et extrapolation temporelles en combinant avec la matrice rainflow	Difficile d'interpoler les coefficients de lissage qui varient en fonction de la localisation,
Basée sur PSD	Interpolation des parties aléatoires du signal de contrainte	Interpolation spatiale des PSD est difficile
Basée sur TFCT	Permet de modéliser les non-linéarités en fréquences, ce qui augmente la capacité de simulation des processus non-gaussiens	Interpolation des coefficients de TFCT n'est pas évidentes. Il faut interpoler les paramètres de simulations. La capacité d'interpolation dépend donc de stratégie de simulation
Comptage rainflow	Interpolation rapide adaptée aux études en fatigue	Le signal reconstruit depuis un comptage Rainflow est complètement aléatoire
Cyclostationnarité	Facilité d'interpolation en séparant le signal en certaines parties	Le nombre des interpolations et le temps de calcul sont grands

1.3.2 Méthode d'interpolation

D'après notre revue, on peut séparer les méthodes d'interpolation en trois types : les méthodes déterministes (MD), les méthodes géostatistiques ou stochastiques (MG) (comme le krigeage, ou la simulation stochastique), les méthodes d'apprentissage automatique (MAA) (*Machine Learning Based Method*).

En raison du développement du domaine d'apprentissage, les MAA comme le réseau de neurones ou les forêts d'arbres décisionnels (*Random Forest*) ont été largement appliquées

durant les dernières années (Bélisle, Huang, Le Digabel, & Gheribi, 2015; J. Li, Heap, Potter, & Daniell, 2011). Cependant, le problème de ces méthodes est qu'elles nécessitent un grand nombre des données. Des données limitées pour l'apprentissage (*Training Data*) ne peuvent pas assurer une prédiction précise. Considérant que les mesures expérimentales de contraintes sur la roue sont limitées dans certains RF (environ des dizaines RF), l'application des MAA nous semble non adéquate à notre problème.

1.3.2.1 Méthodes déterministes

L'interpolation par les MD permet d'attribuer des valeurs pour chaque localisation en évaluant la similarité avec des observations mesurées aux voisinages. Les estimations seront donc considérablement influencées par des voisins les plus proches dans l'ensemble. On peut mentionner certaines méthodes typiques comme l'interpolation par voisins (IPV), l'interpolation polynomiale, l'interpolation par spline, ou encore la pondération inverse à la distance (IDW). Pour plus d'informations des MD, la revue de J. Li et Heap (2014) est recommandée.

Interpolation par voisins

L'approche la plus simple pour l'IPV est l'interpolation au plus proche voisin où la valeur estimée est considérée comme égale au point du voisin le plus proche. Cependant, cette approche est trop simpliste et elle n'est pas adaptée parce qu'elle ignore l'influence d'autres observations que le voisin le plus proche. En particulière, cette approche est le cas spécifique de l>IDW quand les poids de régressions sont 0 et 1 (J. Li & Heap, 2014).

Une autre approche est l'interpolation par voisins naturels (IVN) qui est méthode multi variables basée sur le diagramme de Voronoi (Sibson, 1981). Chaque point est initialement lié à une cellule. Ensuite, les voisins naturels d'un nouveau point x sont définis comme des points tels que leurs cellules sont modifiées par l'apparition d'une nouvelle cellule autour de

x . Les poids d'interpolation sont calculés par la proportion d'intersection de la nouvelle cellule et les cellules initiales. L'IVN est écrite par :

$$IVN(x) = \sum_i^k \pi_{x_i}(x) f_{x_i}(x) \quad (1.9)$$

où $\pi_{x_i}(x) = \frac{V_i}{V_x}$ sont les poids correspondants aux points du voisin naturel x_i , V_x est le volume de la nouvelle cellule créé par x , V_i est le volume de zone d'intersection entre cette nouvelle cellule et la cellule de x_i , $f_{x_i}(x)$ est la fonction de référence qui attribue la valeur dans la position x_i . L'IVN peut fournir une fonction d'estimation continue. Elle est souvent utilisée pour la reconstruction des surfaces lisses (Boissonnat & Cazals, 2002).

Pondération inverse à la distance

La pondération inverse à la distance (PID) est une méthode déterministe qui interpole une valeur dans une position ciblée s par la moyenne pondérée des observations :

$$z^*(s) = \sum_i^n \pi_i^{PID} z(s_i) \quad (1.10)$$

où π_i^{PID} sont les poids calculés par une formule de l'inverse de distance à une puissance q strictement positive :

$$\pi_i^{PID} = \frac{\|s - s_i\|^{-q}}{\sum_i^n \|s - s_i\|^{-q}} \quad (1.11)$$

La PID donne une interpolation rapide, particulièrement adaptée au cas des données qui sont régulièrement distribuées (J. Li & Heap, 2014). Cependant, la performance de PID est réduite s'il y a des regroupements de données (*Clustering*) dans la distribution (ce qui est

possiblement observé dans les mesures de contraintes). Pour améliorer la qualité de l'interpolation par PID dans le cas de la distribution non-uniforme, Z. Li, Wang, Ma, et Wu (2018) ont proposé d'optimiser la distance pondérée en ajoutant un coefficient de réglage. On peut aussi mentionner une version améliorée qui permet de varier les coefficients pondérés en fonction de la caractéristique spatiale des points aux voisinage qui a été proposé par Lu et Wong (2008).

Splines

L'interpolation par splines est liée à une fonction définie par des polynômes en morceaux. Basée sur l'ordre des polynômes, on peut parler des splines classiques comme linéaire (1^{er} ordre), quadratique (2^{ème} ordre), et cubique (3^{ème} ordre). En pratique, on préfère utiliser la spline cubique, car elle est suffisamment riche pour reproduire un cas réel et elle n'exige pas trop de calculs comparée aux autres ordres supérieurs (Erdogan, 2013). Une interpolation par splines classiques peut fournir une courbe qui passe exactement par les observations, mais elle ne donne pas une indication des erreurs d'estimation. Par conséquent, dans plusieurs cas, on utilise une approche par spline de lissage, par exemple la spline en plaque mince (SPM) (Thin Plate Spline en anglais) (Dubrule, 1983; Hutchinson, 1995). Cette approche permet de lisser la courbe d'interpolation avec des petites erreurs acceptables pour chaque observation, au lieu d'obligatoirement passer par tous les points. Il faut noter qu'on met la spline dans le cadre de MDs, mais c'est seulement vrai pour le cas des splines exactes comme les splines classiques. En particulière, la SPM, qui fonctionne comme une extension d'une régression linéaire, est considérée comme une approche stochastique et elle peut fournir une quantification de l'incertitude de l'estimation (Haylock et al., 2008).

L'interpolation basé sur les splines a été appliquée dans plusieurs recherches du domaine des turbines hydroélectriques (Brezovec, Kuzle, & Tomisa, 2006; Poirier et al., 2017; Skjelbred & Kong, 2019). Elle est utilisée également avec des extensions comme la spline laplacienne (Hastie & Tibshirani, 1990), la spline cubique d'Hermite (Erdogan, 2013; Lekkas & Fossen,

2014). De plus, vu que la spline est plus simple que les autres types d'interpolation comme le krigeage ou les MAA, elle est une bonne méthode pour réaliser une interpolation rapide.

1.3.2.2 Krigeage

Les méthodes d'interpolations stochastiques sont les interpolations basées sur les relations statistiques (autocorrélation, covariance...) entre les valeurs de données. Par conséquent, les méthodes stochastiques peuvent aussi fournir l'indication des mesures d'erreur ou d'incertitude de la prédiction, ce qui n'est pas évident pour les méthodes déterministes.

Parmi les méthodes stochastiques, la famille de krigeage est un ensemble de méthodes d'interpolation très employé. Le krigeage est une méthode d'estimation optimale qui a été initialement exprimée par D.G. Krige (Krige, 1951). Une revue de l'origine de krigeage a été introduite par Cressie (1990). En général, le krigeage est une méthode d'interpolation spatiale non biaisée qui permet d'estimer les valeurs manquantes à partir des observations via un processus de régression linéaire :

$$Z^*(s_c) - m(s_c) = \sum_i^n \lambda_i(s) [Z(s_i) - m(s_i)] \quad (1.12)$$

où $Z(s_i)$ est l'observation à la localisation s_i , $m(s_i)$ est la moyenne de la variable aléatoire Z à la localisation s_i , $Z^*(s_c)$ est la valeur estimée à localisation ciblée s_c . Il faut noter que cette équation de régression est une formulation générale. Elle peut être différente pour chaque type de krigeage. Le principe est de déterminer les poids de régressions (pour toutes les observations) qui minimisent la variance d'estimation $\sigma_e(s)^2$.

Le processus de minimisation a besoin d'une carte de covariance qui est exprimée généralement en fonction des distances spatiales de toutes les paires de points (y compris les points cibles et les observations). Dans un plan spatial de mesures, deux observations, qui sont situées l'une proche de l'autre, devraient se ressembler plus que deux observations éloignées. Par conséquent, en se basant sur l'analyse de la dissimilarité entre des points, le

krigeage devient une méthode d'estimation qui conserve des erreurs-variances minimales. Par ailleurs, comme mentionné, le krigeage donne non seulement des valeurs approximatives, mais il fournit aussi une indication du niveau d'incertitude de l'estimation (par exemple, la propagation de l'incertitude via l'analyse des variances d'estimation, l'incertitude de l'interpolation via la simulation stochastique conditionnelle).

La fonction de covariance a besoin d'information de la moyenne qui est souvent inconnue en pratique (Montero, Fernández-Avilés, & Mateu, 2015). Par conséquent, dans plusieurs cas, le krigeage utilise un autre représentant appelé le SemiVariogramme (SV) qui montre la demi-variance entre des paires des mesures en fonction de la distance entre deux observations. Le SV peut être expérimentalement déterminé en utilisant les observations disponibles :

$$\hat{\gamma}(h) = \frac{1}{N(h)} \sum_{(i,j) \in \vartheta(h)} [Z(s_i) - Z(s_j)] \quad (1.13)$$

où $\vartheta(h)$ est l'ensemble des paires de points séparés par une distance h et $N(h)$ est le nombre de paires de points dans $\vartheta(h)$. Cependant, le SV expérimental ne peut pas donner les informations pour tous les incréments h possibles. La solution est de modéliser le semivariogramme par un ou des modèles paramétriques comme le modèle sphérique, le modèle gaussien, le modèle exponentiel, etc. La liste de ces fonctions est facilement trouvée dans des manuels de géostatiques (Deutsch & Journel, 1992; Goovaerts, 1997).

Le krigeage a été d'abord beaucoup utilisé dans le domaine des ressources naturelles ou de la science de la Terre (mine, météo, pétrolier, environnement...), qui sont basés sur les informations géographiques (Deutsch & Journel, 1992; Goovaerts, 1997; Hudson & Wackernagel, 1994; Andre G Journel & Huijbregts, 1976; Kassim & Kottegoda, 1991; Oliver & Webster, 1990; Snepvangers, Heuvelink, & Huisman, 2003; Vann, Jackson, & Bertoli, 2003). Ensuite, le krigeage est largement utilisé dans plusieurs domaines de l'ingénieur, en particulier dans le domaine hydroélectrique (Ben Salah, 2014; Léonard, 2015; Lucifredi, Mazzieri, & Rossi, 2000; Moreau, Bage, Marcotte, & Samson, 2012). En

particulier, le travail de Ben Salah (2014) a utilisé le krigeage pour réaliser l'interpolation spatiale des mesures de contrainte entre différentes localisations de l'aube de la roue hydroélectrique.

Il y a plusieurs types de krigeage qui créent une famille d'estimation flexible. Cela permet de s'adapter à différentes situations d'interpolation. La théorie et les formulaires des types de krigeage sont facilement trouvés dans les manuels géostatistique (Chiles & Delfiner, 2009; Deutsch & Journel, 1992; Goovaerts, 1997; Montero et al., 2015). Les trois types « classiques » de krigeage sont : le Krigeage Simple (KS) qui considère que la moyenne est connue et constante pour toutes les mesures, le Krigeage Ordinaire (KO) où la moyenne est considérée constante et inconnue et Krigeage Universel (KU) où la tendance pour la moyenne est modélisée comme une fonction déterministe variée de façon lisse (Ben Salah, 2014). D'après une soixantaine d'années de développement, plusieurs versions de krigeage ont été étudiées pour des situations différentes d'interpolation. Par rapport à la problématique de notre interpolation, nous présentons, parmi plusieurs types de krigeage, trois types potentiels.

D'abord, nous mentionnons une approche non paramétrique appelée le Krigeage d'Indicatrice (KI) qui permet d'estimer une variable indicatrice (0 ou 1) transformée depuis les observations par rapport à un (ou des) seuil(s) donné(s) (André G Journel, 1983; Suro-Perez & Journel, 1991). Le KI est dit une méthode non paramétrique, car il ne demande aucune hypothèse imposée à la distribution des variables (Marinoni, 2003). De plus, avec la transformation vers la forme binaire (0 et 1), le KI devient moins sensible aux valeurs extrêmes. Vu que le KI peut fournir l'estimation de la probabilité des valeurs liées à un seuil donné, le KI est souvent utilisé pour l'analyse de la qualité ou du risque de pollution (des sols, de l'eau,) (Antunes & Albuquerque, 2013; C.-W. Liu, Jang, & Liao, 2004; Smith, Halvorson, & Papendick, 1993). Dans notre projet, les mesures de contraintes de la roue peuvent parfois être représentées par des distributions à queue lourde générée par de grands pics de signaux. Par conséquent, le KI est une approche potentielle dans notre projet.

Cependant, la difficulté du KI est le choix des seuils et aussi un grand nombre de traitements s'il y a plusieurs seuils donnés, ce qui augmente le coût de calcul.

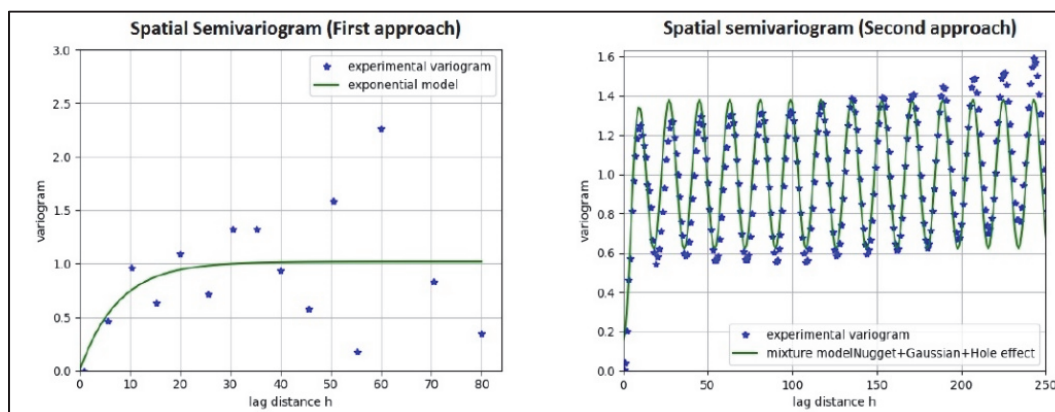


Figure 1.19 Un exemple de semivariogramme expérimental et d'un modèle paramétrique de semivariogramme
Tirée de Pham, Gagnon, Antoni, Tahan, et Monette (2019)

Un autre type de krigeage potentiel est le Krigeage Spatio-Temporel (KST). En pratique, l'exploitation des données n'est pas limitée dans le domaine spatial, mais il faut parfois élargir au domaine temporel. Le principe d'interpolation du KST est le même, mais le KST permet d'estimer une valeur dans une localisation spatiale à un moment temporel donné (Montero et al., 2015). La localisation spatio-temporelle entraîne que le semivariogramme (ou la fonction de covariance) dans le cas de KST soit bidimensionnel (voir Figure 1.20). Par conséquent, elle est plus difficilement modélisable que la version en une seule dimension. Plusieurs approches mathématiques comme la métrique de somme, le modèle « *Separable* », la métrique de « *Product-Sum* » permettent de combiner les deux semivariogrammes spatiales (avec des incréments dans l'espace) et temporelles (avec des incréments dans le temps) (Pebesma & Heuvelink, 2016). Le KST est souvent appliqué dans le domaine météorologique ou le domaine de la science des sols qui nécessitent d'estimer certaines valeurs (précipitations, eau de rétention) de façon quotidienne, mensuelle, ou annuelle (Aryaputera, Yang, Zhao, & Walsh, 2015; Hu, Shu, Hu, & Xu, 2017; Jost, Heuvelink, & Papritz, 2005; Snepvangers et al., 2003). Pour notre projet, les mesures de contraintes contribuent à l'évaluation de la fatigue (besoin d'informations à long terme). Par conséquent,

le KST est une méthode appropriée qui pourrait permettre de réaliser à la fois l'interpolation spatiale entre les RF et, l'interpolation ou l'extrapolation dans le temps.

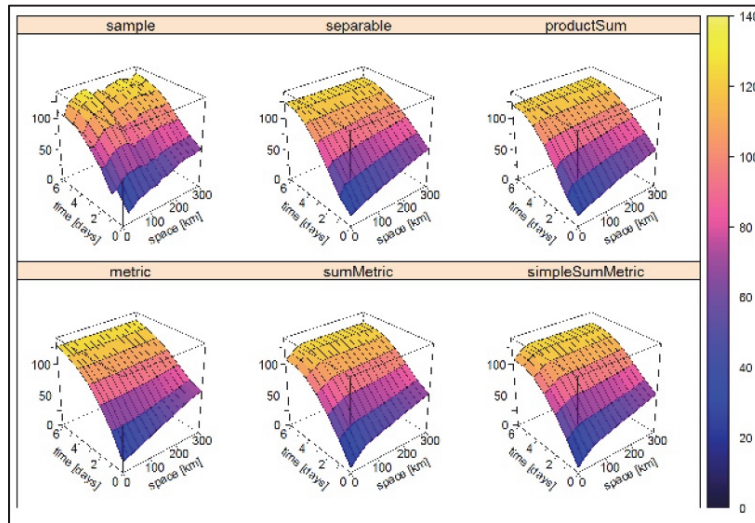


Figure 1.20 Un exemple de différents modèles de semivariogramme spatio-temporel
Tirée de Pebesma et Heuvelink (2016)

Le dernier type de krigeage présenté dans cette revue est le Cokrigeage (CK), qui permet d'ajouter une ou des covariables corrélées pour améliorer l'interpolation de la première variable. La théorie et la formulation du CK ont été présentées depuis les années 90s (Goovaerts, 1998; D. E. Myers, 1982, 1991). La motivation du CK est de résoudre le problème du manque de données en utilisant une combinaison pondérée entre la variable d'intérêt et les covariables. Donc, le choix de covariable influe directement sur la précision d'interpolation. Dans le cas du CK, en plus de l'auto-semivariogramme pour chaque variable, on doit traiter un semivariogramme croisé qui est souvent modélisé par un modèle de co-régionalisation linéaire (Goulard & Voltz, 1992). La modélisation du semivariogramme dans le cas du CK n'est pas une procédure stable. Cependant, le besoin d'utiliser le CK est augmenté en raison du manque de données observées. Certaines recherches ont été proposées pour améliorer la modélisation du semivariogramme croisé (Lark, 2002; Lark & Papritz, 2003). D'après notre point de vue, le CK est une approche potentielle pour notre situation d'interpolation, car les observations expérimentales sont limitées. De plus, certains

phénomènes physiques (comme le phénomène de torche) dans le signal de contrainte pourraient être observés dans d'autres signaux de la turbine. Cela nous donne une chance de trouver une covariable appropriée pour le CK.

1.3.2.3 Conclusion

Pour une conclusion, le Tableau 1.2 donne une comparaison des avantages et des inconvénients des méthodes présentées dans cette revue de littérature via certains critères typiques. D'après cette comparaison, la famille de krigeage semble être la méthode qui convient le mieux à notre situation d'interpolation. De plus, le krigeage est aussi recommandé d'après des expériences de notre partenaire industrielle.

Tableau 1.2 Comparaison entre des méthodes d'interpolation présentées dans la revue

Méthodes	Types	Indication d'incertitude	Points forts	Niveau d'adaptation pour notre projet
IPV	Déterministe	Non	Rapide	Simple pour notre besoin
IDW	Déterministe	Non	Rapide	Simple pour notre besoin
Splines	Déterministe /Stochastique	Oui	Rapide, plusieurs types, donc flexibles.	Potentiel pour une interpolation rapide des paramètres mathématiques
KI	Stochastiques	Oui	Moins sensible aux valeurs aberrantes	Potentiel pour une interpolation principale, mais le temps de calcul est possiblement grand
KST	Stochastique	Oui	Possibilité d'interpoler en plus dans le domaine temporel	Potentiel pour devenir l'interpolation principale du projet en raison de capacité de la prédiction spatio-temporelle
CK	Stochastique	Oui	Interpolation multivariée avec des covariables corrélés	Potentiel pour l'interpolations des phénomènes spécifiques comme la torche

1.3.3 Méthodes de validation

La validation est une étape importante et incontournable pour évaluer la qualité de l'interpolation. Il n'y a aucune méthode de validation qui peut être adaptée à tous les cas. Pour chaque ensemble de données ou chaque objectif de prédiction, il faut utiliser une ou des méthodes spécifiques. En général, le principe consiste à comparer les valeurs estimées par le modèle de prédiction et les valeurs réelles (expérimentales observations ou mesures). Une comparaison classique est la validation graphique visuelle. Cependant, une validation graphique représente un cas qualitatif qui ne donne aucune valeur d'incertitude du modèle de l'estimation. De plus, une validation graphique peut varier entre différentes personnes (Y. Liu, Chen, Arendt, & Huang, 2011). Il faut donc utiliser aussi des métriques quantitatives (ex., les erreurs statistiques). Basées sur les données étudiées, certaines approches pour la validation de l'outil d'interpolation sont présentées dans cette revue de littérature (ex., erreurs d'estimation, validation croisée, *Area Metric*...). Les méthodes présentées sont potentielles, non seulement pour la validation des valeurs interpolées, mais aussi pour aider à optimiser des modèles partiels ou à guider le choix des hyper paramètres clés durant l'interpolation.

1.3.3.1 Mesures de l'erreur de prédiction

Les mesures de l'erreur de prédiction comme l'erreur absolue moyenne (EAM) ou la racine de l'erreur quadratique moyenne (REQM), etc., sont un critère simple et essentiel pour évaluer la qualité de prédiction. Ce type de mesures sont toujours un bon choix pour une première observation de l'erreur du modèle de prédiction. Il y a plusieurs mesures des erreurs qui sont séparées en certains groupes comme les erreurs absolues, les erreurs de pourcentages, les erreurs symétriques... (Shcherbakov et al., 2013). Deux types qui sont couramment utilisés sont la REQM :

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - x_i^*)^2} \quad (1.14)$$

et la EAM:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |x_i - x_i^*| \quad (1.15)$$

où x_i^* est la valeur estimée par le modèle. En particulier, la REQM et la EAM sont aussi utilisées dans l'interpolation par krigeage pour aider à modéliser et choisir les modèles paramétriques de covariances (ou semivariogramme) (Pebesma & Heuvelink, 2016). Dans notre projet, les données étudiées sont les signaux temporels de contrainte. Par conséquent, les mesures de REQM et de EAM permettent de statuer rapidement de la similarité entre le signal expérimental et l'ensemble des signaux estimés. Le grand inconvénient de la REQM est qu'elle est sensible aux valeurs aberrantes (Chai & Draxler, 2014), tandis que la EAM est moins sensible. Ces valeurs aberrantes sont souvent liées aux valeurs extrêmes du signal, qui sont observées dans notre signal étudié. Les avantages et inconvénients de la EAM ou REQM ont été un sujet intéressant par certains chercheurs (Chai & Draxler, 2014; Willmott & Matsuura, 2005).

Au lieu d'utiliser les erreurs absolues, l'estimation de l'erreur relative (ER) est aussi utilisée :

$$ER = \frac{x_i^* - x_i}{|x_i|} \quad (1.16)$$

L'ER ne calcule pas une erreur absolue, mais elle donne une valeur algébrique. Son avantage est de donner une évaluation du niveau de sous-estimation ou surestimation du modèle. Cependant, l'ER est souvent adaptée à l'erreur de prédiction pour une valeur unique (par exemple, l'amplitude maximum du signal) plutôt qu'un ensemble comme un signal complet. En conclusion, chaque type d'erreur possède des avantages et des inconvénients propres, mais en pratique, une combinaison de plusieurs types différents est toujours un meilleur choix pour la validation qu'un type unique.

1.3.3.2 Validation croisée

La validation croisée (VC) est une technique statistique qui permet de vérifier la robustesse d'un modèle pour un ensemble fini de données. Elle est souvent utilisée pour une sélection du modèle (Mosier, 1951). Le principe de la VC est de séparer un ensemble de données en deux groupes : un groupe d'entraînement (GE) pour un processus d'apprentissage du modèle et un groupe de validation (GV) pour vérifier la performance du modèle optimisé. Pour tester la performance, les mesures d'erreur de prédiction (REQM, EAM, ...) ou les autres métriques statistiques peuvent être appliquées pour comparer les valeurs estimées et les valeurs originales. Ce processus de "séparation-entraînement-validation" est répété plusieurs fois afin de trouver le modèle le plus optimisé (qui corresponde à l'erreur finale la plus petite). La Figure 1.21 montre un exemple d'un bon choix du modèle qui donne des erreurs petites entre les valeurs originales du GV et les valeurs estimées par ce modèle en utilisant le GE.

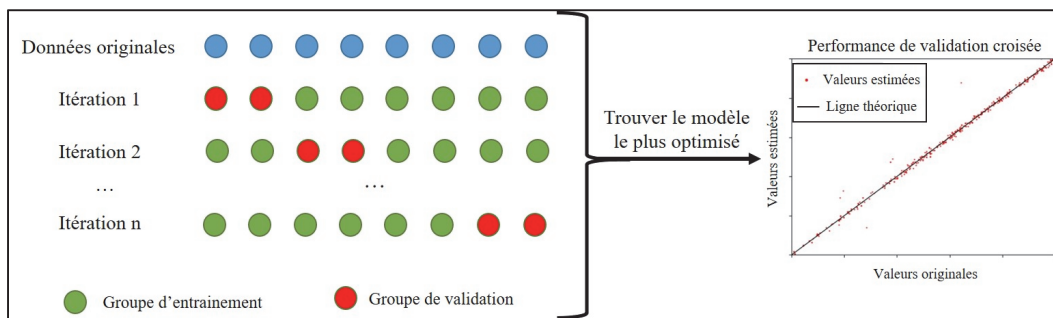


Figure 1.21 Un exemple de l'algorithme de la technique de validation croisée

L'application de la VC a été introduite la première fois par (Mosier, 1951) afin de vérifier la performance d'une régression linéaire. De temps en temps, l'application de la VC est élargie à plusieurs cadres de la sélection du modèle (Arlot & Celisse, 2010; Browne, 2000). En particulière, la VC contribue considérablement à la modélisation du modèle paramétrique du semivariogramme dans l'interpolation par krigeage (Ben Salah, 2014; Deutsch & Journel, 1992; D. Myers, 1993; Pebesma & Heuvelink, 2016). De plus, l'utilisation de la VC contribue à optimiser des paramètres du modèle et permet aussi de vérifier la sensibilité du nombre de données d'entraînement sur le modèle utilisé. Cela est une vérification importante

pour le cas d'un nombre limité des données comme les mesures de contraintes de la roue hydroélectrique.

1.3.3.3 Métrique de surface (Area Metric)

La métrique de surface (MS) est une métrique qui permet de valider un processus ou un modèle de prédiction en déterminant « l'écart » entre les prédictions et les valeurs originales. La motivation de l'MS est d'évaluer la similarité entre deux distributions entières des prédictions et des observations expérimentales (Y. Liu et al., 2011). Par conséquent, l'MS est principalement utilisée pour une mesure quantitative de la surface entre la fonction de répartition (CDF) F^* des prédictions et celle originale F :

$$d(F^*, F) = \int_{-\infty}^{\infty} |F^*(x) - F(x)| dx \quad (1.17)$$

De plus, l'MS peut aussi fonctionner quand il n'y a pas beaucoup de données disponibles, ce qui est souvent observé en pratique (par exemple, dans le cas de mesures expérimentales sur la roue hydroélectrique). Dans ce cas-là, la CDF des données est exprimée de façon empirique par une fonction en marche d'escalier (*Step Function*). Si les deux CDF à comparer ne représentent pas une distribution entière, la formule de l'MS peut être réécrite sous la forme discrète :

$$d(F_n^*, F_n) = \sum_{i=1}^{i_{max}} |F^*[x_i] - F[x_i]| \Delta x_i \quad (1.18)$$

L'MS ne donne pas seulement une valeur quantitative, mais aussi permet d'une validation graphique. La Figure 1.22 montre quelques exemples d'MS avec certains types de CDF à comparer. D'après Chatenet, Tahan, Gagnon, et Chamberland-Lauzon (2016), l'utilisation de l'MS pourrait donner plus d'information de validation qu'une analyse basée sur les deux premiers moments statistiques. Par exemple, pour les mesures de contrainte de la roue hydroélectrique, une des informations importantes à vérifier est la grande fluctuation

dynamique (les valeurs extrêmes) qui est souvent liée aux moments d'ordres supérieurs comme le kurtosis. L'utilisation de l'MS permet d'évaluer ces valeurs extrêmes en observant les écarts dans la zone de queue entre deux distributions.

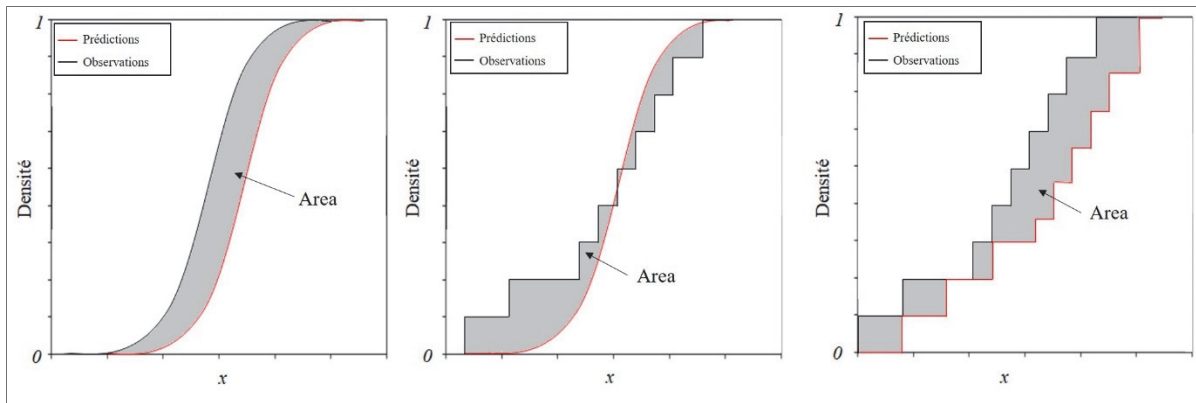


Figure 1.22 Exemples graphiques de la métrique de surface avec différents types de CDF

L'MS est une des métriques pratiques qui contribuent à l'étape de la validation du modèle dans plusieurs domaines d'ingénieur (Ferson, Oberkampf, & Ginzburg, 2008; Jung, Park, Oh, Kim, & Youn, 2015; Y. Liu et al., 2011). En particulier, l'MS est aussi utilisée pour la validation de la simulation numérique de contraintes à partir des mesures expérimentales sur une turbine Francis (Chatenet et al., 2016). La performance de la validation est de temps en temps améliorée via des recherches des extensions d'MS. Par exemple, la méthode de l'MS *u-Pooling* permet de déterminer l'écart entre des distributions dans le cas des multiples sites de validation (Ferson et al., 2008). Une méthode d'MS modifiée permet de séparer les zones de sous-estimation et surestimation entre deux CDF (Voyles & Roy, 2014). Une nouvelle méthode d'MS par intervalle permet de fournir en plus l'incertitude (l'intervalle des valeurs) causée par le manque de valeurs expérimentales (Wang et al., 2018). De notre point de vue, le concept de l'MS pourrait être élargie à d'autres types de fonctions que la CDF. Par exemple, une mesure des zones entre deux courbes du comptage Rainflow serait potentiellement une stratégie dans notre cas quand l'évaluation de fatigue est une des objectifs principaux.

1.3.3.4 Conclusion

Tableau 1.3 Avantages des différentes méthodes de validation

Méthodes	Avantages pour notre projet
Validation graphique	Une validation rapide (spécialement dans le cas de comparaison entre deux spectres rainflow)
Mesure des erreurs (REQM, EAM, etc.)	Des validations rapides et flexibles avec plusieurs types des erreurs
Validation croisée	Une validation robuste pour optimiser des paramètres et pour évaluer la sensibilité au nombre des données. Elle est couramment utilisée dans le krigeage qui est potentiellement la méthode principale d'interpolation dans notre projet
Métrique de la surface	Une validation robuste pour des distributions entières ou partielles. Peut être appliquée à un nombre de données limité Peut être utilisé pour d'autres types de comparaison comme les spectres Rainflow

L'étape de validation pourrait apparaître dans certaines phases du processus d'interpolation comme la validation finale du résultat interpolée, ou la validation du choix d'un paramètre. Par conséquent, nous n'utiliserons pas une méthode unique dans ce projet. Le Tableau 1.3 montre quelques avantages des méthodes mentionnées qui sont potentielles pour notre projet.

CHAPITRE 2

OBJECTIFS, PLAN ET CONTRIBUTIONS

2.1 Bilan de l'originalité

Par rapport à la revue de littérature, la prédiction de contraintes de la roue hydroélectrique est un réel défi scientifique et technologique. De plus, nous pouvons justifier la pertinence de notre projet par l'existence de problématiques qui attendent toujours une solution fiable et précise. Le grand défi, qui peut être relevé dans la revue de littérature, est que la qualité des modèles d'interpolation est directement affectée par le manque de données expérimentales dans un régime de fonctionnement donné. Par exemple, Poirier et al. (2017) ont résolu le problème de la limite temporelle des mesures expérimentales en réalisant une extrapolation dans le temps. Pour réaliser l'extrapolation, ils ont eu besoin d'au moins des mesures expérimentales sur une courte durée. Le problème est la même dans le cas de l'interpolation spatiale dans l'aube (Ben Salah, 2014), qui exige au moins des mesures sur quelques localisations.

Les simulations numériques par CFD peuvent fournir des informations sur plusieurs régimes d'opération. Leur précision dépend des hypothèses ou de modèles imposés. De plus, la vérification de ces simulations a besoin aussi de valeurs expérimentales. Malheureusement, ces mesures expérimentales ne couvrent pas toutes les plages d'opération de la turbine, ce qui correspond à notre objectif de recherche, qui est le développement d'un outil d'interpolation entre des régimes de fonctionnement de la turbine.

On peut dire que le présent projet de recherche a démarré dans un moment opportun où les données et les informations du domaine ont été beaucoup développées du côté scientifique, mais aussi du côté industriel. De plus, la famille des méthodes d'interpolation est de plus en plus diversifiée. Dans notre projet, le mot "interpolation" contient beaucoup d'échelles. Notre équipe tente d'utiliser des méthodes d'interpolation afin de résoudre des problèmes globaux comme la récupération des chargements dans de différents régimes d'opérations à partir des

données existantes. Plus spécifiquement, nous visons le développement d'un outil d'interpolation plus complet et plus détaillé permettant de l'interpolation de différentes caractéristiques du signal (cycles de fatigue, phénomènes physiques cachés, valeurs simulées...) et des interpolations dans plusieurs domaines (spatial, temporel, fréquentiel...).

2.2 Objectifs de recherche

Comme mentionné dans la problématique, l'objectif principal à long terme est d'améliorer notre connaissance du comportement dynamique de la turbine hydroélectrique pour améliorer l'estimation de la durée de la vie en fatigue de la roue. Plus précisément, ce projet se concentre sur la construction de modèle d'interpolation des signaux (contraintes, déformations) sur tous les régimes d'opérations permanents de la turbine en utilisant les mesures expérimentales. Pour notre projet, nous identifions les objectifs spécifiques suivantes :

- identifier des caractéristiques importantes du signal à préserver pendant l'interpolation comme :
 - caractéristique de fatigue (ex. comptage de rainflow...),
 - phénomènes physiques cachés dans le signal,
- développer des processus d'interpolation correspondant aux caractéristiques intéressantes,
- améliorer le processus pour maximiser la qualité de l'interpolation.

2.3 Plan de la thèse

Pour clarifier le chemin de recherche et aussi la continuité de cette thèse, la Figure 2.1 montre un plan général de l'organisation de ce rapport. Le chapitre "Introduction" résume la problématique de la thèse et aussi les liens avec nos partenaires industrielles. Une revue de littérature a été présentée dans le Chapitre 1. Ensuite, à partir de cette revue, l'objectif et l'originalité de la recherche sont détaillés dans le Chapitre 2. Les Chapitres 1 & 2 ont permis de mettre en évidence le besoin d'un outil d'interpolation entre des différents régimes de

fonctionnement de la turbine hydroélectrique. Cette conception de l'interpolation pourrait fournir un potentiel pour améliorer la fiabilité en fatigue de la roue. De plus, le Chapitre 1 fournit aussi une revue de littérature sur des méthodes scientifiques potentielles pour chaque étape nécessaire d'une conception générale de l'interpolation. Cette revue ne parle pas d'un travail particulier, mais il donne une vision générale à partir de laquelle on pourra créer une stratégie d'interpolation adaptant chaque tâche spécifique qui est présentée dans les articles scientifiques.

Le Chapitre 3 présente un article scientifique publié dans la revue *Renewable Energy*. Le processus proposé permet d'estimer directement l'information sur la fatigue en interpolant la matrice de comptage de cycles de fatigue.

L'interpolation telle que proposée nous permet d'obtenir directement l'information de fatigue, mais ne permet pas d'observer les phénomènes physiques impliqués dans un signal de contrainte. On propose ensuite dans le Chapitre 4 un nouveau processus qui permet d'interpoler indépendamment chaque phénomène physique caché dans un signal de contrainte de la roue. L'extraction de ces phénomènes est réalisée en utilisant les opérateurs de cyclostationnarité.

Les Chapitres 3 & 4 nous aident à créer un outil d'interpolation cohérent. Ensuite, la motivation est d'essayer d'améliorer la qualité de la performance de cet outil. Le résidu, après avoir extrait tous les phénomènes, contient la partie stochastique du signal. Cette partie est reconstruite de manière aléatoire, ce qui conduit à un manque de connaissance du comportement stochastique (ce qui est important surtout dans les régimes de petite charge). Par conséquent, le Chapitre 5 propose une combinaison de simulation/interpolation qui permet de générer la partie aléatoire en utilisant la transformée de Fourier à court terme (TFCT) et les courbes de cycle Rainflow.

Finalement, le mémoire est parachevé par une conclusion générale qui reprend les principales contributions et par les recommandations pour les futures perspectives de recherche.

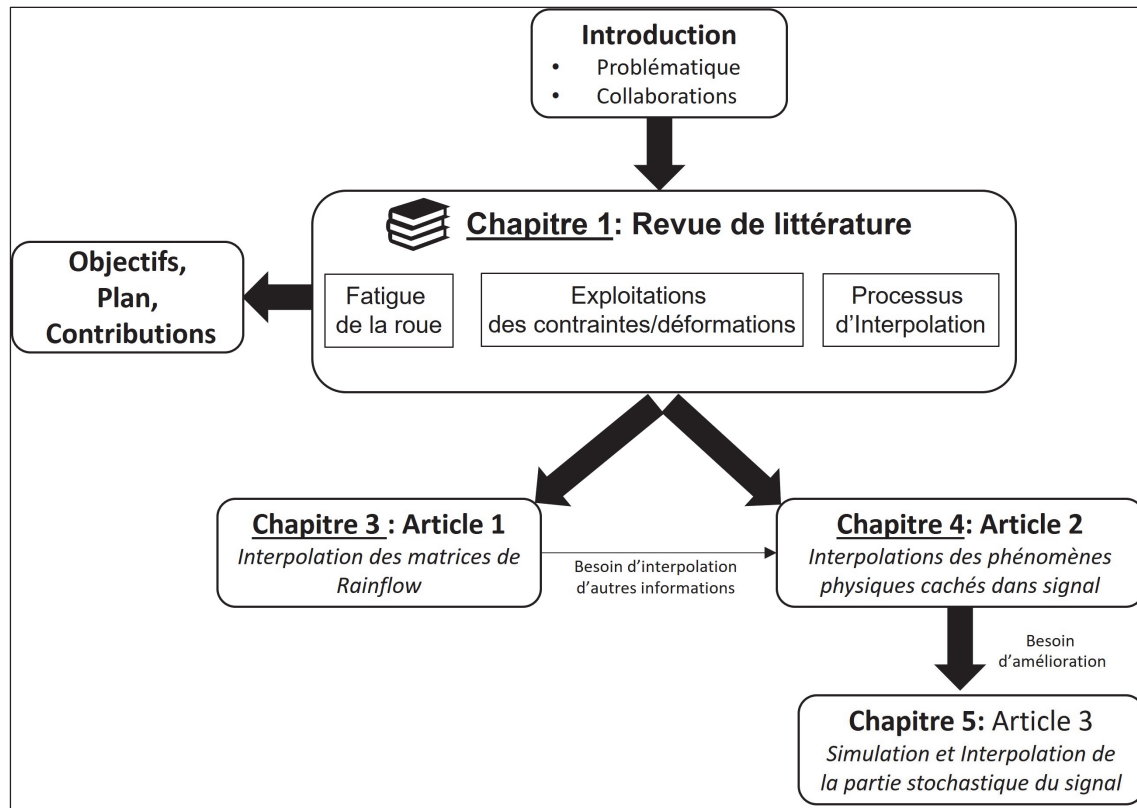


Figure 2.1 Structure du plan de la thèse

2.4 Contributions scientifiques et industrielles

Les travaux effectués dans cette thèse aboutissent à un outil original de l'interpolation entre différents régimes de fonctionnement pour la turbine hydroélectrique. Les contributions scientifiques au domaine hydroélectrique sont :

- un processus original permettant d'interpoler et de générer la matrice de comptage des cycles de fatigue d'un signal de contrainte de la roue est proposé et validé. Ce travail ouvre un accès rapide et direct aux informations importantes dans l'étude de la fatigue. Cette contribution correspond au contenu d'un article de journal publié dans *Renewable Energy* (Pham, Gagnon, Antoni, Tahan, & Monette, 2021) ;
- un processus d'interpolation permettant de reproduire les phénomènes physiques (les périodicités, le phénomène de torche, la stochasticité) contenus dans un signal de contrainte de la roue. Ce travail permet de générer un signal de contrainte dans un régime

de fonctionnement inconnu (non mesuré). Cette contribution est présentée dans un article de conférence SURVISHNO à Lyon (Pham et al., 2019), dans un article de conférence IAHR à Lausanne (2021) et un article de journal qui est soumis pour publication à *Renewable Energy* (depuis le février 2021).

Après avoir traité les étapes importantes pour l'outil d'interpolation comme mentionnées au-dessus, la partie suivante de cette thèse concerne l'amélioration de la qualité de l'interpolation afin de reproduire des informations plus riches pour la prédiction. Voici la contribution scientifique correspondante :

- une simulation stochastique permettant de régénérer la partie stochastique cachée dans le signal de contrainte à partir de la TFCT et des spectres rainflow a été développé et validée. L'application de l'interpolation sur les paramètres de cette simulation permet d'obtenir le signal estimé le plus proche du signal réel. Ce travail correspond au contenu d'un article de journal soumis pour publication en 2021.

Finalement, en marge des travaux scientifiques, voici quelques autres contributions industrielles (spécialement pour les deux partenaires de ce projet) :

- contribution pour réduire l'exigence de mesures expérimentales (ou simulations numériques par CFD) afin de diminuer les coûts;
- contribution pour une planification fiable des périodes d'inspection (maintenance) de la roue, qui est une étape importante dans la phase d'exploitation de la turbine hydroélectrique;
- contribution pour optimiser le dimensionnement (durant la phase de conception de la turbine) de la roue, dont la fiabilité en fatigue est améliorée.

CHAPITRE 3

RAINFLOW-COUNTING MATRIX INTERPOLATION OVER DIFFERENT OPERATING CONDITIONS FOR HYDROELECTRIC TURBINE FATIGUE ASSESSMENT

Quang Hung Pham^{a, b}, Martin Gagnon^c, Jérôme Antoni^b, Antoine Tahan^a, Christine Monette^d

^a École de technologie supérieure (ÉTS), 1100 Notre-Dame Ouest,
Montréal, Québec, Canada H3C 1K3

^b Université de Lyon, INSA Lyon, LVA, EA677, Bâtiment St Exupéry, 25 bis Avenue Jean
Capelle, 69621 Villeurbanne Cedex, France

^c Institut de recherche d'Hydro-Québec (IREQ), 1800 Boulevard Lionel-Boulet,
Varenes, Québec, Canada J3X 1S1

^d Andritz Hydro Canada Inc, 6100 Route Transcanadienne, Point-Claire,
Québec, Canada H9R 1B9

Article publié dans « Renewable Energy » en mars 2021

3.1 Contribution à la thèse

La fatigue de la roue a été mentionnée plusieurs fois comme un mot-clé dans cette thèse. L'évaluation de la fatigue devient une tâche importante à laquelle cette thèse prétend contribuer. Donc, la première idée est "d'attaquer" directement sur le(s) caractères(s) de fatigue du signal de contrainte. La nature physique de la fatigue est la propagation de fissure causée par un grand nombre de cycles de chargements cumulés. Cela nous conduit à la modélisation du comptage de cycles de contrainte. Telle que présentée au Chapitre 1, la technique de rainflow est une des bonnes méthodes pour représenter le comptage de cycles d'un signal et elle est souvent utilisée dans les recherches de la turbine hydroélectrique. En effet, si l'on utilise ces matrices de comptage rainflow comme les valeurs d'entrées de l'interpolation, on pourrait profiter de quelques avantages :

- cette matrice est plus facile à interpoler, comparée à un signal de contrainte complet ;
- cette matrice permet d'accéder directement au terme de fatigue sans besoins d'information sous la forme d'un signal temporel ;

- l'accès à la modélisation de cette matrice de comptage n'est pas difficile quand il y a des techniques pertinentes et des recherches liant à cette matrice dans l'état de l'art.

Par conséquent, l'objectif de l'article présenté dans ce chapitre est de développer un processus pour interpoler des matrices de comptage rainflow entre différents régimes de fonctionnement permanent de la turbine hydroélectrique. En effet, via cet article, nous visons à :

- un processus complet permettant de reconstruire le comptage de cycles de contrainte dans un régime non mesuré en utilisant des régimes mesurés,
- une technique d'interpolation multidimensionnelle (deux axes de matrice et un axe de plage de régimes), qui est inspirée par le krigeage spatio-temporel, permettant de construire une matrice complète dans un régime inconnu,
- une technique de simulation après l'interpolation permettant de générer les cycles de contraintes pour une évaluation de fatigue à long terme.

3.2 Résumé de l'article en français

Les contraintes mesurées sur l'aube jouent un rôle important dans l'évaluation de la fatigue de la roue d'une turbine hydroélectrique. Pourtant, à cause des limitations de mesures à la centrale, ces mesures expérimentales ne peuvent pas couvrir tous les régimes de fonctionnement possible, ce qui conduit à un manque dans l'évaluation de la fatigue pour les conditions non mesurées. Dans cet article, la méthode de krigeage est utilisée afin d'estimer les informations de fatigue manquantes en utilisant les mesures expérimentales disponibles dans des régimes de fonctionnement de la turbine hydroélectrique. La matrice de comptage Rainflow, qui contient les cycles de fatigue du signal de contrainte, est considérée comme l'entrée du processus d'interpolation. Pour la validation, les cycles de fatigue sont aléatoirement générés, en utilisant les distributions gaussiennes, à partir de la matrice interpolée afin de comparer ensuite aux valeurs expérimentales dans un même régime de fonctionnement. L'approche développée aidera à planifier les mesures à la centrale et à contribuer à la réduction de nombre de mesures nécessaires pour l'évaluation de fatigue.

3.3 Abstract

The measurement of stress/strain on blade plays an essential role in the fatigue assessment of the hydroelectric turbine runner. However, a typical measurement campaign does not cover all the possible operating conditions due to the field measurement requirements and limitations, which leads to fatigue evaluation missing for unmeasured conditions. In this paper, kriging interpolation method is used to estimate the missing fatigue information based on the available measured operating conditions of the turbine. The rainflow counting matrix, which contains the fatigue cycles of strain signal, is considered as the input of the interpolation process. For validation, the fatigue cycles are randomly generated, using bivariate Gaussian distributions, from the interpolated matrix for comparison with the experimental values under similar operating conditions. The developed approach will help plan experimental campaign and might contribute to decreasing the measurement requirements for runner fatigue assessment.

3.4 Introduction

Fatigue is one of the damage mechanisms occurring during the operation of a hydroelectric turbine. Today, runner blades are experiencing increasing dynamic strains due to larger operating range of turbines caused by electrical network usage changes (Monette et al., 2016). These changes mean that the turbine must operate under conditions that deviate significantly from the Best Efficiency Point (BEP), which leads to a higher risk of fatigue crack. Monette et al. (2016) highlighted the reduced fatigue life of turbine runner and the high repair costs generated by this enlarged operating range. A general overview of conditions and dynamic phenomena that significantly influence runner fatigue is presented by Gagnon et al. (2013); X. Liu et al. (2016) and Seidel et al. (2014). Steady state operating conditions containing a high stochastic component such as Speed No-Load (SNL) or Deep Part-Load (DPL) or Part load containing vortex rope phenomenon, and transient conditions such as turbine start-up can have major impact on runner fatigue assessment if not properly accounted for.

Fatigue is assessed based on several factors (runner material, stress level, initial crack), and requires a representative strain/stress history. Many approaches are available for improving runner fatigue life estimation. To limit fatigue failure, (Gagnon et al., 2013) proposed a runner reliability model using the High Cycle Fatigue (HCF) onset as the limit. Fatigue damage is linked to the propagation of cracks on the runner, and thus, an investigation of the metallurgical or mechanical behaviour of the material is also needed (El Haddad, Topper, & Smith, 1979; Kitagawa, 1976; Thibault, Gagnon, & Godin, 2014; Trudel & Sabourin, 2014). However, fatigue assessment cannot be carried out if the dynamic load information is missing. Computational Fluid Dynamic (CFD) combined with Finite Element Analysis (FEA) can be used to estimate the static and dynamic behaviour of the structure under such as SNL or DPL (Morissette et al., 2016; Nennemann et al., 2014; Xiao, Wang, & Luo, 2008). Nevertheless, numerical prediction is still challenging when turbulence models are not able to properly describe the behaviour of the turbine (Presas et al., 2019). Moreover, the validation of numerical simulation depends on the availability of the experimental information. However, to recover this missing experimental information, other approaches, such as the interpolation and extrapolation of experimental data, can be used (Ben Salah, 2014; Pham et al., 2019; Poirier et al., 2017). The use of indirect method can also help minimize the measurement requirement, Diagne et al. (2016) utilised the ARMAX (Autoregressive Moving Average with eXogenous input) model to predict the runner dynamic stress using shaft torsion information, which is more readily available. For a detailed review of the fatigue life estimation methodologies used for Francis turbines, please refer to Presas et al. (2019). Our state-of-art summary shows that in situ field strains/stress during or after commissioning provide important and essential information for the study of the hydroelectric turbine runner fatigue.

Finding a solution that can provide missing experimental information and improve knowledge on the dynamic behaviour remains a challenge. Obtaining experimental strains over all the possible states of the hydroelectric turbine operation would be overly costly and time-consuming; hence, experimental measurements are usually limited to a limited set of operating conditions, which leads to difficulties during runner fatigue assessment. The goal

of this paper is to propose a methodology to minimize the requirements for experimental strain measurements. There are three difficulties associated with the measurement of strain for fatigue evaluation:

- limited number of measured locations on runner blades,
- limited measurement length,
- limited number of measured operating conditions.

The first was studied by Salah where the kriging method was used for the interpolation between different locations of the runner blade (Ben Salah, 2014). For the second, several approaches can be used for temporal extrapolation to obtain long-term fatigue assessments. Poirier et al. (2017) extrapolated information on the cyclostationary decompositions of the short-term records of strain signals to evaluate runner fatigue cycles over a longer period of usage. Other researchers extrapolated the rainflow matrices directly over longer periods (Pär Johannesson, 2006; Pär Johannesson & Thomas, 2001; D. Socie, 2001; D. F. Socie & Pompetzki, 2004). However, such strain interpolation between blade locations or a temporal extrapolation can be performed if and only if an estimate of the signal is available (generally in the form of measurements) at a given turbine operating condition. Therefore, the lack of measurements over the complete operating range should be studied as shown in Figure 3.1.

In a previous work the authors developed an interpolation process for a subcomponent of the strain signal between different turbine operating conditions (Pham et al., 2019). Only the periodic component of the signal was interpolated while the complete strain signal was overlooked as it is more complex and contains other hidden structures. The direct interpolation of complete strain signal (measurement form) is a complicated problem because of these hidden structures. However, in cases where only fatigue information is needed for further evaluation, the strain signal could be transformed into a cycle counting form (level crossing, peak counting, rainflow matrices, etc.) to simplify the interpolation. Among cycle counting methods, the rainflow algorithm is commonly used for fatigue assessment of hydroelectric turbines (Monette et al., 2016; Morissette et al., 2016; Nennemann et al., 2014; Pham et al., 2019; Poirier et al., 2017). Therefore, our goal is to interpolate the rainflow

counting matrices (instead of the signal itself), which contain information on the fatigue cycles, over the complete range of operating conditions. Notice that there are two important variables that define the operating conditions of turbine: opening vane level and head level. However, in this study, the aim is to focus on the differences of stress level between different opening vane levels for the same head (the data in the case study also comes from the measurements strain at relatively constant head). The head level might be used in the future research if observed data become available for multiple head levels.

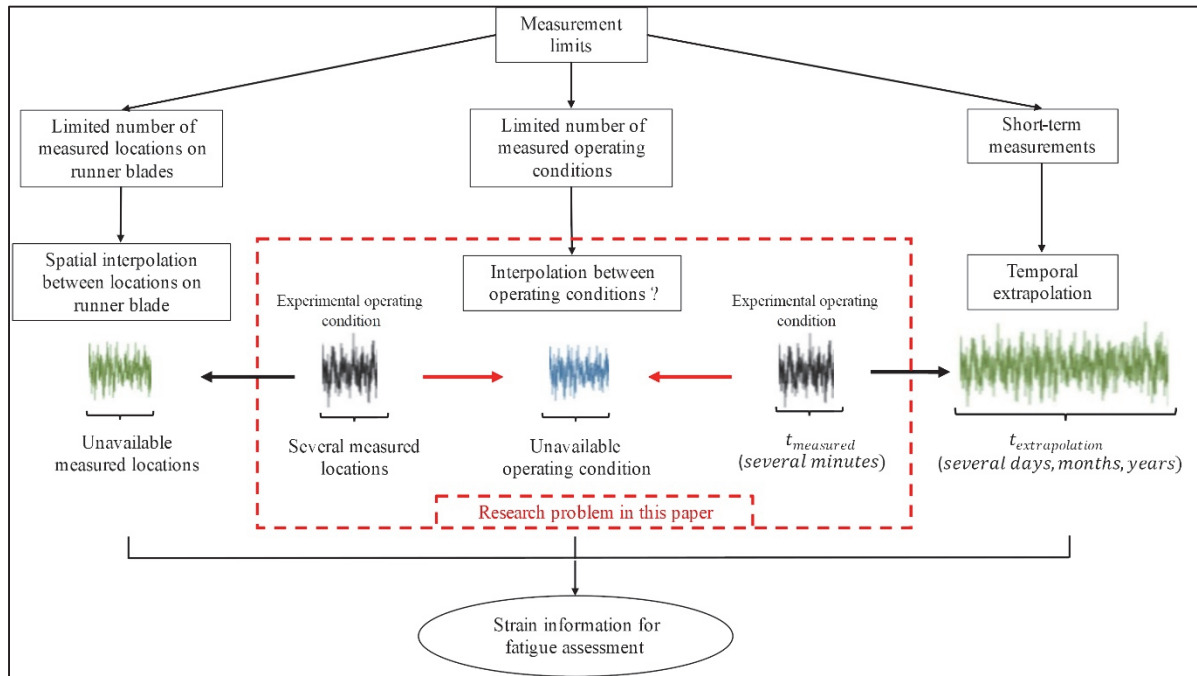


Figure 3.1 Measurement limits and interpolation/extrapolation solutions for the fatigue assessment of hydroelectric turbine runner

The kriging method was chosen for this study to interpolate rainflow counting matrices (histogram form). Then, the fatigue cycle reconstruction was generated via stochastic simulations using bivariate Gaussian distributions. This reconstruction step helps recover the long-term information of fatigue, more cycles are generated, more fatigue information is evaluated. The obtained results were compared with the experimental fatigue cycles to evaluate the correctness of the proposed interpolation methodology. In addition, the signal

was reconstructed from the simulated cycles by using the inverse method proposed by D. Socie (2001) to compare with the measured signal.

The paper is organized as follows: First, the methodology is presented with detailed explanations of the principal methods (rainflow-counting, kriging interpolation, reconstruction simulation). Then, a hydroelectric turbine runner case study is presented followed by interpolation results and a discussion.

3.5 Proposed Methodology

3.5.1 Asymmetric Rainflow Counting Matrices

The rainflow algorithm allows the analysis of signal amplitudes (or range) by assigning extreme (max, min) values of closed loading cycles. To evaluate fatigue, only the range/amplitude information of each cycle is often needed. In this study, the strain signal is transformed into an Asymmetric Rainflow Matrix (ARM), which separates the cycles into two types: increasing and decreasing cycles (or standing and hanging cycles in some references) (Par Johannesson, 1999). These two types of cycles are defined based on the order of occurrence of maximum and minimum values in the signal. To construct the ARM, a rainflow-counting procedure called the Four-Point Counting (FPC) algorithm is used (Dreßler et al., 1997). Considering four consecutive points (three consecutive cycles) in the loading signal, if the cycle range created by the second and the third points is smaller than the range of the two other points, this cycle is reduced and is added to the ARM. An ARM is constructed by two variables: from value (noted as f) corresponding to the amplitude of the first point of the cycle (the first point of the cycle arrives on the axis of time) and to value (noted as t) corresponding to the amplitude of the final point of the cycle (Figure 3.2). In our study, the created ARM is represented under histogram density form where each bin contains the number of cycles or the probability of occurrence of cycles. The ARM is a composition of max-min cycles (above the matrix diagonal) and min-max cycles (below the matrix diagonal). Furthermore, the choice of the asymmetric rainflow matrix allows a distinction to be made between increasing and decreasing cycles during the signal reconstruction. To

properly apply the FPC, the turning points, which represent the peak and the valley of the signal, are extracted before the ARM is constructed. This extraction deletes variations that are neither peaks nor valleys (these small variations in the rainflow approach are considered to have minor impact on the fatigue assessment).

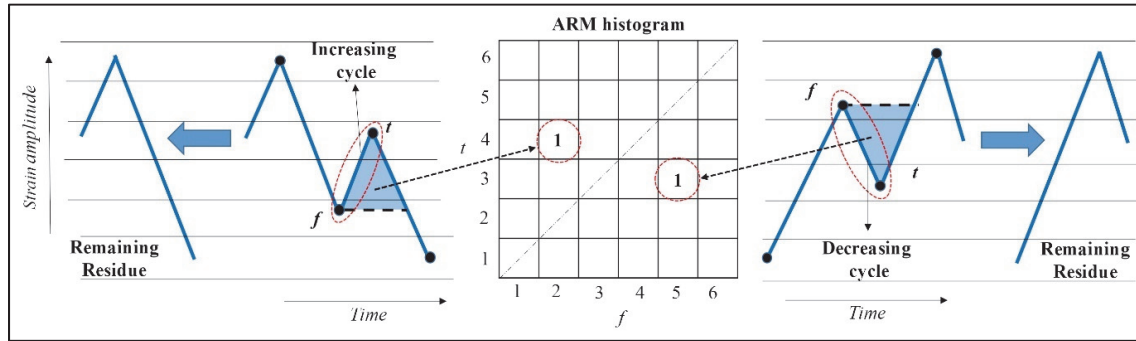


Figure 3.2 Four Point Counting algorithm applied for strain signal

3.5.2 Kriging Interpolation Method

Before presenting the interpolation method, the context must be clarified. The interpolation input is a set of ARM histograms, which are spread across the different turbine operating conditions. The histogram bins are fixed for all the histograms. The goal is to estimate new ARM histograms at any given operating conditions (see Figure 3.3).

The relation between the bins of each ARM matrix is an important bit of information that cannot be ignored in the interpolation phase. Hence, the kriging interpolation method is a suitable method to preserve the spatial covariance within the ARM histogram. However, there is also a covariance to preserve between each operating condition. Thus, a multivariate kriging method, inspired by Spatio-Temporal Kriging (STK), where the temporal and spatial coordinates are treated separately, is applied, and developed for this study. Readers can consult the general definition of the kriging principle in (Deutsch & Journel, 1992; Goovaerts, 1997; Montero et al., 2015; Pebesma & Heuvelink, 2016).

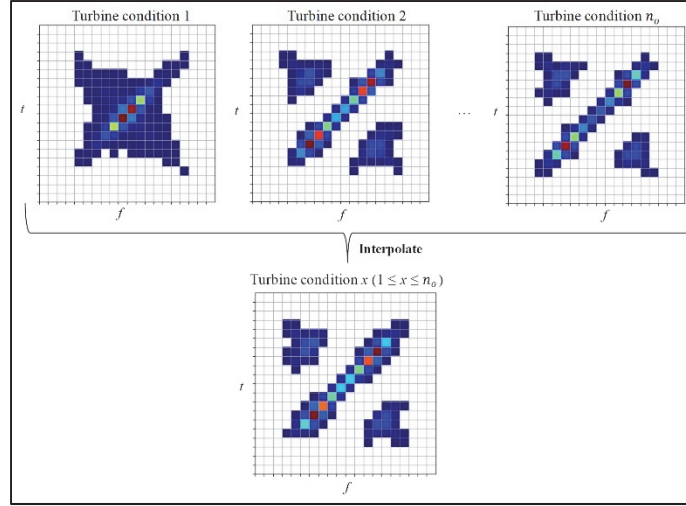


Figure 3.3 Interpolation objective: new ARM histogram (at a given turbine operating condition) is interpolated from all experimental ARM histograms

In our case, the position of bins in the ARM histogram and the operating condition are considered as coordinates. The estimator used in this study can be written as:

$$B^*(p_i, o_x) = \sum_{k \in \vartheta_i} \sum_{l=1}^{n_o} \lambda_{kl}(p_i, o_x) B(p_k, o_l) \quad (3.1)$$

where p_i is a given bin position in the ARM histogram (p depends on a ‘from value’ f_i and a ‘to value’ $t_i : p_i(f_i, t_i)$), o_x is a given operating condition of the turbine, $B^*(p_i, o_x)$ is the interpolated value of bin p_i at the condition o_x , $B(p_k, o_l)$ is the known bin value at location (p_k, o_l) , $\lambda_{kl}(p_i, o_x)$ is the regression weight to optimize during the interpolation process (details below), n_o is the number of experimental turbine conditions, ϑ_i is a set of neighbourhood positions of p_i (including p_i). It must be noted that the spatial covariance map is obtained from the whole histograms of all the experimental turbine conditions, while the equation (3.1) searches for the input in the neighbourhood ϑ_i of the target bin. This reduces the calculation time and avoids the interpolation smoothing effect.

Kriging allows the unbiased estimation of values at given locations by determining the weights λ_{kl} which minimize the error variance $\sigma_e^2(p_i, o_x)$ between the real value B and the interpolated value B^* at the same location, $\sigma_e^2(p_i, o_x) = \text{Var}(B(p_i, o_x) - B^*(p_i, o_x))$. The condition for non-bias relating to the regression weights is $\sum_{k \in \vartheta_i} \sum_{l=1}^{n_o} \lambda_{kl}(p_i, o_x) = 1$; it ensures a constant local mean value in every location. Therefore, to impose this non bias constraint, the minimization system is updated by adding a Lagrange parameter μ :

$$L(\lambda_{kl}, \mu) = \underbrace{\text{Var}(B(p_i, o_x) - B^*(p_i, o_x))}_{\sigma_e^2(p_i, o_x)} + 2\mu \left(\sum_{k \in \vartheta_i} \sum_{l=1}^{n_o} \lambda_{kl}(p_i, o_x) - 1 \right) \quad (3.2)$$

where $L(\lambda_i, \mu)$ is the Lagrange function. Making use of the equation (3.1), the minimization of the equation (3.2) leads to the system that allows to obtain the λ_{kl} and μ :

$$\begin{cases} \sum_{k \in \vartheta_i} \sum_{l=1}^{n_o} \lambda_{kl}(p_i, o_x) \gamma(h_{\|p_k - p_j\|}, u_{|o_l - o_m|}) - \mu = \gamma(h_{\|p_i - p_j\|}, u_{|o_x - o_m|}) \\ \sum_{k \in \vartheta_i} \sum_{l=1}^{n_o} \lambda_{kl}(p_i, o_x) = 1 \end{cases} \quad (3.3)$$

where $j \in \vartheta_i, m = 1, \dots, n_o$ and:

$$\gamma(h_{\|p_a - p_c\|}, u_{|o_b - o_d|}) = 0.5 \times \text{Var}[B(p_a, o_b) - B(p_c, o_d)] \quad (3.4)$$

is the semivariogram, which depends on the increments h and u , p_a and p_c are two positions of the bin in the histogram, o_b and o_d are two turbine operating conditions such that: $h = \|p_a - p_c\|$ and $u = |o_b - o_d|$. Based on the equation (3.4), the semivariogram is defined as the half variance between two values separated by the increments h and u . This semivariogram can be considered as a simpler representative of the covariance map that does not require information on the mean value of B (that is often unknown in practice (Montero et al., 2015)). To obtain the semivariogram at any possible increment (h, u) , some numerical

models can be used to fit to the semivariogram using analysis estimations (e.g., Least Squares method) (Deutsch & Journel, 1992; Pebesma & Heuvelink, 2016). However, to choose the best numerical model, one needs to create a semivariogram from the experimental data. In practice, the quantity of data is not enough to provide values exactly separated by $h = \|p_a - p_c\|$ and $u = |o_b - o_d|$, as in the equation (3.4). Consequently, to create the Experimental Semivariogram (ES), the equation (3.5) is used. This equation allows an estimation of the semivariogram value associated with a set of (h, u) values, where h and u are midpoints of imposed spatial increment intervals:

$$\hat{\gamma}(h, u) = \frac{1}{2N(h, u)} \sum_{(a,b,c,d) \in v(h,u)} [B(p_a, o_b) - B(p_c, o_d)]^2 \quad (3.5)$$

where $\hat{\gamma}(h, u)$ is the experimental semivariogram, $v(h, u)$ is a set of paired points separated such that $\|p_a - p_c\| \in [h - dh, h + dh]$ and $|o_b - o_d| \in [u - du, u + du]$ with $dh > 0$, $du > 0$, $N(h, u)$ is cardinal of $v(h, u)$. The next step is to fit the ES with a numerical model that will make a continuity throughout the increments for the semivariogram. Classes of semivariogram models are presented in Montero et al. (2015); Pebesma et Heuvelink (2016). For example, the sum-metric model (Figure 3.4), which turns out as one of the best fitting models in our study, is expressed as:

$$\gamma_{Smetric}(h, u) = \gamma_p(h) + \gamma_o(u) + \gamma_{joint}(\sqrt{h^2 + (\kappa \cdot u)^2}) \quad (3.6)$$

where $\gamma_p(h)$ is the univariate semivariogram model for the variable p , $\gamma_o(u)$ is the univariate semivariogram for the variable o , and $\gamma_{joint}(\sqrt{h^2 + (\kappa \cdot u)^2})$ is a metric joint variogram model including an anisotropy correction κ (Pebesma & Heuvelink, 2016). Many basic analytical models exist for the univariate semivariograms such as spherical model, the exponential model, the Gaussian model, power model, etc. The schema in Figure 3.5 summarizes the kriging interpolation approach used in our case study.

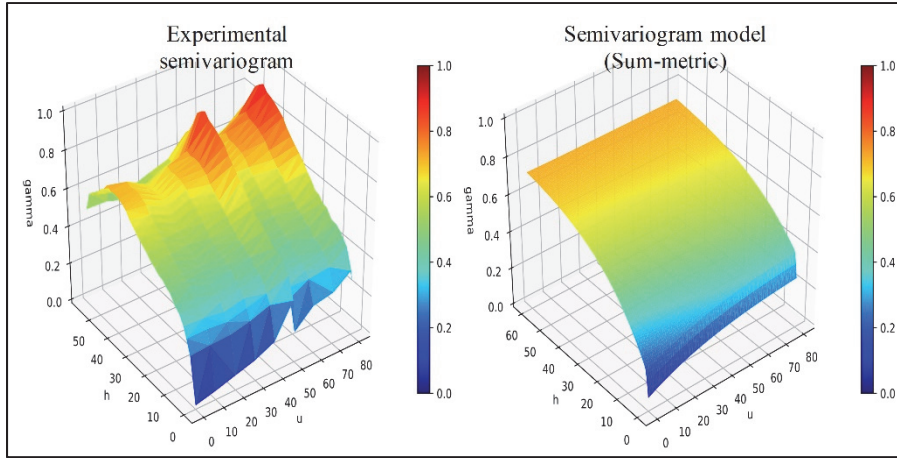


Figure 3.4 Experimental semivariogram and example of semivariogram model - Sum metric (Semivariogram values in this figure are normalized)

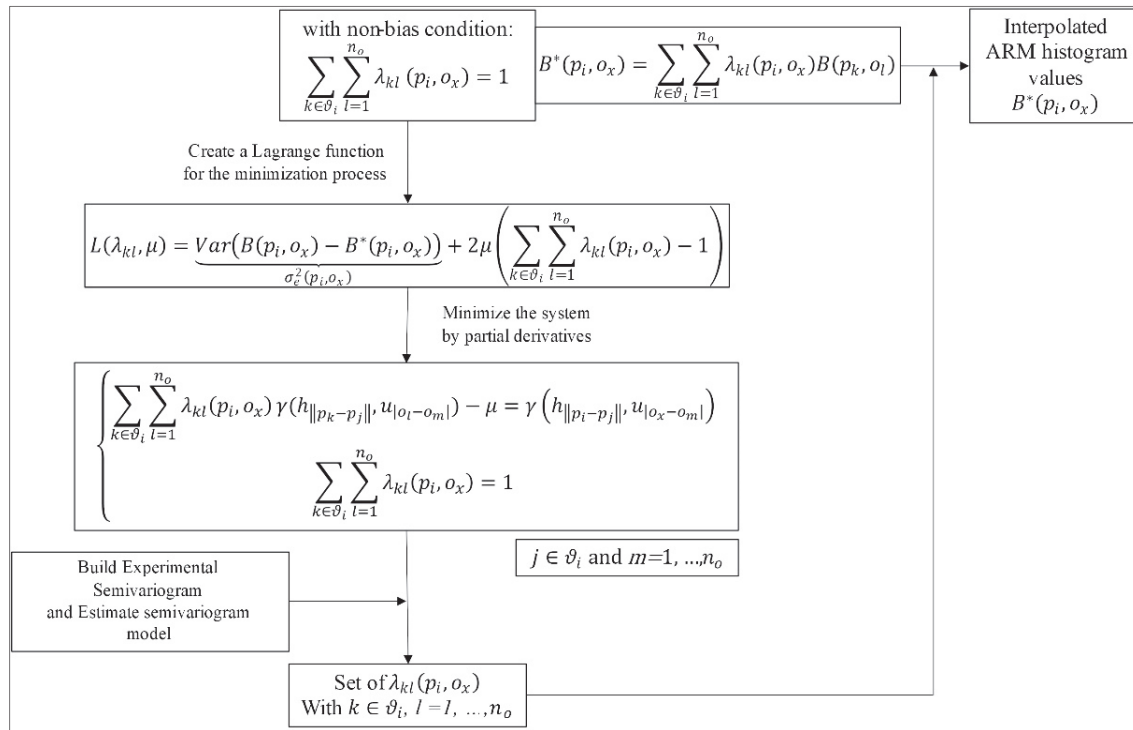


Figure 3.5 Schema of the Kriging process in this paper applied for the studied variables

3.5.3 Simulation Reconstruction

To generate strain cycles from an interpolated ARM histogram, it must be converted to a continuous distribution. A smoothing process for the interpolated histogram is proposed by considering that the cycles inside each bin follow a Bivariate Gaussian Distribution (BGD). The BGD mean corresponds to the coordinates of the center point of the bin and the proposed covariance matrix is $\sigma^2 \mathbf{I}_2$, where $\sigma = \frac{0.5 \times \text{bin diagonal}}{\text{variability parameter}}$ and $\mathbf{I}_2$ is the identity matrix. The variability parameter in the denominator is a coefficient proposed to ensure that the simulated cycle distribution does not create too much variability outside its bin. A variability test is proposed to select this variability parameter. In this study, since the interpolation input comes from the experimental strains measured by gauges, we impose the strain gauge uncertainty as a limited level of variability, which limits the variability of generated point outside each bin. For example, if the relative error of the strain gauge is $\pm y\%$, the imposed limited level for the variability will be $y\%$. It must be noted that the experimental uncertainties are not the focused problem of this study. The test of variability is then launched for several variability parameters. For each ARM, Monte Carlo simulation is generated from the bivariate normal distribution with the mean $[0,0]$ and the covariance $\sigma^2 \mathbf{I}_2$. If the percentage of the sample values outside the bin dimension is approximately equals to the limited level $y\%$ (Figure 3.6), this variability parameter is chosen for the ARM histogram smoothing process. The reader can also optimize the variability parameter by using or adding other uncertainty types (e.g., error of kriging interpolation) if needed for other study. In addition, other approaches, such as the Kernel Density Estimation (KDE), can also be useful for smoothing the histogram (D. Socie, 2001; D. F. Socie & Pompetzki, 2004). Finally, the cycles are randomly generated from the set of BGDs by respecting the probability of occurrence of cycles in each bin (the sum of all the probability values in the histogram is equal to 1). This generation process allows an observation of a set of simulations, which is useful for the industry when evaluating the fatigue. The risk of fatigue will be predicted with a set of possibilities instead of a unique prediction.

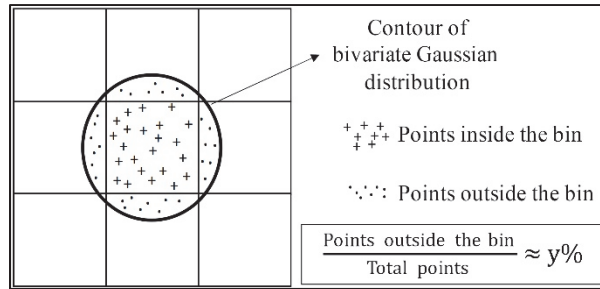


Figure 3.6 Percentage of the sample values outside the bin dimension is approximately equal to $y\%$

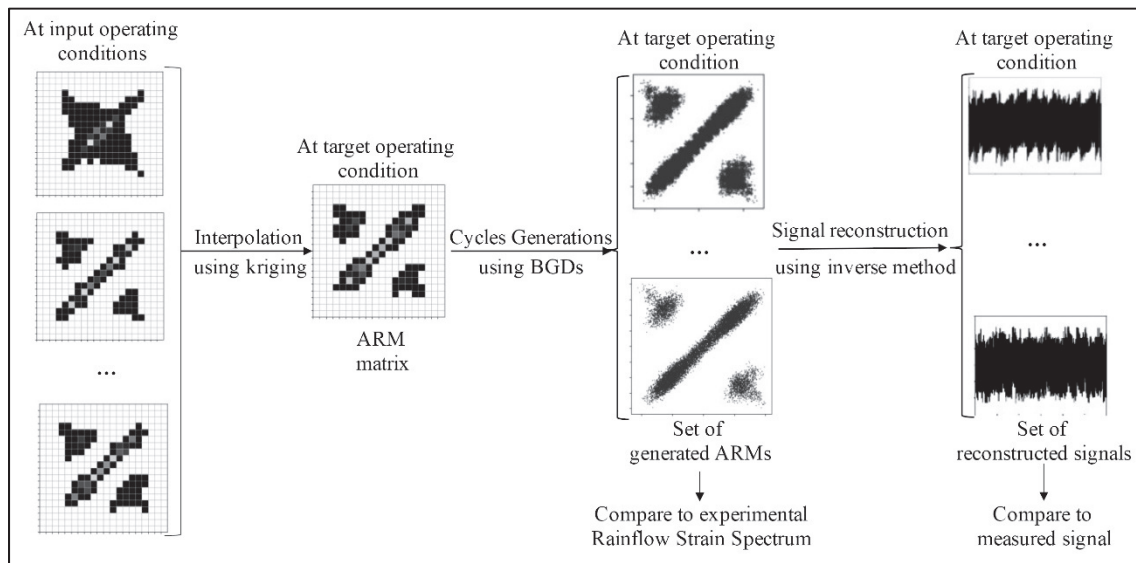


Figure 3.7 Flowchart displays the interpolation scheme

In addition, an inverse method is used to reconstruct the strain signal from the simulated samples (D. Socie, 2001; D. F. Socie & Pompetzki, 2004). The signal reconstruction consists of two steps: First, the largest cycles (the furthest points from the ARM diagonal), either a Decreasing Cycle (DC) or an Increasing Cycle (IC), are used to create an initial reconstructed signal (also called the *remaining residue* in Figure 3.2). Second, all the other cycles are randomly inserted into the remaining residue descending order (from the furthest to the closet point from the diagonal) by respecting the rule proposed by D. Socie (2001). The smaller IC is added to the position of the larger DC (peak and valley values of the smaller cycle fall within the range of the larger cycle) and vice versa. Each cycle added to the remaining

residue is removed in the ARM until there are no cycles left. It is not expected that the reconstructed signal will be similar to the experimental one because only the rainflow information is considered for the interpolation; nevertheless, the signal should represent a time history that generates equivalent fatigue damage. The flowchart in Figure 3.7 illustrates the interpolation scheme used in this study.

3.6 Case Study

The data used in this case study were obtained from experimental measurements conducted using strain gauges, which were installed on the runner blade of a Francis hydroelectric turbine of a power plant in Quebec (Canada). The measured locations on the blade are close to critical locations where the highest stress levels are found. For the purpose of this paper, the experimental uncertainties (ambiance error, human error, error of installation, etc.) are ignored. The interpolation is performed between different steady-operating conditions of the turbine. Transient operating conditions, such as start-ups or stops, are not considered in this research because their specific dynamic behaviours do not correlate with steady-operating conditions. A steady-operating condition is approximately defined as a regime of the turbine under which the guide vane opening, power and turbine head can be considered as quasi constant. For the interpolation, steady states are separated by different opening levels of the turbine guide vane (noted as %OV) (Figure 3.8). The guide vanes control the flow rate into the runner.

Due to the high quantity of cycles in each strain signal, the ARMs are presented under the histogram density form with a prior fixed bin number where each bin contains the probability of occurrence of cycles (Figure 3.8). The choice of input/output bin number is an important parameter. A higher number of bins significantly increases the calculation time, while a smaller one leads to imprecise results because of the large interval of values covered by each bin.

Fitting a model to the experimental semivariogram is an important step of the interpolation process. We highly recommend using the cross-validation technique for this step. This technique removes some point values in the input set and then predicts these values from the remaining points using fitted semivariogram models. This procedure is repeated several times for each semivariogram model to observe the interpolation performance of this model. Moreover, other error estimations such as Root Mean Square Error (RMSE) or Mean Absolute Error (MAE) are additionally used to obtain the best semivariogram model. In our case study, based on a good cross-validation performance (Figure 3.9) and on the small error estimation values (see MAE and RMSE in Tableau 3.1), the chosen semivariogram model was sum-metric model (see the equation (3.6) and Tableau 3.1). For each univariate model, $\gamma_p(h)$ was modeled by sum of an exponential model and a Gaussian model, $\gamma_o(u)$ and $\gamma_{joint}(\sqrt{h^2 + (\kappa \cdot u)^2})$ was respectively modeled by a Gaussian and an exponential model (Tableau 3.1).

$$\begin{cases} \gamma_{exponential}(h) = (Sill - nu) \cdot \left(1 - e^{-\frac{3 \cdot h}{range}}\right) + nu \\ \gamma_{Gaussian}(u) = (Sill - nu) \cdot \left(1 - e^{-\frac{(3 \cdot u)^2}{range^2}}\right) + nu \end{cases} \quad (3.7)$$

where *Sill* corresponds to the average variance where the experimental semivariogram stabilizes, *range* presents the distance from which no more correlations exist between the data, and *nu* is a small value added to the sill parameter to represent a very short-range variability in the dataset (Nugget effect) (Deutsch & Journel, 1992).

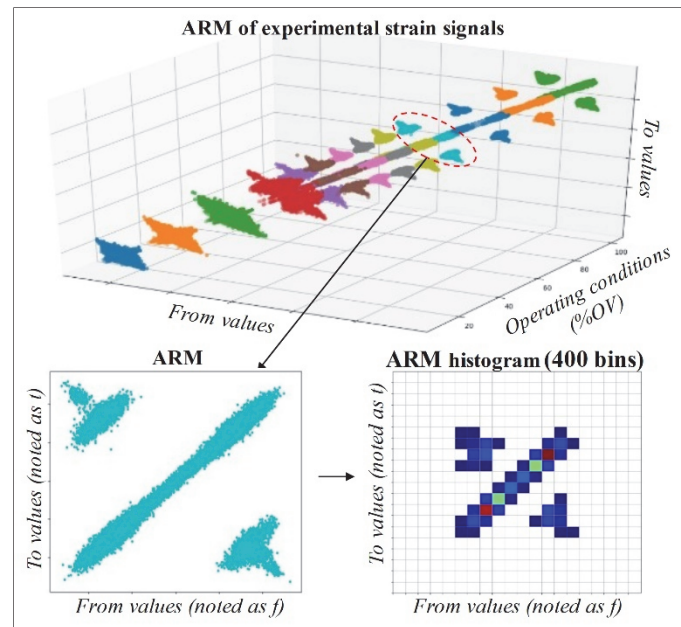


Figure 3.8 ARMs over different operating conditions for this case study and example of ARM histogram (the white bin corresponds to the null value)

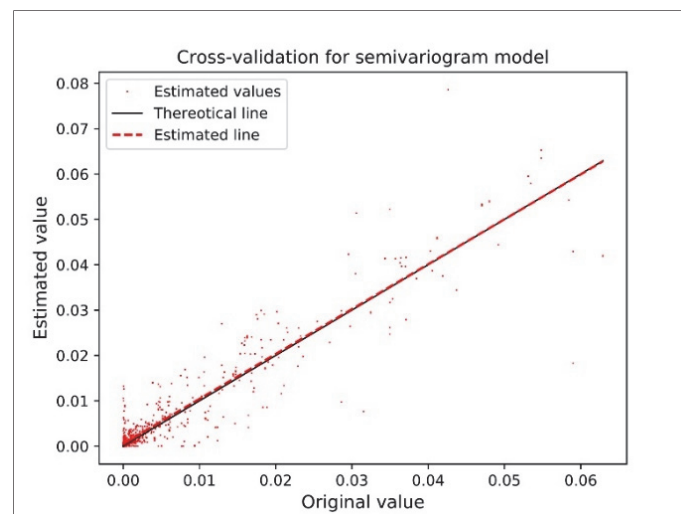


Figure 3.9 Cross validation performance of sum-metric model (its parameters are presented in Tableau 3.1)

Tableau 3.1 Parameters of the proposed interpolation process adapted for this case study

General Parameters	Semivariogram model parameters (fitted to ES as in Figure 3.4)					
Input/Output histogram bins number: 40×40 bins Input operating conditions (%OV): 11, 20, 40, 50, 60, 70, 90, 100 Variability para. ≈ 3.87 Number of points in neighbourhood set: 9	Optimized sum-metric model (RMSE = 0.0993, MAE= 0.0747) $\gamma_{Smetric}(h, u) = \gamma_p(h) + \gamma_o(u) + \gamma_{joint}(\sqrt{h^2 + (\kappa \cdot u)^2})$					
		Model type	Sill	nu	range	κ
	$\gamma_p(h)$	Gaussian	0.116	3e-08	2.011	0.296
		Exponential	0.376	0.009	70.025	
	$\gamma_o(u)$	Gaussian	0.040	0.042	30.007	
	$\gamma_{joint}(\sqrt{h^2 + (\kappa \cdot u)^2})$	Exponential	0.230	7e-08	60.057	

The neighbourhood set ϑ_i (the equation (3.1)) is chosen using the cross-validation technique. Several sets with different numbers of neighbourhood positions are tested and the optimized set is chosen based on the smallest RMSE and based on the similarity between theoretical and estimated line (see ANNEXE I). Notice that the set of neighbourhoods is applied only for the histogram position “ p ”, but not for operating conditions “ o ”. Thus, the points in a neighbourhood set correspond to the closest points from the target points in ARMs. In this case study, the set of 9 nearest positions is chosen for the neighbourhood set ϑ_i in the equation (3.1) (ANNEXE I). Tableau 3.1 shows the optimized parameters used for this case study. To verify the quality of the interpolated result, a validation is made using known experimental operating conditions (not included in the set of interpolation inputs).

3.7 Results and Discussions

The interpolated results (corresponding to each step illustrated in Figure 3.7) and their comparison with experimental ones at 55% OV are presented in the Figure 3.10-Figure 3.12. All the results are normalized to respect the confidential information.

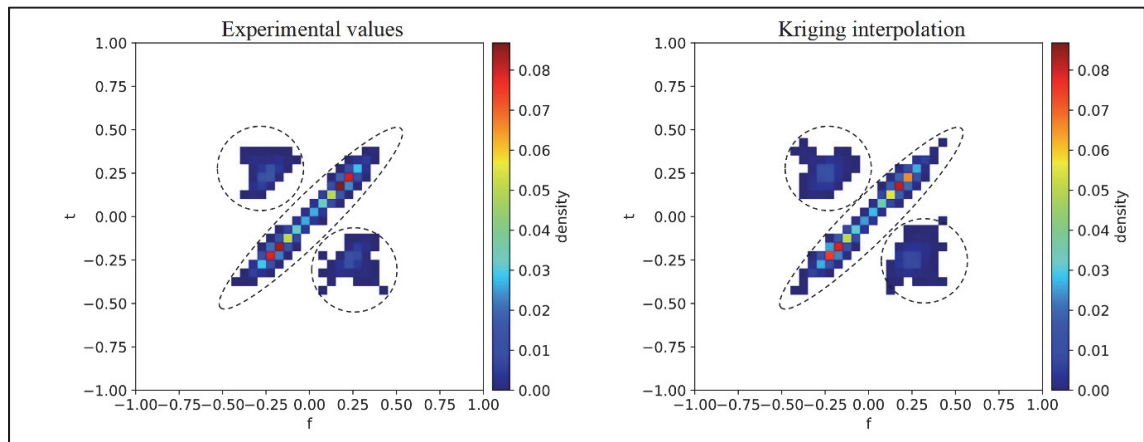


Figure 3.10 Comparison between experimental and interpolated ARM histogram at 55% OV

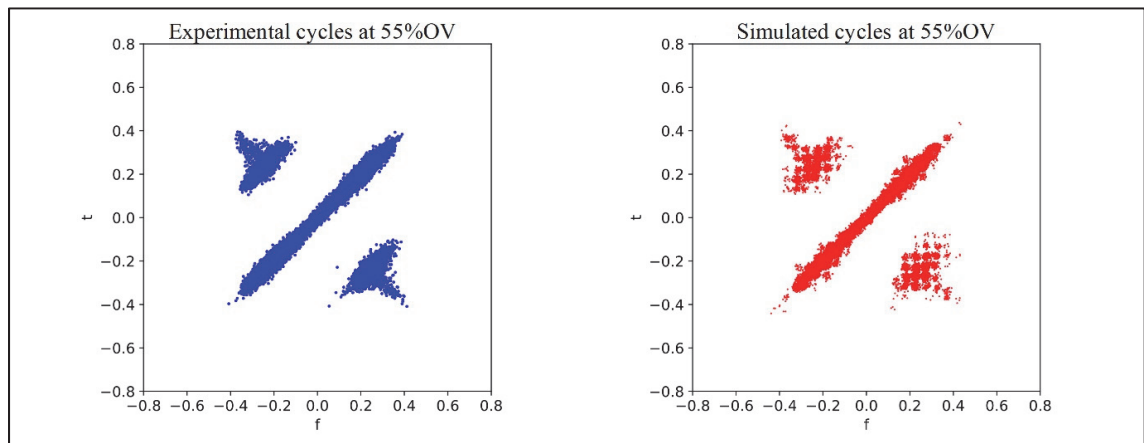


Figure 3.11 Experimental and generated cycles (one simulation) at 55 %OV

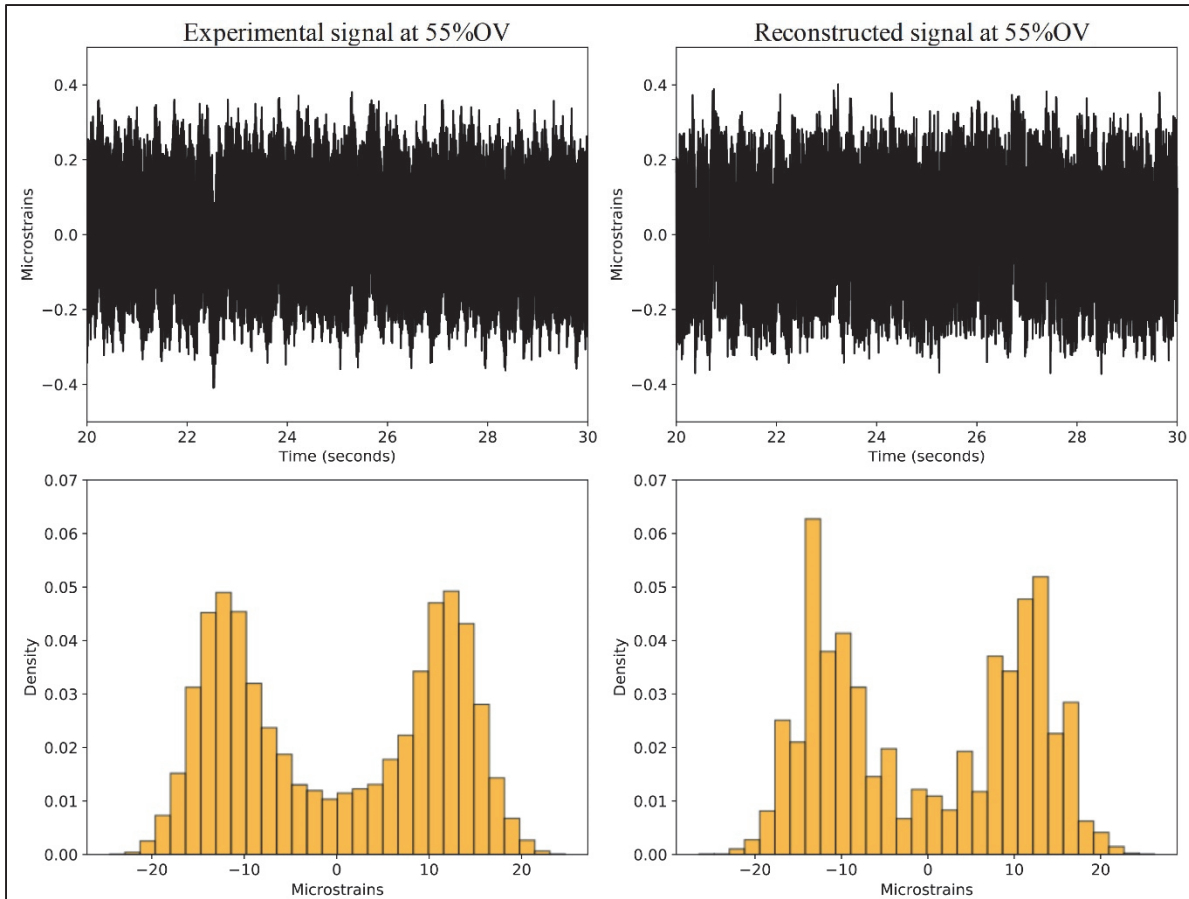


Figure 3.12 Comparison between experimental signal and one reconstructed signal at 55%OV; (Above) 10 seconds of signals; (Below) Histogram of whole signal

The interpolated ARM shows the same number of distribution modes (dashed line in Figure 3.10) as in the experimental ones; this means that this interpolation did not generate large overestimation in positions the null bins. Nonetheless, some differences are observed at the frontier positions of each mode in the interpolation result (Figure 3.10 et Figure 3.11). Graphically, the reconstructed signal (Figure 3.12) is not similar to the experimental one like we have foreseen, however its amplitudes have the same level as the experimental one. To verify the quality of interpolation process, we have compared the rainflow strain spectrums (strain range as a function of cumulative number of fatigue cycles) between 500 simulations (sampled from the set of BGDs) and the experimental values at some operating conditions (Figure 3.13).

For fatigue assessment, the expected goal in this study is to capture the amplitude/range level of fatigue cycle (especially high range cycles which cause more damage). At 55 %OV and 65 %OV, the experimental rainflow (the blue line in Figure 3.13) is mostly within of the set of simulations while there seems to be some overestimations at the high ranges at 80 %OV. This difference might be explained by that there are more operating conditions close to the 55 %OV, 65 %OV in the interpolation input set (see Tableau 3.1).

An unusual variability is observed at 30 %OV. The interpolation result depends on the nature of known data. The more complicated the combination of phenomena hidden in the signal, the less confident we might be in the interpolation result. Therefore, at some partial load conditions of turbine, this unusual variability of the interpolated result occurs when an operating condition with a complicated dynamic behavior is presented next to the interpolation target condition. Some of those conditions contain specific fluctuation coming from the vortex in the draft tube (Seidel et al., 2014), which appears and disappears over limited set of measured operating conditions rendering the interpolation difficult. This problem provides new challenge for further research, in which we would like to recover such fluctuation information.

For the runner maintenance strategy, even if the reconstructed signal is not similar to the experimental one (Figure 3.12), the interpolated strain ranges might be more representative of data variability and uncertainty than reusing experimental measurements from nearby operating conditions. Also, the proposed interpolation can be a solution to reduce the measurement requirements for runner fatigue assessment.

3.8 Conclusions

In this article, an interpolation process using kriging has been presented to estimate the strain range at any steady state operating condition based on rainflow cycle counting for hydroelectric turbine runner fatigue assessment. A set of bivariate Gaussian distributions were used to smooth the histogram, allowing quicker simulations of the loading cycles. It

must be noted that the proposed interpolation process is sensitive to the parameters used (i.e., the semivariogram model, etc.), using the error optimizing methods (such as cross-validation, Mean Absolute Error, etc.) are therefore highly recommended. The results of the case study show that even when the reconstructed signal is not similar to the experimental one, our proposed interpolation might be useful to estimate the dynamic strain level using only the load cycles information of strain signal.

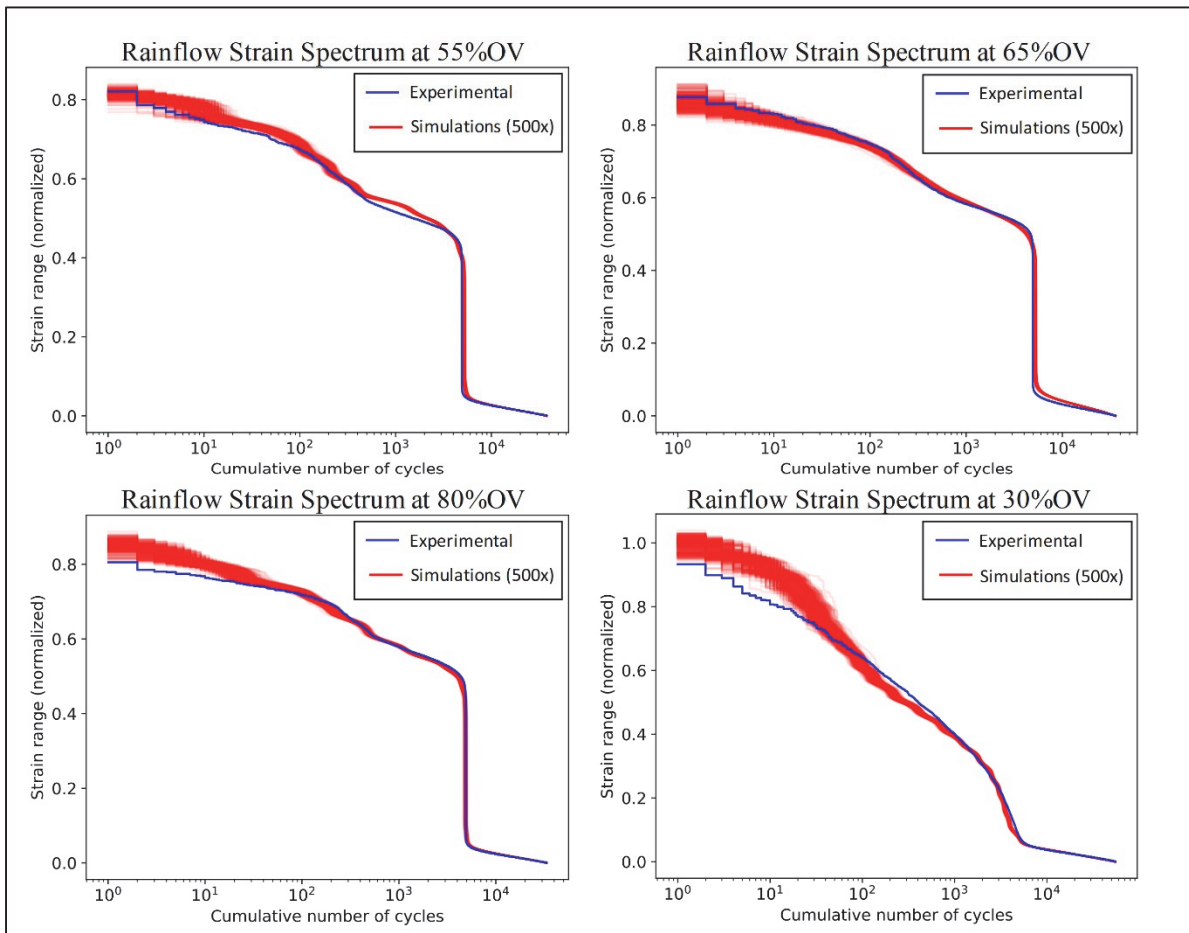


Figure 3.13 Rainflow strain spectrum between experimental values and simulations

The advantage of the proposed method is that it creates a stochastic process accounting for uncertainty hence contributing to fatigue risk assessment by generating loading cycles representative of observed behaviours. Moreover, it can help to optimize the requirements for experimental measurements.

In the result of this case study, the complicated phenomena hidden in strain signal caused the unusual variability at some partial load conditions. For further research, preserving other information of strain signal (not only the strain range information) during the interpolation phase will be interested to improve the quality of missing data recovery at unmeasured operating conditions.

CHAPITRE 4

PREDICTION OF HYDROELECTRIC TURBINE RUNNER STRAIN SIGNAL VIA CYCLOSTATIONARY DECOMPOSITION AND KRIGING INTERPOLATION

Quang Hung Pham^{a, b}, Martin Gagnon^c, Jérôme Antoni^b, Antoine Tahan^a, Christine Monette^d

^a École de technologie supérieure (ÉTS), 1100 Notre-Dame Ouest,
Montréal, Québec, Canada H3C 1K3

^b Université de Lyon, INSA Lyon, LVA, EA677, Bâtiment St Exupéry, 25 bis Avenue Jean
Capelle, 69621 Villeurbanne Cedex, France

^c Institut de recherche d'Hydro-Québec (IREQ), 1800 Boulevard Lionel-Boulet,
Varennnes, Québec, Canada J3X 1S1

^d Andritz Hydro Canada Inc, 6100 Route Transcanadienne, Point-Claire,
Québec, Canada H9R 1B9

Article publié dans « Renewable Energy », novembre 2021

4.1 Contribution à la thèse

L'interpolation des matrices de rainflow proposée dans le chapitre 3 donne une estimation rapide qui se concentre uniquement sur le niveau de fatigue du signal. En effet, le niveau de fatigue est un facteur important, mais en pratique on a besoin de plus d'information que la fatigue (ex., les natures physiques ou les comportements dynamiques du signal). Cela nous a donné la motivation de réaliser les travaux dans ce présent chapitre.

L'idée de cette recherche est d'identifier et d'isoler tous les comportements physiques importants du signal de déformation de la roue et ensuite d'essayer d'interpoler ces composantes en utilisant une méthode d'interpolation appropriée. Pour isoler les phénomènes cachés dans le signal, les outils de décomposition cyclostationnaires sont utilisés. D'après l'état de l'art, les phénomènes physiques extraits sont les périodicités, le phénomène de torche et les phénomènes stochastiques. La famille de krigeage est choisie pour créer la stratégie d'interpolation appropriée à chaque phénomène. Cet article fournit les contributions suivantes :

- un processus complet d'interpolation qui permet de reproduire le signal de déformation de la roue en capturant tous les phénomènes physiques présentés. Le signal reconstruit possède le même comportement dynamique (via visualisation spectrale) et aussi une dispersion acceptable pour le niveau de fatigue,
- une stratégie originale pour reproduire les amplitudes maximales du phénomène de torche du signal en appliquant la méthode de cokrigeage et une autre mesure indirecte (mesurés dans une zone plus facile d'accès que la roue),
- une amélioration de la technique de simulation des cycles de rainflow proposée dans le chapitre précédent. La méthode d'interpolation d'histogrammes de rainflow (voir chapitre 3) est appliquée pour l'interpolation de la partie stochastique. Cependant, au lieu d'utiliser un niveau unique de dispersion pour chaque bin d'histogramme, cette recherche propose des dispersions qui varient en fonction de la localisation de bins. Ces dispersions sont déterminées en basant sur les erreurs d'interpolation qui sont estimées via des simulations stochastiques conditionnelles.

4.2 Résumé de l'article en français

Les mesures de contraintes par les jauges de déformation jouent un rôle important dans l'évaluation de fatigue de la roue d'une turbine hydroélectrique. Pourtant, ces mesures ne peuvent pas couvrir tous les régimes de fonctionnement possible à cause des limitations expérimentales. Par conséquent, l'objectif de cette recherche est d'estimer le signal de déformation sur l'aube dans toutes les plages de régime permanent de la turbine hydroélectrique. Le signal de déformation mesuré durant un régime permanent contient certaines composantes (comme les périodicités, le phénomène de torche et les composantes stochastiques) qui peuvent causer des difficultés pendant la prédiction. Cet article propose une solution pour estimer les signaux de déformation de la roue en interpolant indépendamment chaque phénomène physique entre des régimes de fonctionnement. Ces composantes sont extraites en utilisant les outils de cyclostationnarité. Un cas d'étude est réalisé sur une turbine hydroélectrique Francis pour vérifier la robustesse de l'interpolation.

La méthodologie proposée contribue à l'évaluation de la fatigue et aide à réduire les exigences des mesures expérimentales.

4.3 Abstract

Strain measurements by gauges can play an important role in analyzing the fatigue damage of the hydroelectric turbine runner. However, these measurements cannot cover every steady-state operating condition due to the experimental limitations. Thus, the aim of this research is to predict the strain signal on runner over every possible steady operating condition using available experimental measurements. The strain signal measured during steady state involves several components (such as periodicity, vortex rope component and stochastic components) which can generate difficulties during the prediction. This paper proposes a solution to predict the runner strain signals by independently interpolating each physical phenomenon over different turbine operating conditions. These components are extracted using cyclostationary decomposition operators. A case study is performed on a Francis hydroelectric turbine to verify the interpolation performance. The proposed methodology can contribute to the fatigue assessment and help to reduce the requirements of *infield* measurements.

4.4 Introduction

In recent years, the emergence of renewable energies such as solar and wind have transformed the energy sector. The accompanying changes have led to decreased grid stability, hence the need for hydroelectric power plants to expand the operating range (X. Liu et al., 2016; Monette et al., 2016). Operating conditions such as partial load conditions present a higher dynamic strain/stress amplitude and a more complex dynamic behaviour than typical operations such as the Best Efficiency Point condition (BEP). As such, runner fatigue damage can be significantly influenced by these expanded operating conditions. A review by X. Liu et al. (2016) on fatigue damage demonstrates that runner fatigue damage can lead to bad consequences for the turbine operation if it is not predicted and addressed on time.

Obtaining precise strain/stress information over all possible turbine operating conditions is not exactly a straightforward task. The most reliable and accurate information in this context is derived from experimental measurements using strain gauges. However, high costs and time constraints make it impossible to cover all the possible operating conditions. On the other hand, numerical simulation, performed through a combination of Computational Fluid Dynamic (CFD) and Finite Element Analysis (FEA), can provide runner strain/stress estimates over any operating condition (Morissette et al., 2016; Nennemann et al., 2014; Xiao et al., 2008). Its drawback, though, is that it cannot properly capture the entire dynamic behaviour. Moreover, the accuracy of numerical simulations cannot be verified if experimental measurements are unavailable. Presas et al. (2019) also reviewed the fatigue life estimation methodologies used for Francis turbine runners and highlight the need to improve the fatigue reliability of runners, as well as the understanding of the dynamic behaviour of turbines in terms of strain/stress. In recent years, aside from the development of CFD simulations, some research studies have used mathematical models such as stationary signal models, statistical models, or stochastic simulations ... to predict the dynamic behaviour of runner strain signals (see (Presas et al., 2019)). These models are often created based on experimental measurement information. For example, a temporal extrapolation model based on cyclostationarity can allow the long-term time extrapolation of strain signals from short-term measurements (Poirier et al., 2017). Missing strain data can be estimated using kriging interpolation based on spatial covariance preservation (Ben Salah, 2014). An ARMAX (Autoregressive Moving Average with eXogenous input) model can be used to estimate the runner strain via indirect measurements (such as measurements on a turbine shaft) (Diagne et al., 2016). Notice that the use of such approaches could reduce the need for *in-field* measurements or CFD simulation calculation costs.

In a previous study, the authors have introduced a method to estimate the dynamic strain by interpolating the RainFlow-Counting matrices (RFC) between different operating conditions. The aim was to predict strain signal having similar expected fatigue damage as the reference case. However, the RFC matrix does not capture the physical behaviour of the runner strain signal. The reconstructed signal from interpolated RFC matrices is completely stochastic and

unable to account for periodic components which are deterministic by nature. In another study by the authors, the periodic components were interpolated using a spatio-temporal kriging method. In several instances, the structure of the strain signal is even more complex due to the presence of other phenomena that were not previously taken into account, such as the vortex rope phenomenon, which needs to be interpolated independently (see Figure 4.1). This complexity makes the interpolation of the complete signals more challenging. The goal of this paper is to develop an interpolation scheme that provides an estimate of the strain signal whose expected fatigue damage and physical behaviours are similar to reference measured values, over the complete range of operating conditions. Inspired by the use of signal decomposition for temporal extrapolation (Poirier et al., 2017), the main focus is to independently interpolate each extracted phenomenon component of the strain signal (Figure 4.2). This decomposition facilitates the interpolation process and preserves the dynamic behaviour in the estimated signals.

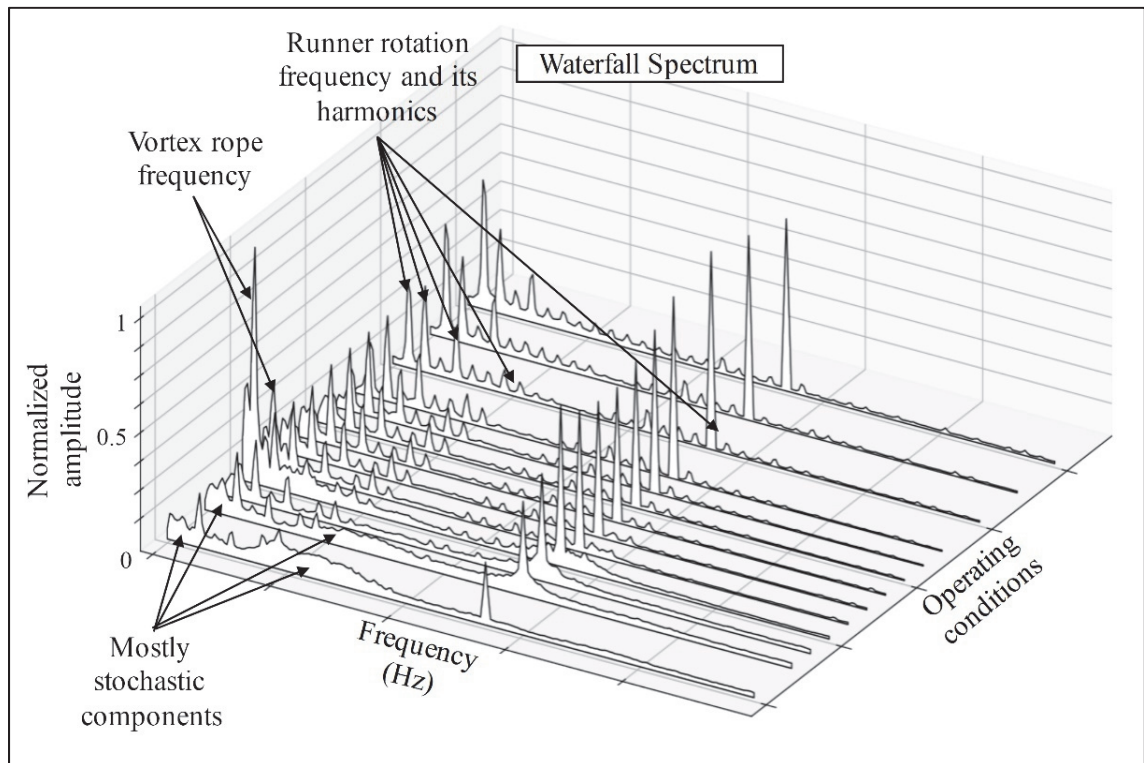


Figure 4.1 Waterfall spectrum with respect to operating conditions displaying the different physical phenomena present in a runner strain signal

Based on the observation of dynamic phenomena that significantly influence runner fatigue (X. Liu et al., 2016; Seidel et al., 2014), the three main components that need to be interpolated are:

- the Periodic Component (PC) related to the rotation speed of the runner and that can be viewed as first-order cyclostationary (CS1), and that is extracted using the synchronous average operator,
- the Vortex-Rope Component (VRC) related to outflow behavior in the draft tube at some partial load conditions (Dörfler et al., 2012; Seidel et al., 2014), which can be viewed as secondary-order cyclostationary (CS2), and that is extracted using the operator proposed by Poirier et al. (2017). This phenomenon is only available in some partial load conditions,
- the Residue Component (RC) which mostly contains the random component of the signal.

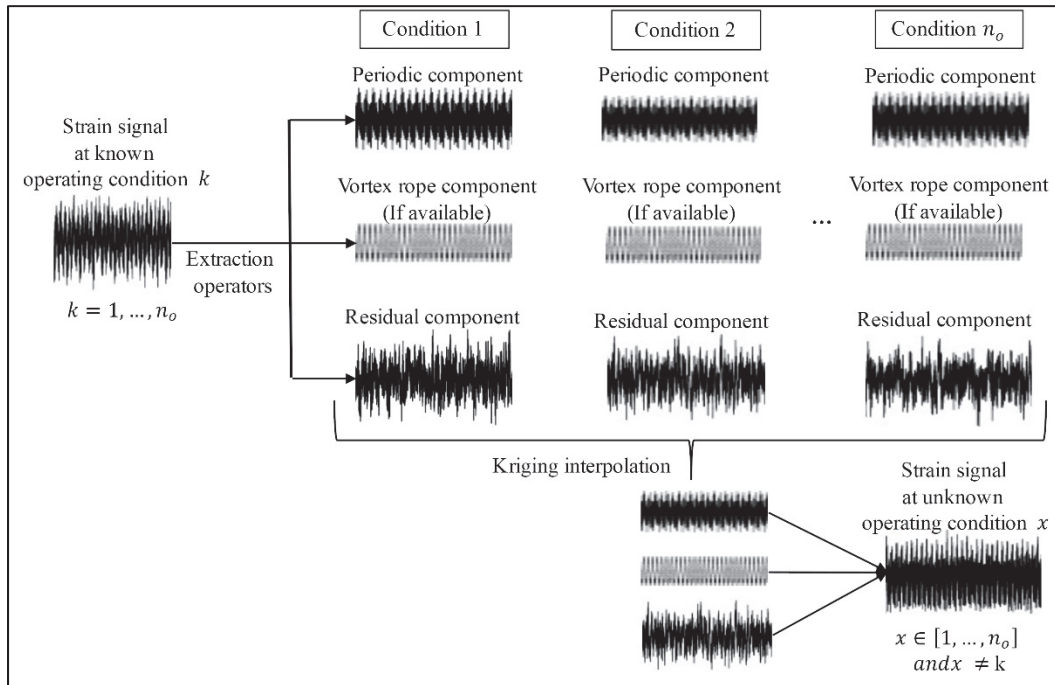


Figure 4.2 Proposed strategy to predict a missing strain signal

Thus, a strain signal $S(o, a)$ can be expressed by:

$$S(o, a) = \begin{cases} P(o, a) + V(o, a) + R(o, a) & \text{if VRC is available} \\ P(o, a) + R(o, a) & \text{if VRC is not available} \end{cases} \quad (4.1)$$

where P, V, R are respectively the periodic component, the vortex rope component, and the residue, o is the operating condition, and a is the rotation angle. Note that instead of using the temporal domain the rotation angle domain is used to represent the signal at each operating condition to improve the quality of the extraction method.

Kriging is the interpolation method family chosen in this study. The PC signal is interpolated using Spatio-temporal kriging. The VRC signal is interpolated using CoKriging which uses a secondary signal as reference to improve the estimation. Eventually, the RC signal is transformed into an RFC matrix and then interpolated by a multivariate kriging. The interpolation quality is validated both individually, for each phenomenon, and collectively, for the complete signal, using measured values. Note that, based on the available measurements, this research focuses on the interpolation between steady operating conditions which are defined by different guide vane opening levels at the same head.

The paper is organized as follows: First, section 4.5 introduces the extraction and interpolation methodologies mentioned above. Then, sections 4.6 and 4.7 discusses and validates the proposed methodology on a case study with a hydroelectric turbine runner.

4.5 Interpolation Methodology

4.5.1 Kriging Interpolation method

Kriging is a probabilistic interpolation method that uses the covariance of a dataset. Kriging provides not only an interpolation estimate, but also the variance of interpolation errors, which contributes to the uncertainty analysis. Moreover, kriging results can be combined with conditional stochastic simulations to provide a set of likely estimated values. The

performance of kriging has been demonstrated in a few research works in the hydroelectric turbine domain (Ben Salah, 2014; Pham et al., 2021; Pham et al., 2019).

In this study, the PC, VRC and RC are extracted from the experimental dataset and interpolated by independent kriging schemes tailored to their respective nature. Each kriging scheme has its own postulates. However, the general and common steps can be described as follows: (i) impose the regression interpolation model; (ii) model the covariance or semivariogram based on the input dataset; (iii) minimize the interpolation error variance using the covariance/semivariogram model to estimate weights in the equation of step (i); (iv) use the obtained weights to estimate the unknown value via the equation of step (i). Readers are referred to Chiles et Delfiner (2009); (Deutsch & Journel, 1992; Goovaerts, 1997; Montero et al., 2015; Pebesma & Heuvelink, 2016) for more information on kriging interpolation.

A key quantity used in kriging, the semivariogram, is defined as the half-variance between pairs of values and depends only on the distance increment of two locations. In kriging, the semivariogram is preferred to the covariance function because the mean value is not required as it is in the latter, and also because its existence is guaranteed for a wider class of stochastic processes.

The following subsections introduce the different extracted components and the associated kriging schemes.

4.5.2 Periodic Component

The PC, which relates to a phenomenon synchronous to the rotation of a rotating machinery, is a typical feature of hydroelectric turbines. At a steady operating condition, the rotation speed is considered constant. This PC is thus a periodic signal in time, with a period equal to the duration of one rotation of the runner. However, the PC is often difficult to observe directly because it is masked by interferences in the strain signal, such as stochastic

components. Its estimation can be achieved within the framework of cyclostationarity which enables the extraction of “hidden” periodicity (Antoni, 2009; Antoni et al., 2004). The PC is well modelled by a CS1 process, i.e. a process whose expected value (in the ensemble sense) is periodic. Therefore, it can be extracted using the Synchronous Average operator (SA):

$$\hat{P}[o, i_a] = \frac{1}{K} \sum_{k=0}^{K-1} S[o, m + kN] \quad (4.2)$$

where $\hat{P}[o, i_a]$ is the extracted periodic part (at operating condition o) as a function of the angular index i_a , $S[o, i_a] = S(o, 2\pi i_a/N)$ is the strain signal resampled in the angular domain, see Antoni et al. (2004); (Bonnardot, 2004), with K cycles and full-length L , $N=L/K$ is the cycle length, and $m=\text{mod}[i_a, N]$. Note that, to simplify the notation, $\hat{P}[o, i_a]$ is expressed with $P(o, a)$ for the other following equations.

The Spatio-Temporal Kriging (STK) is chosen for interpolating the PC because it has shown good performance in one of our previous research (Pham et al., 2019). STK independently processes the “spatial” (here the operating condition “ o ”) and “temporal” (here, the rotation angle “ a ”) coordinates when building the semivariogram. The regression equation in STK reads:

$$P^*(o, a) = \sum_{k=1}^{n_0} \sum_{l=1}^{l_k} \lambda_{kl}(o, a) P(o_k, a_l), \quad (4.3)$$

where $P(o_k, a_l)$ is a known value of the PC extracted by the SA at location (o_k, a_l) , where exponent $*$ means the target value, λ_{kl} is the regression weight, n_0 is the number of known operating conditions (input set) and l_k is the number of rotation angles of extracted periodic parts at operating condition o_k . The regression weights are obtained via the minimization of the cost function:

$$L(\lambda_{kl}, \mu) = \sigma_e^2(o, a) + 2\mu \left(\sum_{k=1}^{n_0} \sum_{l=1}^{l_k} \lambda_{kl}(o, a) - 1 \right) \quad (4.4)$$

where μ is the Lagrange coefficient associated with the non-bias condition $\sum_{k=1}^{n_0} \sum_{l=1}^{l_k} \lambda_{kl}(o, a) = 1$, $\sigma_e^2(o, a)$ is the error variance between the real value P and the interpolated value P^* at the same location (o, a) , $\sigma_e^2(o, a) = \text{Var}(P(o, a) - P^*(o, a))$. Making use of the equation (4.3), the minimization of the equation (4.4) leads to a system that allows the obtention of λ_{kl} and μ :

$$\begin{cases} \sum_{k=1}^{n_0} \sum_{l=1}^{l_k} \lambda_{kl}(o, a) = 1 \\ \sum_{k=1}^{n_0} \sum_{l=1}^{l_k} \lambda_{kl}(o, a) \gamma(u_{|o_k - o_l|}, v_{|a_l - a_j|}) - \mu = \gamma(u_{|o - o_l|}, v_{|a - a_j|}) \end{cases} \quad (4.5)$$

where γ is the semivariogram value, $u_{|o_k - o_l|}$ is the operating condition increment, and $v_{|a_l - a_j|}$ is the rotation angle increment, $j = 1, \dots, l_k$ and $i = 1, \dots, n_0$.

To obtain semivariogram value, one first creates the Experimental Semivariogram (ES) from the experimental dataset:

$$\hat{\gamma}(u, v) = \frac{1}{2N(u, v)} \sum_{(k, l, i, j) \in v(u, v)} [P(o_k, a_l) - P(o_i, a_j)]^2 \quad (4.6)$$

where $v(u, v)$ is a set of paired points separated such that $|o_k - o_l| \in [u - du, u + du]$ and $|a_l - a_j| \in [v - dv, v + dv]$, and $N(u, v)$ is the cardinal of set $v(u, v)$. However, the $\hat{\gamma}$ is not used in the equation (4.5) because the ES cannot provide a continuous semivariogram value set at all the possible increments. Therefore, a parametric semivariogram model is fitted to $\hat{\gamma}(u, v)$. Note that other estimation methods such as maximum likelihood and cross

validation, can also be used to determine the semivariogram model (Montero et al., 2015; Pebesma & Heuvelink, 2016). Finally, the interpolated values are obtained via the equation (4.5) using the value of fitted semivariogram model γ . For other detailed explanations, readers can refer to Pham et al. (2019).

4.5.3 Vortex Rope Component

The VRC generates a quasi-periodic pressure fluctuation in the turbine caused by the outflow vortex rope in the draft tube of the turbine. This outflow condition leads to a backflow in the centre of the draft tube cone and a vortex of helical shape (Dörfler et al., 2012; Seidel et al., 2014). This phenomenon can be observed in partial load conditions via a high-pressure fluctuation at a low frequency. Generally, in fixed-blade turbines, such as the Francis turbine, this phenomenon is found in conditions where the flow is between 50% and 85% of the BEP flow (Dörfler et al., 2012). The maximum amplitude is difficult to capture accurately with measurements or with numerical simulation, since the phenomenon tends to suddenly appear and then disappear in the neighbouring conditions. Therefore, this component is very sensitive to interpolation.

The VRC is extracted from the strain signal using a methodology based on CS2 proposed by Poirier et al. (2017). The method allows the creation of a new angular domain (linked to the VCR frequency) around a given frequency (corresponds to the runner rotation frequency in this study).

The movement of this vortex affects not only the runner, but also the other components of the turbine. The VRC can therefore be observed in other signals such as shaft bending, shaft torsion, draft tube pressure, etc. This leads to the idea of adding an exogenous signal in the interpolation phase to add more information to the runner strain estimation. Cokriging is therefore used to interpolate the vortex rope phenomenon. This method allows the addition of a secondary signal which correlates to the signal of interest via the linear regression:

$$V^*(z) = \sum_{i=1}^{n_1} \lambda_i(z) V_1(z_i) + \sum_{j=1}^{n_2} \alpha_j(z) V_2(z_j) \quad (4.7)$$

where V^* is the target VRC with location z that depends on the two coordinates (o, a) , V_1 is the experimental VRC value (also called the primary signal or the signal of interest), V_2 is the experimental VRC value derived from a secondary information with similar dynamic behaviour as V_1 , λ_i and α_i correspond to regression weights, and n_1 and n_2 are the numbers of experimental operation conditions for V_1 and V_2 . There are two non-bias constraints: $\sum_{i=1}^{n_1} \lambda_i(z) = 1$ and $\sum_{j=1}^{n_2} \alpha_j(z) = 0$. Cokriging requires a cross-semivariogram which is created from primary and secondary information V_1 and V_2 . Since the input quantities of two VRC are different, the pseudo-cross semivariogram is used (the equation (4.8)). In this case, there is only one type of spatial increment $h_{CK} = \|z_i - z_k\|$ used in the pseudo cross-semivariogram

$$\gamma_{12}(h_{CK}) = \frac{1}{2} \times \text{Var}[V_1(z_i) - V_2(z_k)] \quad (4.8)$$

of VRC V_1 and V_2 . In order to create the semivariogram from the experimental dataset, the experimental pseudo cross-semivariogram (Papritz, Künsch, & Webster, 1993) is built:

$$\hat{\gamma}_{12}(h_{CK}) = \frac{1}{2N(h_{CK})} \sum_{(i,k) \in v(h_{CK})} \{[V_1(z_i) - \bar{V}_1] - [V_2(z_k) - \bar{V}_2]\}^2 \quad (4.9)$$

where $\bar{V}_1$ and $\bar{V}_2$ are the means of the experimental observations of V_1 and V_2 , $\|z_i - z_k\| \in [h_{CK} - dh_{CK}, h_{CK} + dh_{CK}]$ with $dh_{CK} > 0$. The next step is to fit a parametric model of the cross-semivariogram. The most common approach is the use of the Linear Model of CoRegionalization (LMCR), as explained in Goulard et Voltz (1992). LMCR ensures that all fitted semivariogram models are negative semi-definite, i.e., with positive semi-definite covariance. After suitable parametric semivariogram models are obtained, the ordinary

Cokriging formulas are applied to obtain the interpolated VRC (Goulard & Voltz, 1992; D. E. Myers, 1982, 1991).

4.5.4 Residue Component

The final part to interpolate is the residue left after the extraction of the PC and VRC. The RC is mainly random in nature. Based on previous research, the RC can be predicted by interpolating the RainFlow Counting (RFC) matrices (as explained in CHAPITRE 3). The rainflow counting algorithm, which allows the analysis of signal amplitudes by assigning extreme values (maximum and minimum) of loading cycles, is often used to analyze fatigue levels. The residual components are transformed into RFCs, and then interpolated between operating conditions. This method preserves fatigue cycle information, but the reconstructed signal is completely random. Therefore, in previous works, the interpolation of full RFC matrices showed a representation of strain cycles for fatigue analysis but was unable to reconstruct physical phenomena present in the original signal. This RFC interpolation method is better suited when the signal is mostly random. Note that an RFC is represented by a 2D histogram, where each bin contains the occurring probability of cycles. Using a fixed number of bins helps to decrease the interpolation time. The interpolation method used is based on a multivariate kriging where the semivariogram is processed with two coordinates: the first is the histogram bin position “ p ”, which depends on *From-To* (noted as f and t) values (*From* and *To* values are respectively the amplitude of the first point and of the final point of a cycle) in the RFC, while the second is the turbine operating conditions “ o ”. Applying for the RC interpolation, one notes $B(p_k, o_l)$ is the value of bin p_k at operating condition o_l , and then the linear regression and the Lagrange function of kriging are:

$$B^*(p_i, o_x) = \sum_{k \in \mathcal{D}_i} \sum_{l=1}^{n_o} \lambda_{kl}(p_i, o_x) B(p_k, o_l) \quad (4.10)$$

$$L(\lambda_{kl}, \mu) = \sigma_e^2(p_i, o_x) + 2\mu \left(\sum_{k \in \vartheta_i} \sum_{l=1}^{n_o} \lambda_{kl}(p_i, o_x) - 1 \right) \quad (4.11)$$

where B^* is the interpolated value, $\lambda_{kl}(p_i, o_x)$ is the regression weight, n_o is the number of experimental turbine conditions in the input set, ϑ_i is a set of neighbourhood positions of p_i (including p_i), and $\sigma_e^2(p_i, o_x) = \text{Var}(B(p_i, o_x) - B^*(p_i, o_x))$ is the error variance. After minimizing the cost function (the equation (4.11)), a system is obtained allowing to determine the regression weights:

$$\begin{cases} \sum_{k \in \vartheta_i} \sum_{l=1}^{n_o} \lambda_{kl}(p_i, o_x) \gamma(h_{\|p_k - p_j\|}, u_{|o_l - o_m|}) - \mu = \gamma(h_{\|p_i - p_j\|}, u_{|o_x - o_m|}) \\ \sum_{k \in \vartheta_i} \sum_{l=1}^{n_o} \lambda_{kl}(p_i, o_x) = 1 \end{cases} \quad (4.12)$$

where $j \in \vartheta_i$, $m = 1, \dots, n_o$, et $\gamma(h_{\|p_k - p_j\|}, u_{|o_l - o_m|})$ is semivariogram value at increments $h_{\|p_k - p_j\|}$ and $u_{|o_l - o_m|}$. Then, the ES is estimated from the experimental dataset and then a parametric model is fitted (using LMCR) to obtain the semivariogram values γ used in the equation (4.12). The interpolated values are estimated via the equation (4.10) using the set of λ_{kl} found in the equation (4.12).

After the RFC interpolation, the estimated RFC matrix is then smoothed using Bivariate Gaussian Distributions (BGD). Cycles inside each bin are assumed to follow a BGD of which mean corresponds to the coordinates of the centre point of the bin (or to increase the dispersion, one can take a position around the bin center via a uniform distribution) and the covariance matrix is $\sigma^2 \mathbf{I}_2$ (where $\sigma = 0.5 \times \text{bin diagonal} / \text{variability parameter}$, $\mathbf{I}_2$ is the identity matrix). Strain cycles are then generated by respecting the occurring probability of cycles in each bin (interpolated values). The sum of occurring probabilities in the RFC is equal to 1. A *variability parameter* is proposed to ensure that the simulated cycle distribution

can cover the full bin but does not add too much variability outside the bin. This parameter is selected by performing a Monte-Carlo simulation with a limited level of variability. In this study, the kriging error is used as the limited level variability. It must also be noted that the kriging error variance cannot be used because it only shows the uncertainty propagation, not the interpolation error. To obtain the kriging estimation error, Conditional Sequential Simulation (CSS) based on the kriging interpolation is used (Deutsch & Journel, 1992; Nowak & Verly, 2005). The estimation error is presented in the form of a coefficient of variation (%) (see Figure 4.9). The coefficient of variation is the ratio between the standard deviation and the mean of the set of CSS. Therefore, via the Monte-Carlo simulation, a variability parameter is obtained for each bin. Finally, from the interpolated result, several samples (full RFCs) are generated, and each sample can generate a corresponding signal by applying the inverse method proposed by D. Socie (2001).

4.5.5 Complete Interpolation Methodology

Figure 4.3 summarizes the proposed interpolation methodology. The interpolated results are gathered to obtain a complete interpolated strain signal (sum of all the interpolated components). The final output set corresponds to a set of possible values of estimations which are obtained using kriging and conditional simulations. The quality of this interpolation methodology is verified and validated via a case study on a Francis turbine runner in the following section.

4.6 Case Study

Tableau 4.1 Input and Output sets of the interpolation process

	Input (V_o)	Output (V_o)
Periodic component	11, 30, 35, 45, 50, 55, 60, 70, 80, 100	20, 40, 90
Vortex rope component	For the runner strain: 35, 45, 50, 55 For the shaft bending: 40, 45, 50, 55	40
Residue	11, 30, 35, 45, 50, 55, 60, 70, 80, 100	20, 40, 90

The data used in this case study was derived from experimental measurements by strain gauges, which were installed on runner blades of a Francis hydroelectric turbine of a power plant in Quebec (Canada). The measured locations on the blades are close to critical locations where the highest stress levels are expected in the blades. The dataset contains the measured runner strain at steady-state operating conditions ranging from SNL to full load conditions. Each steady state is named after the percentage of turbine guide vane opening (i.e., noted as V_o), which controls the flow rate going through the runner. For example, in the following, $20V_o$ corresponds to an operating condition in which the vanes open to about 20% of the maximum opening. Tableau 4.1 shows the input and output sets of the interpolation in this case study. About 70% of the experimental dataset is used as inputs, and the remainder is used for validation.

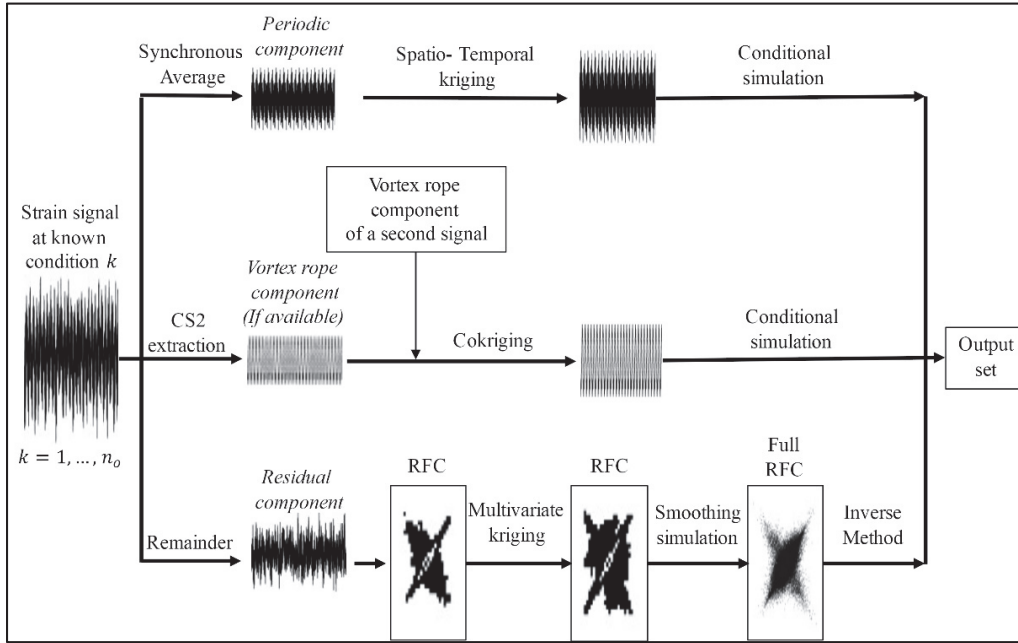


Figure 4.3 Flowchart of the proposed interpolations strategy

Three operating conditions representing typical dynamic behaviours of the turbines were selected for validation:

- $90V_o$: This operating condition contains the most ‘stable’ runner dynamic behaviour and is close to the BEP. The fatigue damage at this condition is often lower than at other conditions,

- $40V_0$: This operating condition contains the largest measured strains associated with the vortex rope phenomenon. A high strain amplitude at rope frequency can increase the risk of fatigue damage at this operating condition,
- $20V_0$: This operating condition often contains the highest runner dynamic strain and is more stochastic, with a larger quasi-random residue. The high extreme value at this condition can lead to a high risk of fatigue damage.

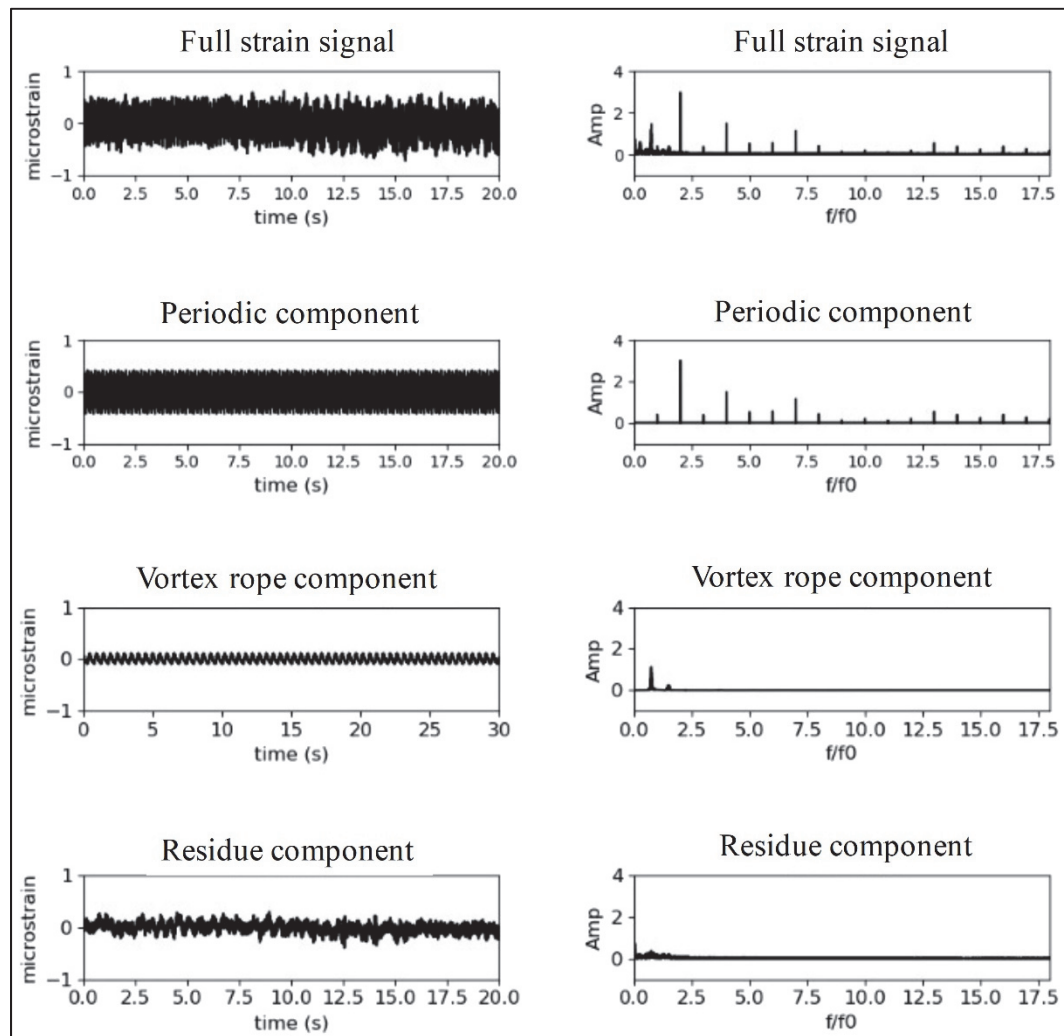


Figure 4.4 Example of component extractions displayed in time (left column) and frequency (right column) domains

Each phenomenon (periodicity, vortex rope, residue) is extracted at all the operating conditions of input set before interpolation (Figure 4.4). For prediction at the $20V_0$ and $90V_0$,

the input conditions need only two interpolation components (periodic and residual part) while the prediction at the $40V_0$ needs three interpolation components (because of the presence of the vortex rope).

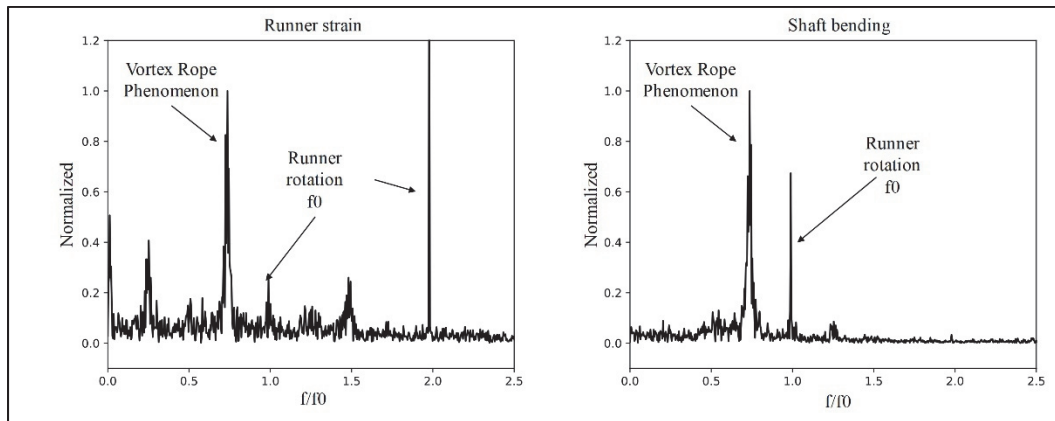


Figure 4.5 Spectra of the vortex rope phenomenon (normalized) in a runner strain (left) and in a shaft bending (right)

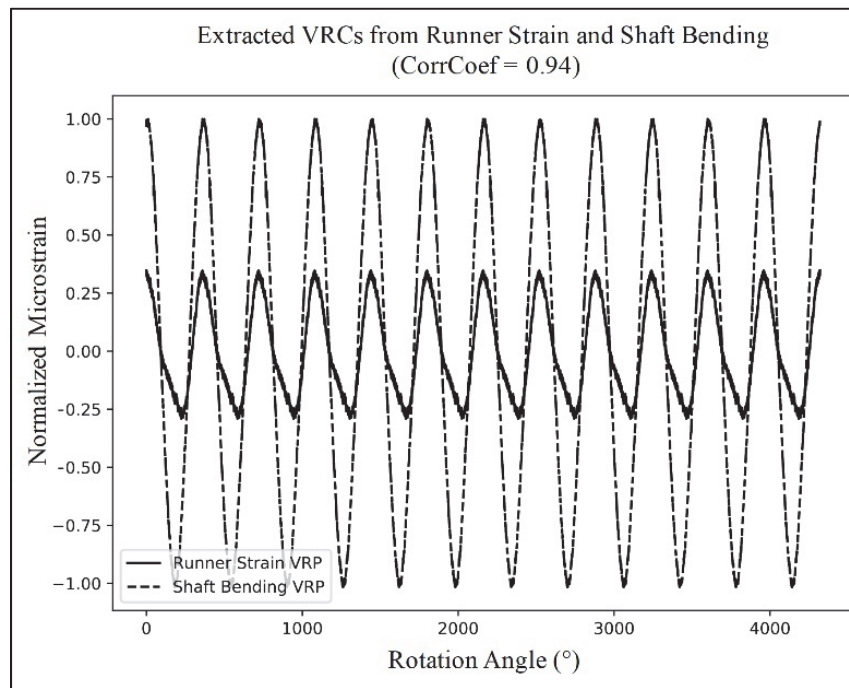


Figure 4.6 Extracted VRC of runner strain and shaft bending at the same operating condition

Since the VRC only occurs in some specific operating conditions and the CK needs secondary information, the input set for VRC interpolation is different from that for other component interpolations (Tableau 4.1). In this study, the shaft bending measurement is chosen as the secondary information because the VRC is visible in this signal (Figure 4.5). Note that CK interpolation requires secondary information that has a good correlation with the information of interest. The extracted VRC in the shaft bending signal has higher relative amplitude than the extracted VRC of the runner strain, but they are highly correlated (the Pearson correlation coefficient is about 0.94; see Figure 4.6). Both VRCs were normalized to obtain similar amplitudes, which avoids an unusual form of pseudo cross-semivariogram.

4.7 Results and Discussion

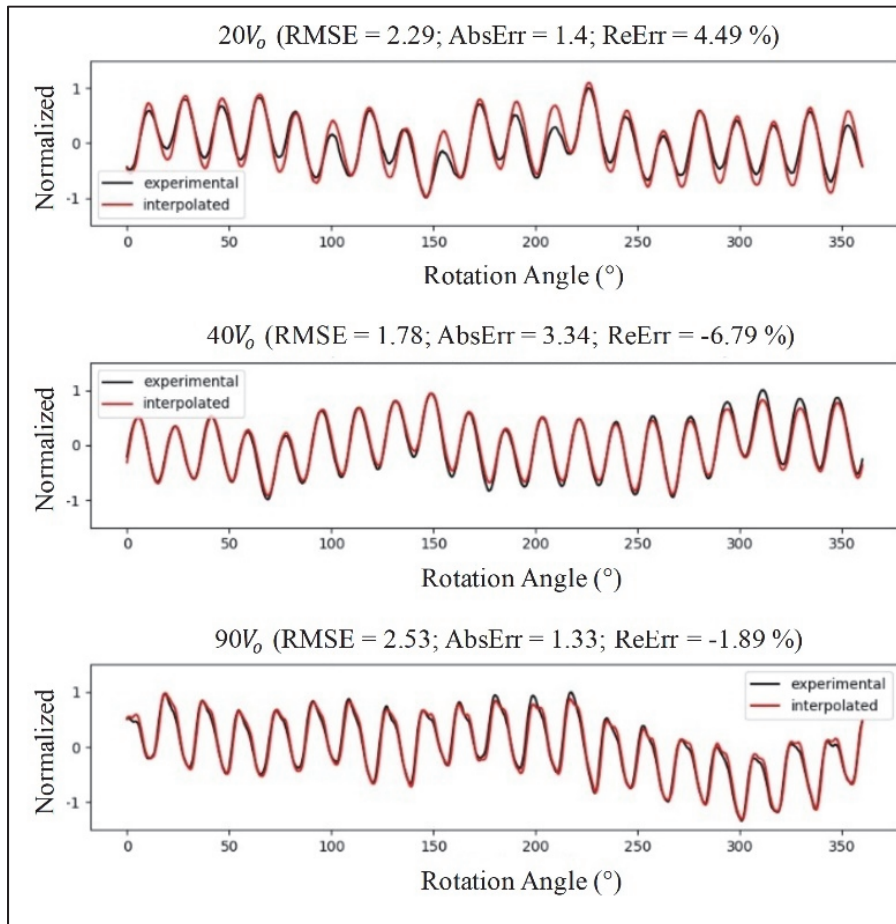
In this section, all amplitudes are normalized for confidentiality purposes, while the error values are calculated from real values. The error values used in this case study are Root-Mean-Square Error (RMSE), Absolute Error (AbsErr) and Relative Error (ReErr) values. AbsErr and ReErr are only applied for the estimation error of the maximum range. Tableau 4.2 displays the interpolation errors for each component, as well as the complete signal. Note that the interpolation of the residue generates a set of simulated signals. Therefore, the errors of the residue and the complete signal are computed as the median error of the simulation set. The negative symbol in ReErr values signifies the underestimation.

4.7.1 Interpolation of Periodic Component

The interpolation results are graphically close to the experimental one (Figure 4.7), with RMSE values small enough to confirm that the STK is suitable for this case study. Moreover, all the relative errors are lower than 7 %. The estimation errors at partial load, such as $20V_0$ and $40V_0$ are higher than those for the “stable” operating condition such as $90V_0$. The proposed kriging method correctly captured the periodic components contained in the runner strain signal both in terms of wave form and amplitude.

Tableau 4.2 Estimation error values of each component and complete signal

Error Type	RMSE			AbsErr (μStrain)			ReErr (%)		
V_o Phenomena	20	40	90	20	40	90	20	40	90
Periodic part	2.29	1.78	2.53	1.40	3.34	1.33	4.49	-6.79	-1.89
Vortex rope		2.15			0.86			-2.86	
Residue	10.25	7.87	6.35	1.48	4.50	1.33	0.90	8.23	3.13
Complete signal	12.12	21.77	25.95	3.40	3.99	4.24	-1.46	-2.00	-3.86

Figure 4.7 Interpolation results for periodic components at $20V_o$, $40V_o$, $90 V_o$

4.7.2 Interpolation of Vortex Rope Component

For the interpolation of the vortex rope component, auto-semivariograms and a pseudo-cross semivariogram (including fitted LMCR) are shown in Figure 4.8. The similar form of these three semivariograms confirms that the VRC from shaft bending is well correlated to the VRC of the runner strain. The interpolated VRC amplitudes (Figure 4.8) are close to the experimental values. In the measurements, the highest VRC amplitude is observed at $40V_0$, which is the target output condition. The proposed interpolation captured this condition signal amplitude with minimal error (RMSE = 2.15, AbsErr = 0.86, ReErr=-2.86%) using only the strain signal from neighbouring operating conditions with much smaller amplitudes than at $40V_0$ and shaft bending at the target condition. Indeed, shaft bending can be measured in the stationary domain of the turbine and is therefore easier to obtain and less costly than the runner blade strains. Therefore, using the shaft bending VRC as secondary input in CK plays an important role in reaching a good interpolation result.

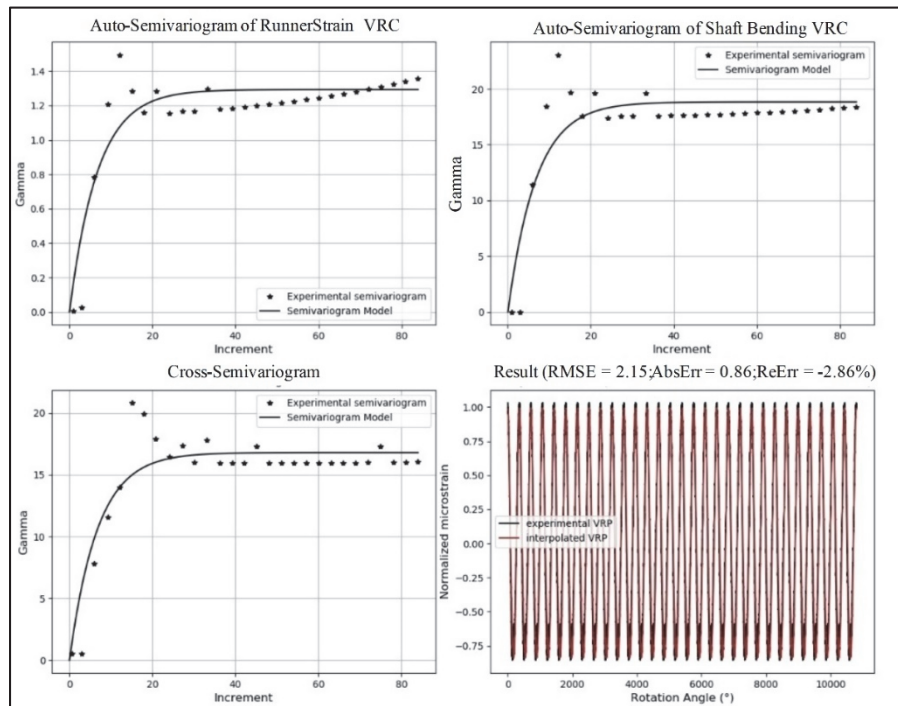


Figure 4.8 Interpolation results for the VRC (including semivariograms)

4.7.3 Interpolation of Residue

Figure 4.9 shows the kriging interpolation of the RFC at each target operation condition. Based on 500 CSS, the kriging error for each bin is presented in the form of a coefficient of variation (%). Coefficients of variation are mostly under 15%, while higher coefficients are distributed at the border of the RFC. These errors come from the incoherence of values at the border positions between the RFCs in the input set. Moreover, the presence of many null bin values around these border positions can also influence the interpolation results. Based on the kriging values obtained, a set of runner strain cycles is generated by smoothing the interpolated RFC. Figure 4.10 presents one of the simulations and its corresponding reconstructed signal. To verify the interpolation quality of this residual component, the rainflow strain spectra of 500 simulations were compared to the experimental one (Figure 4.10). Higher ranges were correctly captured with relative errors smaller than 10% (Tableau 4.2). A slight overestimation seems to be observed at $40V_o$, which may be due to the high dispersion when smoothing the RFC.

4.7.4 Complete Interpolated Signal Verification

The final step is to validate the estimation error for the complete signal (sum of all the interpolated components). The verification goal is to confirm that the amplitude/range and the physical behaviour are well reproduced.

The residue is randomly reconstructed, which leads to higher variability in the complete signal estimation (see the RMSE values in Tableau 4.2). However, note that the errors of the highest strain range prediction are small ($|\text{ReErr}| < 10\%$, see Tableau 4.2). The frequency spectra of the interpolations are compared in Figure 4.11 with experimental measurements. Mostly all the phenomena are observed in the interpolated results, which ensures the preservation of the dynamic behaviour after interpolation.

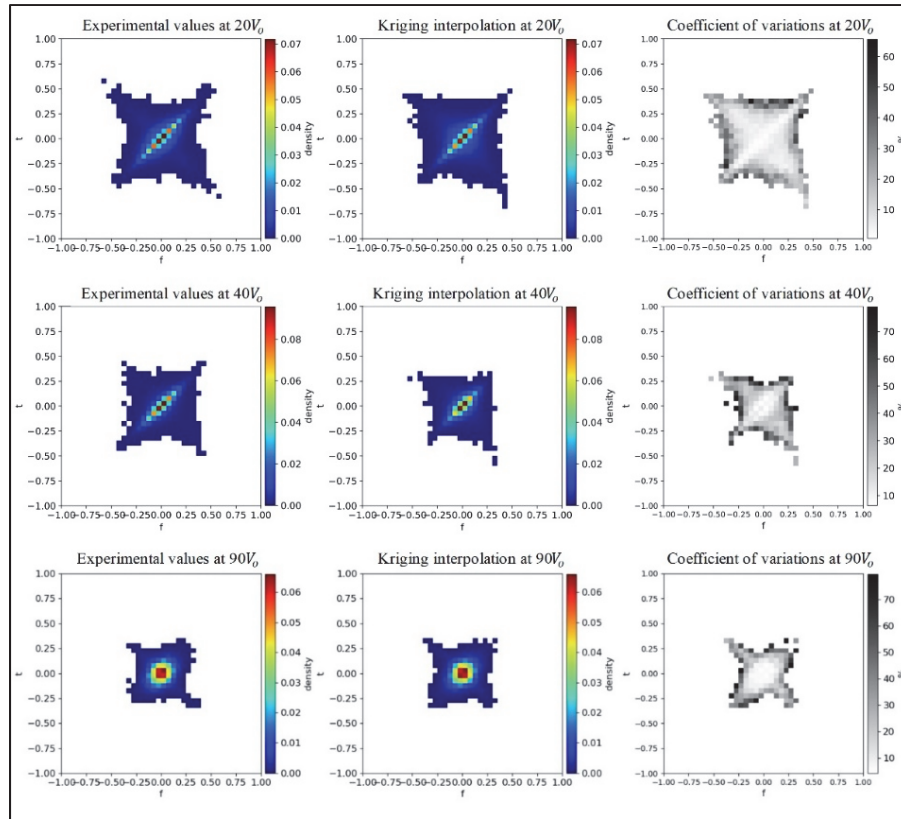


Figure 4.9 Kriging interpolation (middle) of the RFC (left) and the coefficient of variations (right)

For each component, there is also a set of possible estimations obtained from the conditional simulations. Using a set of simulations allows a better evaluation of fatigue damage uncertainty as compared to a single interpolation. As observed in Figure 4.12, the proposed interpolation methodology properly predicts the highest amplitude cycles, while there are higher estimation errors for the lower amplitude cycles. These differences were anticipated because each phenomenon was interpolated independently, leading to some incoherence at lower amplitudes during the summation of the complete signal. However, in practice, the lowest amplitude cycles have only a minor to negligible influence on the fatigue assessment of the hydroelectric turbine runner.

According to the obtained results, the interpolations for the PC and VRC demonstrate a better performance than for the residue. Nonetheless, a higher error in the residue estimation does not cause a significant problem, and mostly influences the lower amplitude fatigue cycles. In

practice, the proposed interpolation methodology could help optimize the requirements for the experimental measurement for a given expected level of precision.

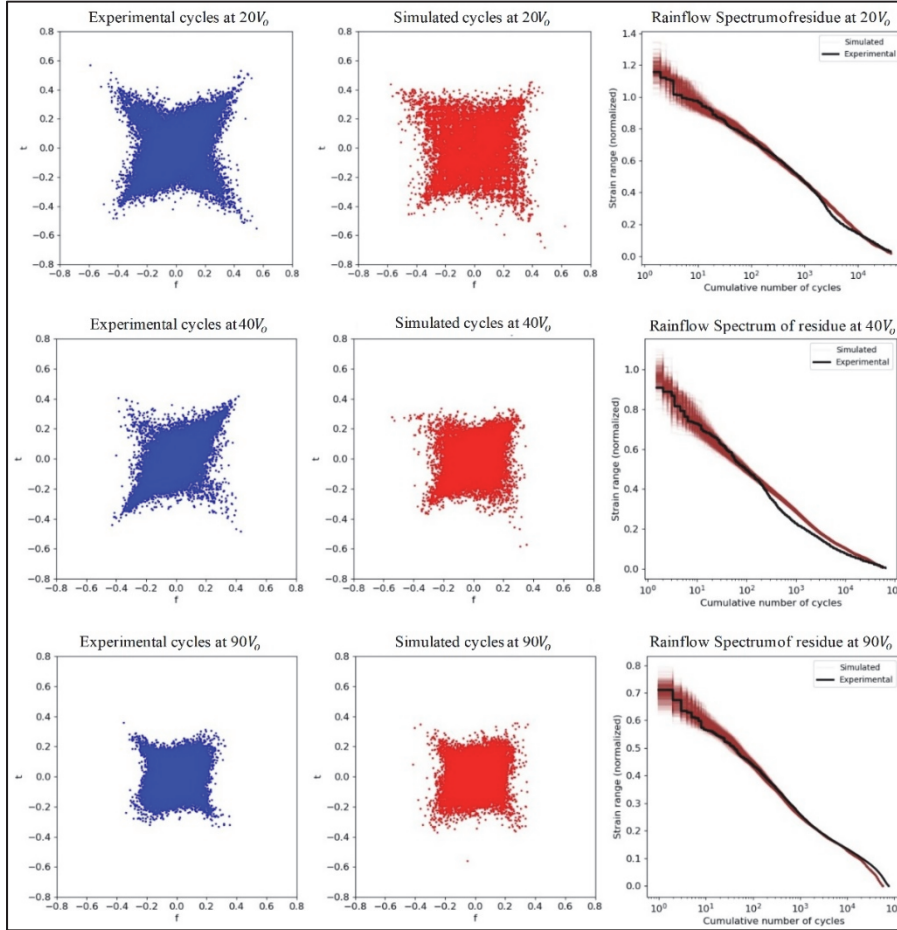


Figure 4.10 Simulation of residue at $20V_o$, $40V_o$, $90V_o$
(left column: experimental RFCs, middle column: simulated RFCs,
right column: simulated rainflow spectra)

4.8 Conclusion

In this article, an interpolation process is proposed to facilitate the estimation of the runner strain signal over all possible steady-operating conditions. Each physical phenomenon is extracted and is interpolated independently using different kriging approaches. The synchronous periodic part is predicted using spatio-temporal kriging. The cokriging process, which uses shaft bending as secondary information, is applied to interpolate the vortex-rope

phenomenon. The residue components are indirectly interpolated in the form of a rainflow counting via a multivariate kriging process. The result of the case study on a Francis turbine runner indicates that this proposed interpolation tool can properly capture high strain amplitudes, which is helpful in the evaluation of runner fatigue. Moreover, this interpolation tool ensures the observation of physical strain signal phenomena.

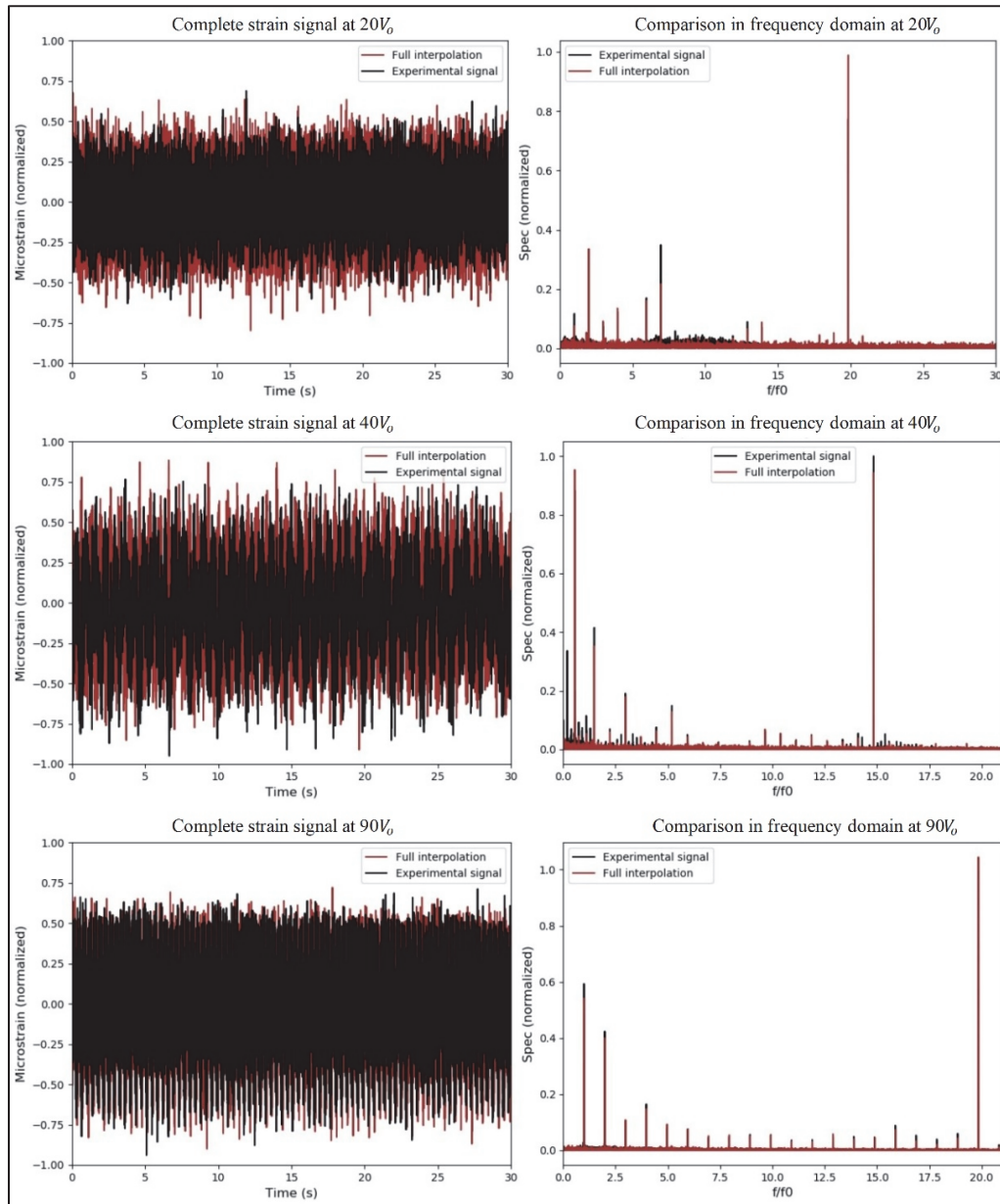


Figure 4.11 Comparison of interpolation result of complete strain signal displayed in the time (left column) and frequency (right column) domains

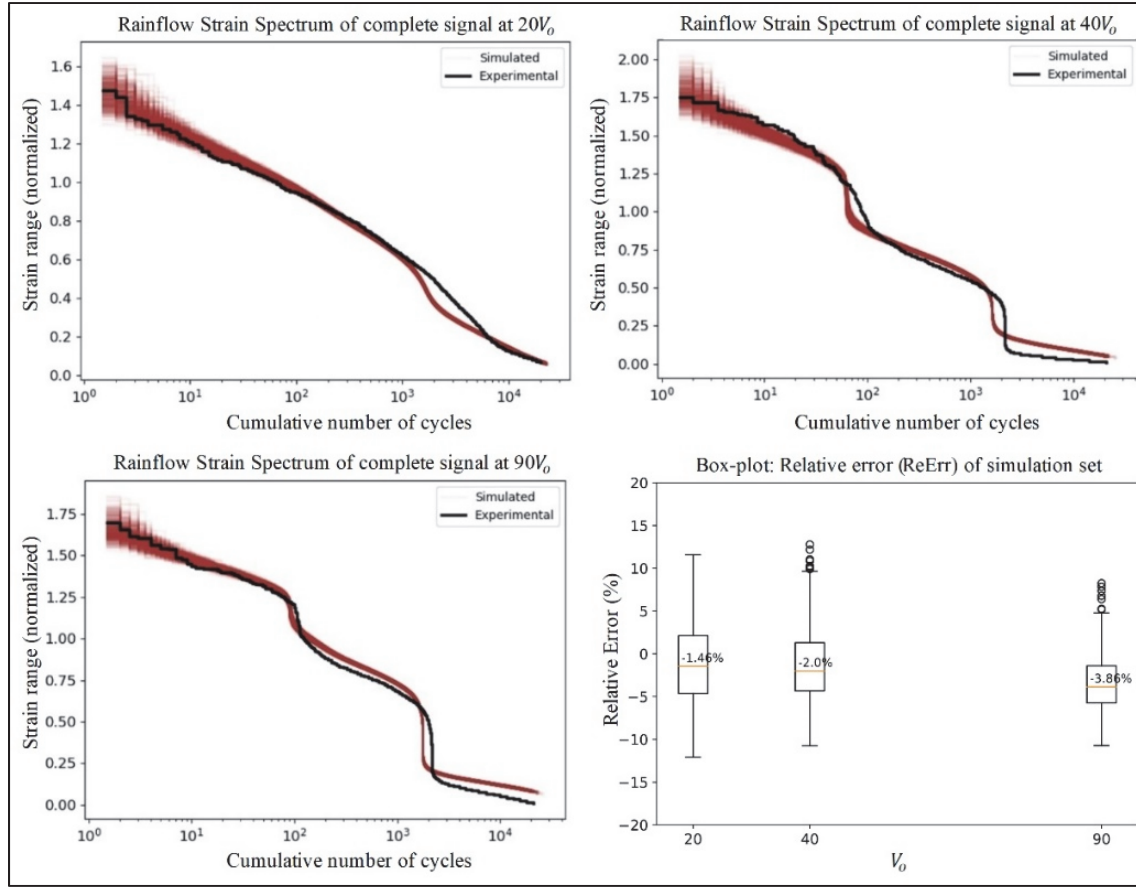


Figure 4.12 Rainflow Strain Spectra of 500 simulations and experimental value for full signals at $20V_o$, $40V_o$, $90V_o$

This interpolation process has the advantage of creating a mathematical model that quickly reproduces the information (amplitude range and physical behaviours) of the strain signal at missing operating conditions. This can help optimize the experimental measurement requirement or reduces the calculation costs for CFD simulations. Therefore, the proposed interpolation process can become a support tool during the turbine design or turbine operation phase.

CHAPITRE 5

SIMULATION OF NON-GAUSSIAN STOCHASTIC PROCESSES WITH PRESCRIBED RAINFLOW CYCLE COUNT USING SHORT-TIME FOURIER TRANSFORM

Quang Hung Pham^{a, b}, Jérôme Antoni^b, Antoine Tahan^a, Martin Gagnon^c, Christine Monette^d

^a École de technologie supérieure (ÉTS), 1100 Notre-Dame Ouest,
Montréal, Québec, Canada H3C 1K3

^b Université de Lyon, INSA Lyon, LVA, EA677, Bâtiment St Exupéry, 25 bis Avenue Jean
Capelle, 69621 Villeurbanne Cedex, France

^c Institut de recherche d'Hydro-Québec (IREQ), 1800 Boulevard Lionel-Boulet,
Varenes, Québec, Canada J3X 1S1

^d Andritz Hydro Canada Inc, 6100 Route Transcanadienne, Point-Claire,
Québec, Canada H9R 1B9

Article soumis pour publication, octobre 2021.

5.1 Contribution à la thèse

Les méthodes proposées dans Chapitre 3 et 4 ont permis de créer une base complète d'outils pour l'interpolation entre des régimes de fonctionnement. Le défi scientifique qui demeure est d'améliorer la qualité de l'interpolation dans le but de maximiser la capture des informations à partir des signaux de contraintes/déformations disponibles. Plus spécifiquement, dans Chapitre 4, les parties résiduelles, après avoir extrait les périodicités, sont reproduites aléatoirement en inversant la matrice de comptage rainflow. Avec une telle approche, le niveau de fatigue a bien été reproduit, mais certains comportements stochastiques n'ont pas été capturés (ex., les flux instables qui apparaissent pendant des régimes de petit chargement et qui créent de grandes fluctuations d'amplitudes à basses fréquences). Par conséquent, l'objectif de la contribution présentée dans le Chapitre 5 est justement de reproduire ces comportements.

L'idée du départ est de créer une simulation d'un processus stochastique non-Gaussien (NG) qui représente les parties stochastiques de la déformation de la roue hydroélectrique. Le

principe de base de cette simulation est de régénérer les coefficients de la TFCT du signal étudié. Les coefficients de la TFCT sont générés via une combinaison entre des séquences de variables indépendantes et identiquement distribuées et un modèle non-linéaire qui dépend de la fréquence. La reproduction du niveau de fatigue est assurée en utilisant une procédure de minimisation de l'écart entre le comptage rainflow simulé et celui attendu (à partir du signal original). Finalement, les paramètres clés peuvent être interpolés sur la plage de régimes de fonctionnement de la turbine, ce qui permet de compléter l'outil d'interpolation tel que défini par le mandat de notre projet.

Cet article apporte des contributions scientifiques et industrielles :

- une méthode originale pour la simulation d'un processus stochastique NG qui utilise la TFCT est développée et validée. D'après notre état de l'art, il n'y a pas encore de recherche publiée qui utilise la TFCT pour simuler un processus non-Gaussien. L'utilisation de la TFCT permet d'imposer plus facilement une PSD donnée, ce qui représente plus adéquatement les différentes natures stochastiques du signal des déformations. De plus, elle ouvre aussi une possibilité de capturer la distribution de probabilité conjointe, au lieu de capturer seulement la distribution marginale comme certaines autres simulations NG;
- la méthode développée procure une possibilité d'interpolation des paramètres de simulation ce qui permet ensuite de reproduire les informations plus riches sur la prédiction de la partie stochastique (résiduelle) de contraintes/déformation de la roue hydroélectrique;
- la méthode reproduit mieux les différents comportements complexes de la déformation sur la roue hydroélectrique. La méthode peut être appliquée dans l'industrie pour la génération de longs signaux de déformations représentatifs et précis, ce qui contribue à l'évaluation de fatigue de la roue.

5.2 Résumé de l'article en français

Les comportements stochastiques représentés par des processus non gaussiens sont omniprésents dans plusieurs domaines scientifiques. Spécifiquement, l'analyse de ces comportements est essentielle dans les analyses d'endommagement de la structure où l'apparition aléatoire des valeurs extrêmes est primordiale. En pratique, les informations disponibles sur ces processus ne sont pas suffisantes pour obtenir une connaissance complète à cause de limitations expérimentales (signaux courts dans le temps et obtenus sur seulement quelques régimes). Par conséquent, une simulation stochastique est incontournable pour la prédiction de ces comportements. Cette recherche présente une méthode de simulation pour recréer des processus stochastiques non gaussiens qui permettent de capturer une densité spectrale donnée et un comptage de cycles rainflow attendue. L'idée capitale est de modéliser le comportement non gaussien en utilisant une non-linéarité sur les coefficients de la transformée de Fourier à court terme. La densité spectrale est directement imposée comme telle, tandis que la non-linéarité est optimisée en minimisant l'écart entre les diagrammes du comptage rainflow du modèle et celui de l'origine (expérimental). De plus, il est possible d'interpoler les paramètres de simulation entre de différents régimes de fonctionnement de la structure. Ceci permet de simuler adéquatement le processus stochastique dans des régimes non mesurés. La méthodologie proposée est validée sur un cas d'étude du signal de déformation mesuré sur la roue d'une turbine hydroélectrique.

5.3 Abstract

Random phenomena described by non-Gaussian processes are ubiquitous in many scientific fields. Properly understanding their characteristics is crucial in safety analysis, where the random occurrence of extreme values is often critical. In practice, the information available for fully characterizing a stochastic process is often limited by experimental constraints. Thus, prediction based on simulations comes as a valid alternative. Motivated by applications in structural fatigue assessment, this research presents a simulation method for non-Gaussian stochastic processes that is able to jointly reproduce a prescribed power spectral density and a rainflow cycle counting. The main idea is to model non-Gaussian behaviour by a frequency-

dependent non-linearity in the short-time Fourier transform. As such, the power spectral density is directly imposed, while the non-linear model is optimized to match the rainflow cycle count. It is shown that the model is flexible enough to be interpolated between different operating points, so as to allow the simulation of stochastic processes at non-measured operating conditions. The full methodology is eventually illustrated for assessing the fatigue assessment of a hydroelectric turbine runner.

5.4 Introduction

Non-Gaussian (NG) random processes often occur in cases with complex loading or in cases with strong excitations such as gas or fluid flow excitation in wind turbines and hydroelectric turbines and can lead to accelerated damage. Therefore, the simulation of this NG random process is essential step to explore the stochastic behaviour of a structure. Thus, this NG simulation has become a concern for many researchers (Bocchini & Deodatis, 2008; Cai & Lin, 1996; Cao et al., 2000; Deodatis & Micaletti, 2001; Grigoriu, 1998; Masters & Gurley, 2003; Poirier et al., 2017; Yamazaki & Shinozuka, 1988). In particular, the simulation of NG processes aims at providing a mean to better understand their effect on structural fatigue and are for instance used to generate long-term loading signals (stress, strain, vibration...) needed in cycle counting. In general, simulations are useful substitutes whenever direct measurement are limited or impossible.

The simulation of a NG stochastic process is expected to reproduce some of its critical properties. As far as fatigue analysis is of concern, the engineering practice primarily focuses on the power spectral density (PSD). Spectral-based methods were early introduced to simulate stochastic processes (Grigoriu, 1998; Shinozuka & Deodatis, 1991; Yamazaki & Shinozuka, 1988) with a prescribed PSD, while assuming that the probability density function (PDF) is Gaussian. These popular methods have been used in numerous applications, such as the fatigue analysis of hydroelectric turbines (Poirier et al., 2017) and long-span bridges (Cao et al., 2000). A more difficult goal is to jointly prescribe a PSD and a non-Gaussian PDF (Cai & Lin, 1996; Deodatis & Micaletti, 2001). Although imposing a NG-

PDF can be achieved by various transformations, the latter will also modify the PSD in general. A natural approach is therefore to iteratively impose the PSD and the PDF until convergence is reached. The general methodology due to Grigoriu (1995) makes use of the concept of translation field. A review of such methods is for instance given by Bocchini et Deodatis (2008); (H. Li & Zhang, 2013; Masters & Gurley, 2003). However, in general, reproducing a pair of prescribed PDF and PSD does not guarantee to properly describe every NG stochastic process. This is because the PDF, as it is meant in the aforementioned works, is the marginal distribution over time and not the joint PDF over all-time instants. In other words, processes with different joint PDFs may well have the same marginal PDF. The consequence is that the simulated stochastic processes may not properly account for the actual fatigue damage engendered to a structure. In fatigue analysis, one characteristic that critically matters is the distribution of high-amplitude cycles. Some authors have addressed this problem by extrapolating the rainflow cycles count or by using the extreme value theory (Pär Johannesson, 2006; Pär Johannesson & Thomas, 2001; D. F. Socie & Pompetzki, 2004). The present work follows a different route and introduces a technique for simulating NG processes with an imposed rainflow cycle count (RCC). Taking the RCC – which returns the cumulative distribution of load cycles (Marsh et al., 2016) – is indeed one of the most relevant characteristics of a stochastic process in fatigue analysis.

The present work focuses on the simulation of a NG stochastic process that contributes to the fatigue damage of a structure by representing the extreme amplitude fluctuations. The strategy is to simulate a stochastic process from its inverse Short-Time-Fourier-Transform (STFT), where the STFT coefficients are generated to match a given RCC. The method consists in first generating independent and identically random coefficients, then to apply a frequency-dependent nonlinearity, and finally to correlate them by application of a filter. The parameters of the nonlinear unit and of the correlation filter are first estimated from a preliminary STFT, as obtained from a partial observation of the stochastic process. The two key elements of the proposed methodology are the use of the STFT, on the one hand, and the use of a specific nonlinear unit, on the other hand.

In the state-of-the-art, the STFT is widely used as an analysis tool in domains such as speech analysis, acoustics, radar/sonar, etc. (Griffin & Lim, 1984; Nawab et al., 1983; Schafer & Rabiner, 1973; Sturmel & Daudet, 2011; Zhu et al., 2019). Its potential for signal synthesis has been early demonstrated, in particular from partial information of the STFT magnitude (Nawab et al., 1983; Sturmel & Daudet, 2011) or ridges (Zhu et al., 2019). However, as far as the authors know, the STFT has never been used to simulate NG stochastic processes. The STFT provides an invertible time-frequency representation of any signal, which by construction is easily amenable to the imposition of a given PSD. At the same time, the STFT makes it possible to introduce a nonlinearity in the time-frequency domain, which may be frequency – dependent as well as dynamic – i.e., involving time derivatives or time lags -- thus resulting in a joint NG PDF.

The choice of the nonlinear unit is crucial to match a given RCC or, more generally, a given joint PDF. For instance, static nonlinearity is often used due to their simplicity to match a marginal PDF. Furthermore, the choice between a pre-nonlinearity model (e.g., Hammerstein model (Narendra & Gallman, 1966)) or a post-nonlinearity model (e.g., Wiener model (Schetzen, 1980)) also influences the simulation process. The post-nonlinearity model makes it easier to impose a NG-PDF in the simulation process, but it modifies the PSD in the output. On the contrary, the pre-nonlinearity model easily allows the imposition of a prescribed PSD, but it then modifies the PDF. The simulation method proposed in this paper is of the latter type, yet with a dynamic nonlinearity that proceeds from a generalization of the autoregressive conditionally heteroskedastic model (ARCH), coined λ -ARCH. The advantage of introducing an λ -ARCH model in the STFT is its ability to represent a large class of NG stochastic processes, with complex structures in the time-frequency domain. A typical example is a broadband stochastic process with a nonlinearity localized in a frequency band, or one with different nonlinearities in distinct frequency bands.

A last contribution of the paper is to demonstrate that the key parameters of the model used for simulation can be interpolated over different operating conditions, so as to simulate a signal at a non-measured condition.

The performance of the proposed method is demonstrated to assess the fatigue endured by a hydroelectric turbine runner. The hydroelectric turbine is a typical example where experimental measures are difficult to set up to obtain and only possible at a limited number of operating conditions due to experimental constraints; therefore, unmeasured operating conditions can conveniently be replaced by simulations to properly characterize the whole operating range.

The paper is organized as follows. Section 5.5 introduces the STFT decomposition of a NG stochastic process and a method to simulate it based on the principle of λ -ARCH model. Section 5.6 then briefly explains how to interpolate a simulation model between operating configurations. Finally, the case study of a hydroelectric turbine runner case study is discussed in section 5.7.

5.5 Simulation Methodology

5.5.1 Signal Model

5.5.1.1 STFT decomposition

As mentioned in the introduction, the aim of the present work is to introduce a method to simulate realizations of a NG stochastic process, as it would be measured in practice, for instance at the *output* of a system. Therefore, it makes sense to consider discrete time from the onset. Let $s(n) \in \mathbb{R}$ denote a discrete-time, zero-mean stationary NG stochastic process, sampled with frequency F_s , with autocorrelation function

$$R_s(\tau) = \mathbb{E}\{s(n)s(n - \tau)\} \quad (5.1)$$

where $\mathbb{E}$ is the expected value operator. The PSD is $S_s(f) = \mathcal{F}\{R_s(\tau)\} = \sum_{\tau \in \mathbb{Z}} R_s(\tau) e^{-2j\pi\tau f/F_s}$. The STFT of the stochastic process $s(n)$ is defined by the random field

$$S_{ik}(R, \Delta) = \sum_{n=iR}^{iR+N_w-1} s(n)w(n-iR)e^{-2j\pi nk\Delta}, \quad S_{ik} \in \mathbb{C}, \quad S_{ik} = S_{i,-k}^* \quad (5.2)$$

or S_{ik} for short, where i, k stand for the time and frequency indices, respectively. In the above equation, $\varphi_{ik}(n) = w(n-iR)e^{-2j\pi nk\Delta}$ is a time-frequency kernel rooted on a window function $w(n)$ with length N_w (assumed even), R is the shift between two consecutive kernels, and $\Delta = 1/N_w$ is the frequency resolution. Alternatively, the stochastic process $s(n)$ can be expanded from its STFT as

$$s(n) = \sum_{i \in \mathbb{Z}} \sum_{k=0}^{N_w-1} S_{ik} \varphi_{ik}^*(n) = 2 \sum_{i \in \mathbb{Z}} \sum_{k=1}^{N_w/2} \Re\{S_{ik} \varphi_{ik}^*(n)\} \quad (5.3)$$

provided that the window satisfies the condition $\sum_{i \in \mathbb{Z}} w(n-iR) = 1$ (assumption H1). The latter condition is for instance satisfied with a Hann window and $R = N_w/4$, or can be forced in general by setting $w(n) = v(n)/\sum_{i \in \mathbb{Z}} v(n-iR)$ based on any positive and smooth function $v(n)$, provided that R is small enough. It is noteworthy that the second equality in the equation (5.3) holds for any zero-mean real stochastic process, with zero Nyquist frequency.

The idea is to simulate a stochastic process from its expansion given by equation (5.3), where S_{ik} is directly sampled in a random field with properties yet to be defined. As explained in the introduction, this is likely to offer flexibility for generating NG processes. Indeed, under assumption H1, the STFT is a highly redundant transform – there are about N_w/R as many STFT coefficients as time samples in the original signal which provides a high number of degrees of freedom for simulating a stochastic process.

From now on, the issue then reduces to generating the random field S_{ik} , such as to reproduce a joint PDF as similar as possible to those of $s(n)$. The approach is to model S_{ik} as a correlated random field at the output of a nonlinear unit.

5.5.1.2 Intrinsic correlation

The first constraint when simulating S_{ik} is to match the PSD of $s(n)$. This is equivalent to imposing a correlation structure to S_{ik} . From equations (5.1) and (5.2), the STFT coefficients should verify

$$\mathbb{E}\{S_{ik}S_{i-p,k-q}^*\} = \sum_{\tau \in \mathbb{Z}} R_s(\tau) e^{-2j\pi\tau(k-q)\Delta} S_w(\tau, i, p, q) \quad (5.4)$$

where

$$S_w(\tau, i, p, q) = \sum_{n \in \mathbb{Z}} w(n - iR) w(n - \tau - (i - p)R) e^{-2j\pi nq\Delta} \quad (5.5)$$

Now, applying the change of variable $m = n - iR$, the latter becomes

$$\begin{aligned} S_w(\tau, i, p, q) &= e^{-2j\pi iq\Delta R} \sum_{m \in \mathbb{Z}} w(m) w(m - \tau + pR) e^{-2j\pi mq\Delta} \\ &= e^{-2j\pi iq\Delta R} C_w(\tau, p, q) \end{aligned} \quad (5.6)$$

Finally,

$$\mathbb{E}\{S_{ik}S_{i-p,k-q}^*\} = e^{-2j\pi iq\Delta R} \sum_{\tau \in \mathbb{Z}} R_s(\tau) e^{-2j\pi\tau(k-q)\Delta} C_w(\tau, p, q) \quad (5.7)$$

which shows that the correlation of the STFT coefficients is a periodic function of i with period $1/(q\Delta R)$, yet independent of k . Reproducing this correlation structure is necessary in simulations for properly matching the PSD of $s(n)$. When the frequency resolution is set

such that $1/\Delta$ is larger than the effective duration⁶ of $R_{ss}(\tau)$ (assumption H2), the following approximation actually holds,

$$\begin{aligned}\mathbb{E}\{S_{ik}S_{i-p,k-q}^*\} &\approx \delta_{q0}C_w(\tau, p, 0) \sum_{\tau \in \mathbb{Z}} R_s(\tau) e^{-2j\pi\tau k\Delta} \\ &= \delta_{q0}C_w(\tau, p, 0)S_s(f_k),\end{aligned}\tag{5.8}$$

with δ the Kronecker symbol, which indicates that the correlation between frequency components can be ignored as compared to correlation between time instants. Under assumption H2, the correlation of the STFT coefficients is now a function of frequency k only. the equation (5.8) draws a direct connection between the variance of the STFT coefficients and the PSD $S_s(f_k)$ at frequency $f_k = k\Delta F_s$. This suggests the normalisation

$$S_{ik} = \sqrt{P_s(f_k)}U_{ik}\tag{5.9}$$

where $P_s(f_k) = \mathbb{E}\{|S_{ik}|^2\}$, such that the variance of the coefficients U_{ik} is now independent of k .

Based on these results, the correlation structure given by the equation (5.8) may be conveniently forced in simulations by using the autoregressive (AR) model

$$\left\{ \begin{array}{l} S_{ik} = \sqrt{P_s(f_k)}U_{ik} \\ U_{ik} = \sum_{p=1}^P U_{i-p,k}a_p + Z_{ik} \\ \mathbb{E}\{Z_{ik}Z_{i-p,k-q}^*\} = \delta_{p,0}\delta_{q,0}\sigma_Z^2 \end{array} \right.\tag{5.10}$$

⁶ The effective duration is defined as the shortest time d such that $|R_s(\tau)| < \epsilon R_s(0)$ for all $|\tau| > d$, with $0 < \epsilon < 1$ a given small value. This means that magnitude of $R_s(\tau)$ is considered insignificant for time-lags greater than d

with orders P , Z_{ik} an uncorrelated and stationary sequence with variance σ_Z^2 , and $\{a_p; p = 1, \dots, P\}$ a set of AR coefficients. Indeed, substituting the above equation for U_{ik} in the autocorrelation $\mathbb{E}\{U_{ik}U_{i-p,k-q}^*\}$ verifies the equation (5.8) by construction. The free parameters of the above model are $P_s(f_k)$ and a_p . How to properly set their values to match the target PSD will be explained in subsection 5.5.2.

5.5.1.3 Nonlinear unit

Forcing the particular correlation structure (5.4) by means of the AR model (5.6) makes it possible to impose a given PSD. The aim is now to impose a given NG PDF by application of a nonlinear unit to the random coefficients Z_{ik} . Introducing a nonlinearity at this stage offers the possibility to generate a large family of joint NG PDFs, since different nonlinear units can be used depending on the frequency index k . The nonlinearity unit used in this paper is inspired by the autoregressive conditional heteroscedasticity (ARCH) model. ARCH is often used to describe leptokurtic stochastic processes (whose PDF has higher tails than the Gaussian) with high fluctuations (Hughes, King, & Kwek, 2004), i.e. with a variance that is conditionally time-varying. In the present context, a generalization is proposed, coined λ -ARCH. It reads

$$Z_{ik}|\{Z_{i-1,k}, \dots, Z_{i-J,k}\} \sim \mathcal{N}(0, \sigma_{ik}^2) \quad (5.11)$$

where $Z_{ik}|\{Z_{i-1,k}, \dots, Z_{i-J,k}\}$ stands for the random variable Z_{ik} at time i conditioned on the random variables $\{Z_{i-1,k}, \dots, Z_{i-J,k}\}$ at previous time instants, and σ_{ik}^2 is a time-varying and frequency-dependent variance that follows the AR equation

$$\sigma_{ik}^{2\lambda_k} = \alpha_{0k} + \sum_{j=1}^J \alpha_{jk} |Z_{i-j,k}|^{2\lambda_k} \quad (5.12)$$

with $0 \leq \lambda_k$, $\alpha_{0k} > 0$, $\sum_{j=1}^J \alpha_{jk} < 1$.

The intuition beyond the λ -ARCH model is that there exists a parameter λ and a set of coefficients $\{\alpha_{jk}; j = 0, \dots, J, k = 1, \dots, N_w/2\}$ such that $|Z_{i,k}|^{2\lambda_k} - \alpha_{0k} - \sum_{j=1}^J \alpha_{jk} |Z_{i-j,k}|^{2\lambda_k}$ is a Gaussian stochastic process uncorrelated with respect to time i . It can be shown that Z_{ik} defined by equations (5.11) and (5.12) is a stationary NG process with λ -moment

$$\mathbb{E}\{|Z_{i,k}|^{2\lambda_k}\} = \frac{\alpha_{0k}}{1 - \sum_{j=1}^J \alpha_{jk}} \quad (5.13)$$

As a particular case when $\lambda = 1$, one recovers the ARCH model with variance $\sigma_{Z,k}^2$ returned by the equation (5.13). The particular case with $\lambda = 0$ reads $\sigma_{ik}^2 = \exp(\alpha_{0k} + \lambda \sum_{j=1}^J \alpha_{jk} \ln |Z_{i-j,k}|^2)$.

The exact class of NG distribution that can be generated by the proposed model is an interesting but difficult question to answer. It can be easily implied by equations (5.11) that such distributions are characterized by a finite variance and a finite moment of order 2λ . In particular, this implies leptokurtic distributions when $\lambda = 2$. More generally, it should be noted that the large set of free parameters in the λ -ARCH model will provide high flexibility to approximate a diversity of NG joint PDFs. An important particular case is when the ARCH parameters are independent of frequency k , or piece-wise constant with respect to large frequency bands, which results in a substantial reduction of the number of free parameters (assumption H3).

The simulation of the stochastic process $s(n)$ then proceeds as described in Algorithm 5.1 .

Algorithme 5.1 Simulation of a NG stochastic process
from the STFT

INPUT: $P_k, \{a_p; p = 1, \dots, P\}, \{\alpha_j; j = 0, \dots, J\}, \lambda_k$
OUTPUT: $s(n)$

1. **Loop** on k
2. Initialize i.i.d sequence $\{Z_{J-1,k}, \dots, Z_{0,k}\}$
3. **Loop** on i
 - Evaluate $\sigma_{ik}^{2\lambda_k}$ from the equation (5.12)
 - Draw Z_{ik} in $\mathcal{N}(0, \sigma_{ik}^2)$
 - Evaluate U_{ik} from the equation (5.10)
 - Set $S_{ik} = \sqrt{P_k} U_{ik}$
4. Evaluate $s(n)$ from the equation (5.3)

5.5.2 Model identification

The signal model introduced in the previous section involves several parameters to identify experimentally, from a given measurement.

5.5.2.1 Identification of the correlation structure

Let the finite length signal $\{x(n); n = 0, \dots, L - 1\}$ denote the partial observation of one realization of the stochastic process $s(n)$. The first step is to compute the STFT coefficients X_{ik} of $x(n)$ from the equation (5.1). Note that index i will range from 0 to $I = \lfloor (L - N)/R \rfloor + 1$. Under assumption H2, the variance of the STFT coefficients can then be estimated as

$$\hat{P}_s(f_k) = \frac{\sum_{i=0}^{I-1} |X_{ik}|^2}{I} \quad (5.14)$$

Once the spectral signature of the process is captured, the STFT coefficients are equalized:

$$V_{ik} = \frac{X_{ik}(R, \Delta)}{\sqrt{\hat{P}_s(f_k)}} \quad (5.15)$$

The coefficients in the AR model of the equation (5.10) can now be estimated by least squares. Let $\mathbf{V}_k = (V_{P,k} \ V_{P+1,k} \ \dots \ V_{I-1,k})^T$, $\mathbf{a} = (a_1 \ a_2 \ \dots \ a_p)^T$ and $\mathbf{B}_k$ the matrix with elements $B_{ip} = V_{i-1+p-p,k}$. Therefore, the least-square estimator of $\mathbf{a}$ is returned by

$$\hat{\mathbf{a}} = \arg \min_{\mathbf{a}} ||\mathbf{V}_k - \mathbf{B}_k \mathbf{a}|| = \left(\sum_{k=1}^{N_w/2} \mathbf{B}_k^T \mathbf{B}_k \right)^{-1} \sum_{k=1}^{N_w/2} \mathbf{B}_k^T \mathbf{V}_k \quad (5.16)$$

The next step is to use the estimated autocorrelation coefficients to whiten the random field V_{ik} , such as to produce the uncorrelated sequence:

$$\hat{Z}_{ik} = V_{ik} - \sum_{p=1}^P \hat{a}_p V_{i-p,k}, \quad (5.17)$$

from which the λ -ARCH model can now be identified.

5.5.2.2 Identification of the nonlinear unit

Let $\mathcal{B}$ denote a set of frequency indices where the nonlinear unit applies uniformly (i.e., where the coefficients of the λ -ARCH model are constant). Then, for a given value of λ , apply the Box-Cox transformation (Sakia, 1992):

$$E_{ik}^{(\lambda)} = \begin{cases} |\hat{Z}_{ik}|^{2\lambda}, & \text{otherwise} \\ \ln |\hat{Z}_{ik}|^2, & \lambda = 0 \end{cases} \quad (5.18)$$

for $k \in \mathcal{B}$. From the equation (5.12), it can be shown that $E_{ik}^{(\lambda)}$ follows the autoregressive equation

$$E_{ik}^{(\lambda)} = \alpha_0 + \sum_{j=1}^J \alpha_j E_{i-j,k}^{(\lambda)} + \eta_{ik}, \quad k \in \mathcal{B} \quad (5.19)$$

where η_{ik} stands for a zero-mean residual error uncorrelated with $E_{ik}^{(\lambda)}$. This offers a direct means to identify the coefficients $\{\alpha_j; j = 0, \dots, J\}$. Taking the average $\langle E_{ik}^{(\lambda)} \rangle = \sum_{i=0}^{I-1} E_{ik}^{(\lambda)} / I$, one has the set of equations

$$\begin{cases} \alpha_0 = \langle E_{ik}^{(\lambda)} \rangle \left(1 - \sum_{j=1}^J \alpha_j \right) \\ E_{ik}^{(\lambda)} - \langle E_{ik}^{(\lambda)} \rangle = \sum_{j=1}^J \alpha_j (E_{i-j,k}^{(\lambda)} - \langle E_{ik}^{(\lambda)} \rangle) + \eta_{ik} \end{cases} \quad (5.20)$$

The first equation returns an estimate of α_0 given the α_j 's. The second equation can be solved for the α_j 's, for instance by least-squares as in the previous paragraph (Vaidyanathan, 2007). The complete identification scheme is resumed in Algorithm 5.2 under assumption H3.

Algorithm 5.2 Identification of model parameters from
an observed signal $x(n)$

INPUT: $x(n)$, P , J , λ

OUTPUT: $\{\hat{a}_p; p = 1, \dots, P\}$ and $\{\hat{\alpha}_j; j = 0, \dots, J\}$

1. Compute the STFT coefficients X_{ik} from the equation (5.2)
2. Evaluate $\hat{P}_s(f_k)$ from the equation (5.14)
3. Evaluate V_{ik} from the equation (5.15)
4. Compute the AR coefficients a_p from the equation (5.16)
5. Evaluate $\hat{Z}_{ik}$ from the equation (5.17)
6. Get $E_{ik}^{(\lambda)}$ from the Box-Cox transform equation (5.18)
7. Estimate the AR coefficients α_j from the equation (5.20)

For the input of Algorithm 5.2, the parameter λ of the Box-Cox transformations can be optimized by maximizing the log-likelihood function, which ensures that Box-Cox

transformation distribution is closest to a Gaussian distribution form. Another simpler strategy is to *a priori* assign it a given value, based on the user experience.

Another parameter can be modified to optimize the nonlinearity such as $\hat{\alpha}_{0k}$. Therefore, in the next section, an additional step will be proposed to improve the capacity of reproducing a prescribed RCC by optimizing $\hat{\alpha}_{0k}$.

5.5.3 Weighted Area Metric for Rainflow Counting

So far, the λ -ARCH identified in section 5.5.2 might not reproduce exactly a prescribed RCC, since the latter does not appear as a target in Algorithm 5.2. There are several reasons why discrepancies might exist. First, the selected model orders P and J might not coincide with the true (unknown) values, a situation that will introduce modeling errors. Second, estimation errors will also alter the model parameters $\{a_p; p = 1, \dots, P\}$ and $\{\alpha_j; j = 0, \dots, J\}$, and will even propagate from the former to the latter. One solution to mitigate these errors is to use the output of Algorithm 5.2 as initial values and then to update these values so as to match a given RCC.

It is noteworthy that there exist several distances dedicated to the comparison of PDFs that can be used in the equation (5.16), such as those based on information theoretic criteria. A distance tailored to fatigue analysis is designed in the next section.

Let θ denote a set of model parameters to be updated. It may include all the free parameters listed in subsection 5.5.1, or a subset of them, such as typically the parameters of the λ -ARCH. Let also R_{cc}^* denote the target RCC and $R_{cc}(\theta)$ the RCC of simulated signal, which depends on θ . The aim is to update the values θ as initially estimated by Algorithm 5.2 by further minimizing some distance between the target RCC and the simulated one, i.e.

$$\hat{\theta} = \arg \min_{\theta} D(R_{cc}^*; R_{cc}(\theta)) \quad (5.21)$$

according to some distance D . In the field of fatigue analysis, the Area Metric (AM) is a popular measure of the distance between two cumulative distribution functions (Y. Liu et al., 2011). The AM between two RCCs is expressed as

$$D(R_{cc}^* ; R_{cc}(\theta)) = \sum_{i=1}^{i_{max}} \beta_i |R_{cc}^*(\Delta n_i) - R_{cc}(\Delta n_i; \theta)| \Delta n_i \quad (5.22)$$

where Δn_i is the number of cycles corresponding to an amplitude level i and β_i is a user-defined weight (see Figure 5.1). In order to give more importance to high amplitude cycles, which are the most critical in fatigue, it is proposed to set $\beta_i = \log(1/n_i)$, where $n_i = \sum_{j=1}^i \Delta n_j$ is the cumulative number of cycles sorted in descending amplitudes.

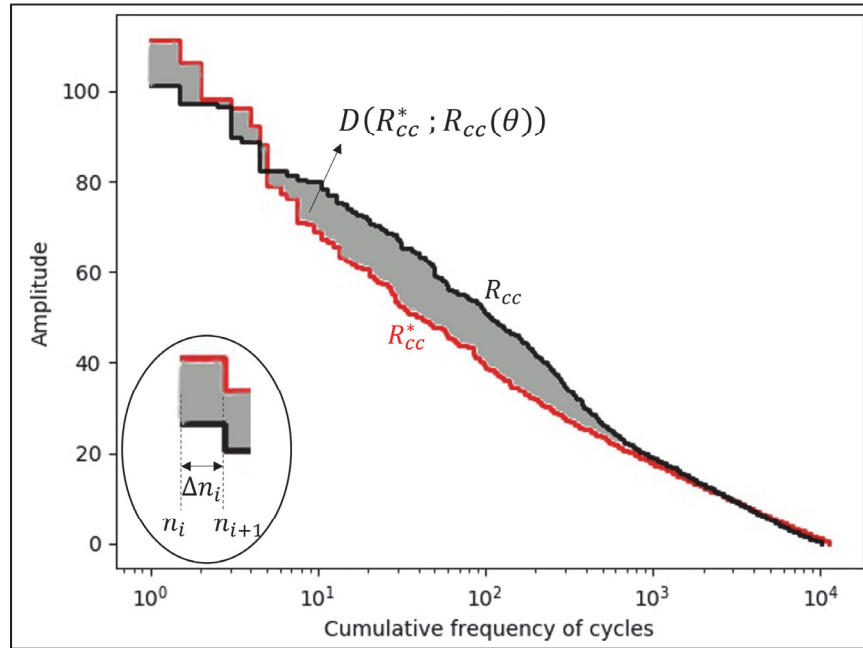


Figure 5.1 Weighted area metric applied to two Rainflow Counting Counts

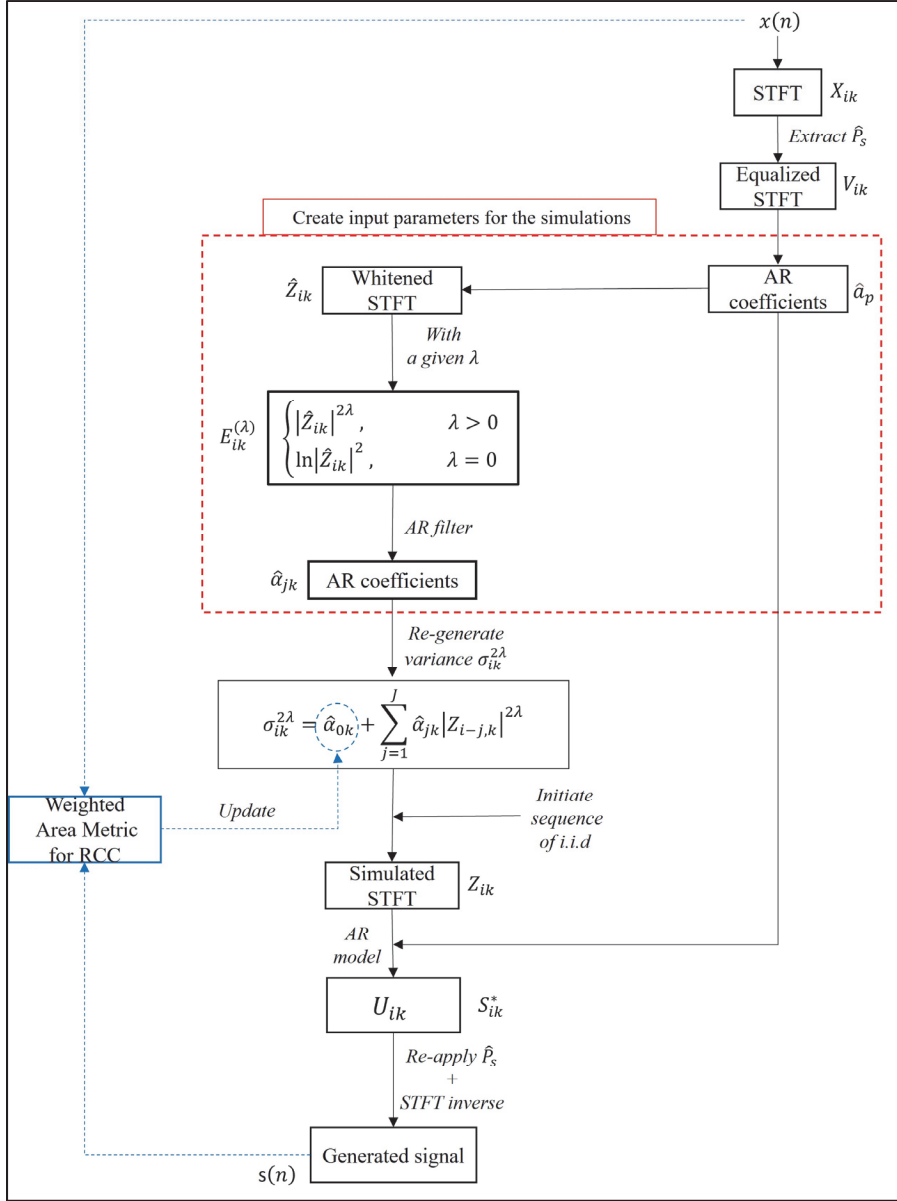


Figure 5.2 Flowchart of the proposed simulation process

To reproduce the correct RCC is of concern, it is proposed to update the coefficient $\hat{\alpha}_0$ such that

$$\sigma_{ik}^{2\lambda} = \hat{\alpha}_0 + \sum_{j=1}^J \hat{\alpha}_j |Z_{i-j,k}|^{2\lambda}, \quad k \in \mathcal{B} \quad (5.23)$$

minimizes $D(R_{cc}^* ; R_{cc}(\theta))$. Initially, $\hat{a}_0$ is first the value returned from Algorithm 5.2 when no updating is required, whereas *increasing* (*resp. decreasing*) this $\hat{a}_0$ will increase (reps. decrease) the effect of the nonlinear unit. Since only one parameter is involved, the minimization in the equation (5.21) is easily achieved for instance with a grid-search method. As shown in section 5.7, this simple updating rule will turn out efficient enough in the applications of interest in this work. Finally, the flowchart of the complete algorithm, including model identification, model updating, and simulation of the stochastic process is shown in Figure 5.2.

5.6 Interpolation Methodology

As explained in the introduction, the simulation of a stochastic process is particularly useful for configurations where experimental data could not be measured. In order to deal with this situation, the idea is to interpolate the model parameters (instead of signals) over different operating conditions where data exist and then to simulate the stochastic process at the operating conditions where measurement are missing. Although all the model parameters introduced in section 5.5.1 can be jointly interpolated, it is wise to keep their number as low as possible in order to guarantee a robust interpolant as well as to reduce the algorithmic complexity. Under the stationary assumption, the autocorrelation coefficients a_p happen to mainly depend on the chosen window parameters. Therefore, they can reasonably be assumed constant over different operating conditions. On the contrary, the interpolation of $P_s(f_k)$ is important if the PSD is likely to vary over different operating conditions. In the case study of this section, it will similarly be assumed that the coefficients α_j of the λ -ARCH model are reasonably constant, except for the coefficients $\hat{a}_0$, λ introduced in the equation (5.23). However, based on the user experience, λ can be fixed using a constant, then only $\hat{a}_0$ need to be interpolated for the non-linear unit of the model.

Since $\hat{a}_0$ is a scalar, it is easily interpolated, for instance with cubic spline interpolation (Dubrule, 1983; Späth, 1995). The interpolation of $P_s(f_k)$ is more intricate as it is a function of frequency. For this kind of interpolation, the Spatio-Temporal Kriging (STK) is chosen

(Montero et al., 2015; Pebesma & Heuvelink, 2016; Sherman, 2011). Let o and f denote the operating condition and frequency variables, respectively. Then, based on known values of the PSD $P_s(o_m, f_l)$ at locations (o_m, f_l) , $m = 1, \dots, M$, $l = 1, \dots, L$, the STK interpolant reads

$$P_s^*(o, f) = \sum_{m=1}^M \sum_{l=1}^L \lambda_{ml}(o, f) P_s(o_m, f_l) \quad (5.24)$$

where superscript $*$ denotes the target value, λ_{ml} is the regression weight, M is the number of known operating conditions and L is the number of studied frequencies at operating condition o_m . With the preservation of the covariance map (or semivariogram) created and modelled from the input set, the weights λ_{ml} in the equation (5.24) are estimated by minimizing the error variance $Var(P_s(p_x, o_y) - P_s^*(p_x, o_y))$. For more details of STK method, the reader can consult references (Montero et al., 2015; Pebesma & Heuvelink, 2016; Sherman, 2011). Finally, a conditional simulation based on kriging parameters is applied to obtain a set of possible interpolated value (Deutsch & Journel, 1992; Nowak & Verly, 2005).

5.7 Case Study

In hydroelectric turbines, the water flow entering the runner creates an extreme environment that can make high stress/strain fluctuation (Seidel et al., 2014). Moreover, with the change of electrical network usage, the turbine must run in some operating conditions at partial load, which increases the high peak in stress/strain signal and the high risk of fatigue on runner. Therefore, a hydroelectric turbine case study will be performed to verify the performance of the proposed simulation/interpolation.

The studied data is a set of strain signals obtained from measurements at steady states on a Francis turbine runner (Breeze, 2018) (a reaction type turbine) in Quebec, Canada. In this case study, to focus on the random part, effort is directed to simulate the residue part remained after extracting the synchronous periodic part. Available strain measurements were

obtained at different vane opening levels with a relatively constant head. Therefore, the operating conditions in this case study are defined using only the vane-opening level (noted as V_o). For example, $20V_o$ corresponds to an operating condition that vanes open about 20% of the maximum opening level (full loading). In the following, the simulation and interpolation based on the proposed methodology will be performed and discussed. Note that some values are shown with normalized values to respect the confidential information.

5.7.1 Simulation Results and Discussion

The first aim of the case study is to illustrate and validate the proposed simulation method on experimental data. Each measurement is available under the temporal signal form (90 seconds). However, the main objective is to contribute to the fatigue analysis of structure which needs the long-term information. In particular, the hydroelectric runner strain measurements by gauges are often available for a limited measurement time. Therefore, to verify the performance of the proposed simulation, a temporal signal generation is highly required. For more detail, in this case study, only 15 seconds of measurement is used for the training set of simulation, and then we try to re-simulate the entire signal of 90 seconds. In this case study, the set $\mathcal{B}$, where the nonlinearity is applied uniformly, corresponds to the set of frequencies lower than $100f_o$, with f_o the frequency corresponding to the steady synchronous runner rotation. Based on the author's experiences on this studied signal, this frequency set contains most of the non-Gaussian characteristic of the signal. The stochastic processes to simulate correspond to the random fluctuations of strain at a steady state regime in which the runner rotates at a constant speed but with no electrical power output (called Speed No Load). This operating condition produces complicated random behaviours, which can significantly contribute to fatigue damage (X. Liu et al., 2016; Seidel et al., 2014).

Stochastic simulations are run with the parameters shown in Tableau 5.1. By default, a Hann window can be safely used as it provides a good compromise between the reduction of frequency leakage and resolution. This comes with 75% overlap ($R = N_w/4$) in order to allow the inversion of the STFT. The window length has been set large enough so as return

an estimated PSD with high enough frequency resolution, i.e., such that increasing N_w would not produce any significant change in the shape of the PSD but only increase its variance. Extensive simulation has shown that an order of 10 for the AR model is good enough in most cases, provided that assumption H2 is met. This turned out almost independent of the types of PSD and PDF. The order of the λ -ARCH and λ have similarly been set to the smallest values that didn't result in significant changes in the AM metric. Using these parameters can allow for a minimal calculation cost while still preserving most interest information.

Tableau 5.1 Set of parameters in the case study

<i>Simulation Parameters</i>	<i>Values</i>
Window function	Hann ($N_w = 2^{12}$)
R	$N_w/4$
Orders of AR model for STFT	$P = 10$
λ	1
J	5
<i>Estimated parameters</i>	<i>Values</i>
$\hat{a}_p$ (real values)	0.0033, 0.008, -0.0014, -0.0093, -0.0007, 0.0024, 0.0003, -0.0030, 0.0031, 0.0019
$\hat{\alpha}_j$ ($j = 1, \dots, J$)	0.189, 0.049, 0.028, 0.039, 0.047
$\hat{\alpha}_0$	4.95

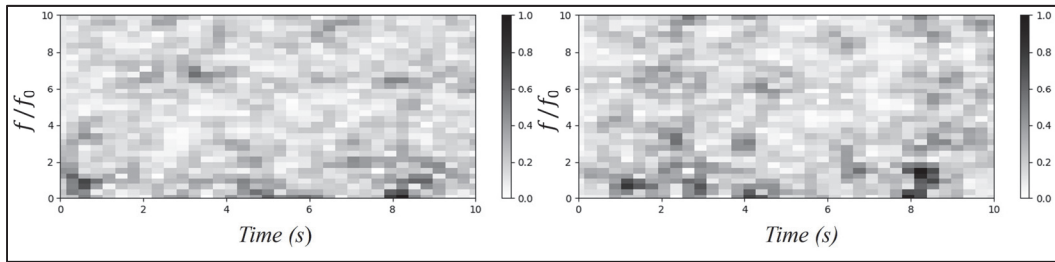


Figure 5.3 (Left) Experimental STFT and (right) STFT of a simulated signal

Figure 5.3 displays one of generated STFTs. Even if the non-linearity is applied uniformly in a wide set of frequencies, the simulations properly generated the high amplitude zone of STFT. In order to verify the performance of the proposed methodology to reproduce RCC for fatigue assessment, the temporal signal generation is compared to two other types of simulation schemes. The first was used for temporal signal generation of residual strain

signal of runners (same signal type with this case study) by Poirier et al. (2017). They suppose that the residual strain signal, which is remained after extracting all periodicities, follows a purely Gaussian distribution and then they applied the spectral-based simulation method (using only a prescribed PSD), as described in Yamazaki et Shinozuka (1988). This Gaussian hypothesis leads to an assumption that certain extreme high peaks are considered as less representative for the actual deformation of the blade (Poirier et al., 2017). However, for fatigue analysis, the extreme values cannot be ignored. In particular, the work of Poirier et al. used a similar type of signal (hydroelectric runner strain), thus, this comparison can show the improvement in the generation of temporal signal for hydroelectric turbines. The second is an iterative simulation using the prescribed PSD and PDF. The second method is an iterative simulation scheme with convergence condition⁷, based on the concept of translation field proposed by Grigoriu (1995), that imposes both a prescribed PSD and a marginal PDF.

Tableau 5.2 Distance metrics calculated from three simulation sets for the case study

	Area metric of RCC	Jensen-Shannon divergence for PDF	Itakura-Saito distance for PSD
Spectral-based simulation	116.14 ± 15.64	0.040 ± 0.008	0.017 ± 0.004
Iterative simulation	90.52 ± 10.89	0.014 ± 0.002	0.024 ± 0.001
Proposed simulation	66.52 ± 9.02	0.034 ± 0.003	0.013 ± 0.001

Figure 5.4 and Tableau 5.2 shows comparisons between three simulations. In this case study, the proposed method has the smallest value of Itakura-Saito distance (IS), which can highlight the advantage of using pre-non-linear model where the PSD is easily imposed. The Jensen-Shannon divergence (JS) of the spectral-based simulation is the highest as the latter is not able to reproduce the NG behaviour of the studied data with presence of extreme values (see distribution tails – Figure 5.4). The JS variability of the proposed simulation is smaller than the spectral-based simulation and is higher than the iterative simulation. Such value is normal because the PDF is not the target input of the proposed simulation. On the other side,

⁷ For each iteration, the convergence point is set when the error (e.g., Itakura-Saito distance) of PSD is smaller than 5% or when a given high number of iterations is reached.

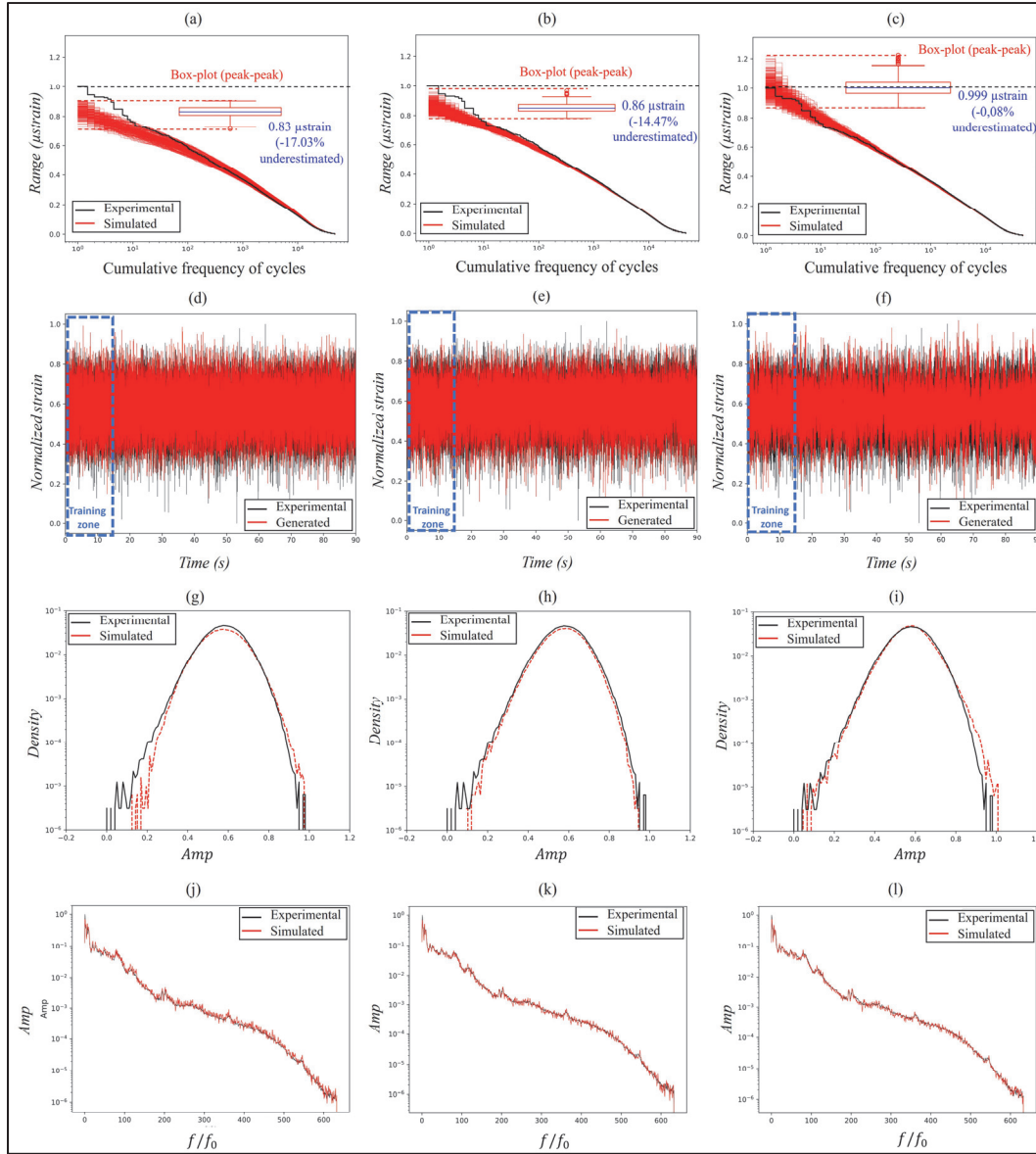


Figure 5.4 Comparisons of the generation performance of the spectral-based simulation (left column), the iterative simulation (middle column) and the proposed simulation (right column) according to the rainflow cycle counts (a, b, c), one of generated signal (d, e, f), their PDF (g, h, i) and their PSD (j, k, l)

Figure 5.4 and Tableau 5.2 show that, even though the reference methods are based on the reproduction of the target PSD and marginal PDF, none of them cover the high amplitude cycles in the RCC. The proposed method does quite well in this respect with a small error (set mean error $< 1\%$ compared to the median of the simulation set, see the boxplot in Figure

5.4). Moreover, the smallest AM value in the proposed simulation also illustrates the advantage of the optimization using RCC.

5.7.2 Interpolation results and Discussion

The second objective of the case study is to illustrate the capability of the proposed method to simulate strain signals at non-measured operating conditions by interpolating between models (section 5.6). Here, one aims to interpolate the model at $V_0 = 20, 45, 70$ by using the identified parameters at $V_0 = 11, 30, 40, 50, 55, 60, 65, 80, 100$. In fact, there is no interested information for the fatigue assessment at frequencies higher than $100f_0$, the signal is downsampled by a factor 7 to reduce the calculation time. The value of λ is fixed as 1 for all the V_0 .

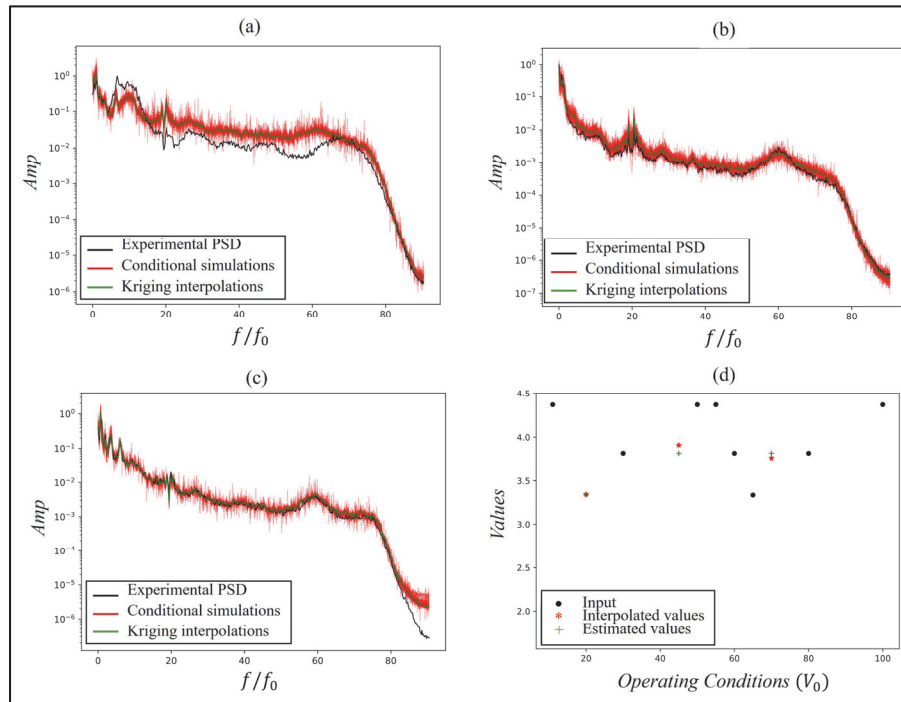


Figure 5.5 Interpolation sets of PSD at $20V_0$ (a), $45V_0$ (b), $70V_0$ (c), and interpolation of $\hat{\alpha}_0$ (d)

Figure 5.5 displays set of interpolations of PSD at three targets operating conditions. The variation of interpolation of the lower load condition is higher than the higher load condition.

The reason is that the $70V_0$ is close to the best efficiency point of the turbine operation, thus, the fluctuation at this condition is more stable than at $20V_0$, $45V_0$.

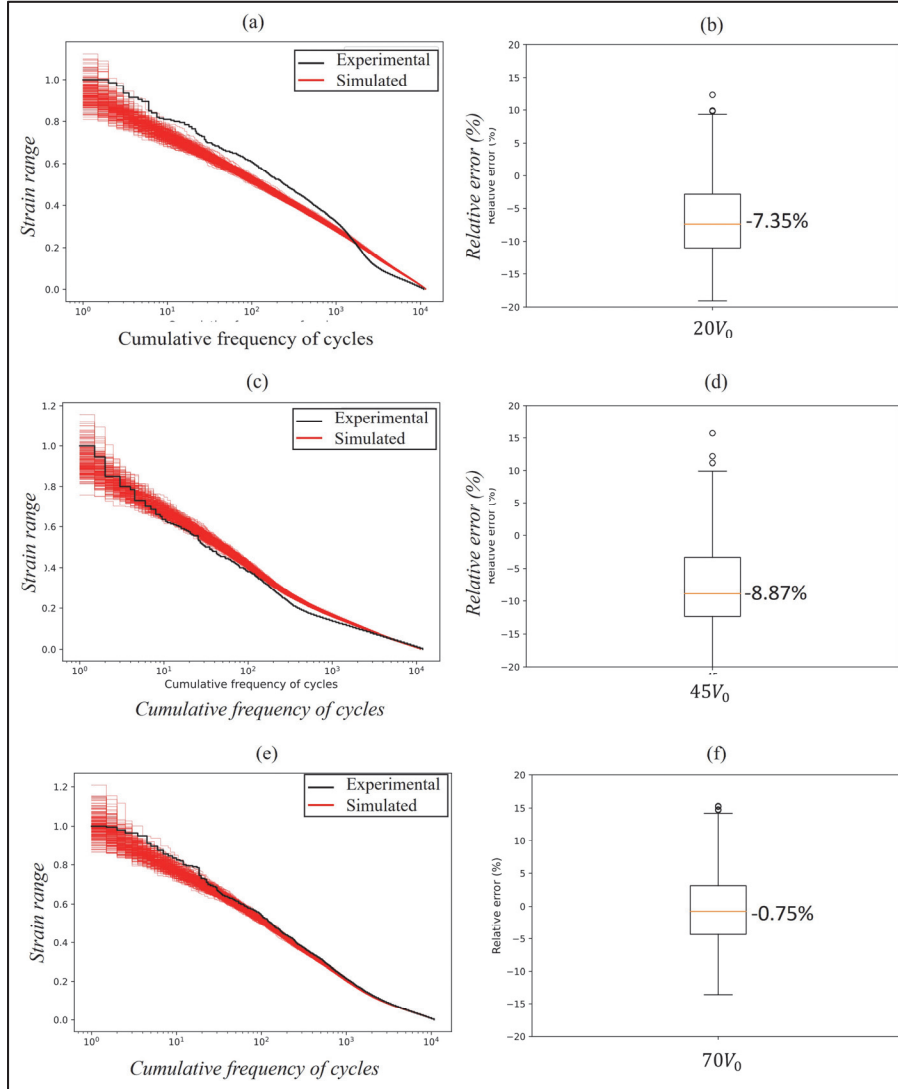


Figure 5.6 RCC and relative error for maximum range of simulations set using interpolated parameters at $20V_0$ (a, b), $45V_0$ (c, d) and $70V_0$ (e, f)

Notice that the interpolation quality depends on the number of inputs, while in the case of experimental measurement, this number is often limited. The quality of the interpolation is assessed hereafter by means of the RCCs of 200 simulations run with the interpolated model, as shown in Figure 5.6. The RCCs show errors at low amplitude while the higher amplitude

is properly covered by the simulation set. There are some outliers far from the set, which might be generated because of the error of interpolation phase. However, the number of the outliers is minor compared to the whole simulation set.

5.8 Conclusion

This paper has introduced a method for simulating non-Gaussian stochastic processes with a prescribed RCC, as typically required in reliability and fatigue analysis. The main idea is to generate the STFT coefficients of the stochastic process after the application of a frequency-dependent non-linearity in the form of a so-called λ -ARCH model. In order to properly reproduce high amplitude cycles, which truly matters in fatigue, the simulation parameters are tuned by minimizing the weighted area metric between the theoretical and the simulated RCCs. Besides, it has been shown that the model used in simulations can be interpolated between unobserved operating conditions, thus offering the possibility to predict signals where measurements are missing. The proposed methodology has been applied to analyze the fatigue of a hydroelectric turbine runner. This case study represents a typical type of data measured from a complex structure where the extreme peaks are sensible for the fatigue assessment. This can show the capacity of this proposed method for application in practice. Future work is needed to theoretically delineate the classes of non-Gaussian processes that can be simulated with the proposed method.

Obviously, the application of the proposed method can find applications in other domains than fatigue analysis, since it is flexible enough way to model a large variety of non-Gaussian processes. In particular, the principle used to reproduce a prescribed RCC can be applied to other types of distributions rooted on the joint PDF of a stochastic process.

CONCLUSION

Ces travaux présentés sont liés à une problématique industrielle qui est le manque d'informations (mesures expérimentales) de l'état de contraintes mécaniques sur une roue d'une turbine hydroélectrique. Les mesures ne sont disponibles que ces courtes périodes temporelles (de l'ordre quelques minutes) et que sur des plages limitées de régimes de fonctionnement. Ce manque peut entraîner de mauvaises estimations de la fatigue de la roue (dommage). Le mandat du projet est le développement d'un outil d'interpolation des signaux de contraintes entre de différents régimes de fonctionnement. Spécifiquement, nous nous intéressons à la turbine du type Francis, majoritairement employée au Québec. La motivation est de créer des modèles d'interpolation avec une erreur acceptable et qui peuvent devenir une alternative validée à la mesure expérimentale.

Nous avons créé un outil complet d'interpolation qui permet de reproduire quasiment tous les phénomènes physiques sur toute la plage de fonctionnement de la turbine à partir d'un nombre limité de mesures expérimentales. L'outil inclut différentes étapes comme le prétraitement du signal, la méthode d'interpolation comme telle et la validation de l'estimation. La famille des méthodes du krigeage est employée comme principal outil pour l'interpolation. Le choix du type de krigeage dépend du prétraitement de signal. En effet, ce dernier décide la forme des entrées. Trois types de prétraitement proposés sont :

- la matrice de comptage rainflow est employée, elle procure l'information sur la distribution des cycles de fatigue;
- les décompositions cyclostationnaires sont employées pour séparer les parties périodiques et stochastiques du signal;
- un processus non-gaussien basé sur la TFCT est employé pour reproduire le comportement stochastique du signal.

Les interpolations proposées pour ces trois jalons correspondent, grosso modo, à nos trois contributions (articles de journaux scientifiques avec comité de lecture) soumises. Les trois

approches d'interpolations sont indépendantes. Néanmoins, les employer conjointement procure un outil complet d'interpolation.

Un premier article publié sur la revue *Renewable Energy* détaille une méthode d'interpolation à partir de la matrice de comptage rainflow (CHAPITRE 3). L'utilisation de cette matrice a simplifié les informations à interpoler en isolant seulement des informations pertinentes à l'estimation de la fatigue (dommage). Le krigeage multidimensionnel, qui est inspiré par le krigeage spatio-temporel, est employé comme une méthode appropriée pour l'interpolation de cette matrice. L'avantage de cette interpolation proposée est de créer une rapide estimation des niveaux d'amplitude de contraintes dynamique de la roue qui contribuent directement à l'analyse de la fatigue, sans besoin de génération d'un signal complet de contrainte.

Au CHAPITRE 4, un nouveau processus d'interpolation basé sur les décompositions cyclostationnaires du signal de déformation est présenté. La contribution a été soumise dans la revue *Renewable Energy*. Trois composantes, nommées parties périodiques, phénomène de torche et partie résiduelles, sont extraites par des opérateurs cyclostationnaires. Elles sont ensuite indépendamment interpolées par différents types de krigeage avant d'être réassemblées. En particulière, l'interpolation de matrice rainflow présentée au CHAPITRE 3 y est améliorée et appliquée pour l'interpolation des parties résiduelles. De plus, nous présentons aussi une méthode qui utilise le CoKrigeage à partir de signaux secondaires (ex., la flexion de l'arbre) qui sont plus faciles à obtenir que les mesures avec des jauges de déformations. Nous avons démontré l'efficacité de cette approche pour interpoler les phénomènes de torche. Ce nouveau processus d'interpolation ouvre une opportunité pour mieux estimer le comportement dans des zones spécifiques sur la plage de fonctionnement de la turbine où le nombre de mesures est limité. Le résultat montré dans cet article indique que les interpolations proposées permettent de capturer non seulement les amplitudes (niveau de fatigue), mais aussi des phénomènes physiques typiques du signal de déformation de la roue hydroélectrique.

Enfin, une troisième contribution soumise sur la revue «Probabilistic Engineering Mechanics» est présentée CHAPITRE 5,. La contribution décrit une méthodologie originale pour la simulation d'un processus stochastique non gaussien en utilisant la TFCT. La motivation de ce travail est dictée par le souci d'améliorer la qualité de l'estimation des parties résiduelles du signal de déformation non captées par le modèle une fois les périodicités extraites. Ces parties résiduelles contiennent certains comportements stochastiques spécifiques que les travaux précédents au CHAPITRE 4 ne peuvent pas reproduire que les travaux précédents au Chapitre 4 ne peuvent pas reproduire (ex., les fluctuations des amplitudes sur de basses fréquences dans certains régimes de petit chargement). L'idée est de générer les coefficients de TFCT en les combinant avec un modèle non linéaire sous la forme de modèle ARCH. Cette approche de modélisation permet une capacité de reproduire les PDF conjointes et la PSD prescrite, ce qui garantit la reproduction fidèle de la nature physique attendue dans les signaux simulés. De plus, il est aussi possible de reproduire les grandes amplitudes (valeurs extrêmes) via un processus de minimisation d'écart entre les courbes de cycles Rainflow simulées et expérimentales. L'interpolation de paramètres clés de cette simulation procure ainsi un outil d'interpolation complet. De plus, vu que le comportement stochastique non gaussien peut aussi être observé dans les autres domaines scientifiques et industriels, cette simulation proposée pourrait élargir ses applications.

En conclusion, ces trois principales contributions présentées dans ce rapport ont permis de construire un outil d'interpolation des déformations de la roue sur toute la plage de fonctionnement de la turbine hydroélectrique. L'interpolation de matrice de rainflow fournit une estimation rapide permettant d'atteindre les niveaux du signal importants pour l'étude de la fatigue, tandis que l'interpolation des décompositions cyclostationnaires fournit une interpolation complète qui maximise les informations estimées sur les déformations de la roue comme telle. En plus, la simulation du processus stochastique non gaussien par TFCT réussit à représenter les comportements stochastiques du signal. Ces trois méthodes d'interpolation contribuent ainsi à atteindre l'objectif final du projet qui est l'amélioration de l'estimation de la fiabilité des roues en assurant la simulation cohérente et représentative. En

conséquence, cet outil peut s'appliquer à l'industrie d'hydroélectricité comme un support de l'analyse de l'endommagement de la roue, ce qui est une phase importante lors de la conception et l'exploitation de la turbine. D'après notre connaissance dans l'état de l'art, l'idée de créer un outil d'interpolation entre des régimes de fonctionnement de la turbine hydroélectrique est nouvelle et originale. Par conséquent, nous espérons que nos travaux pourront créer une base solide pour d'autres recherches scientifiques et ouvrir également des opportunités pour d'autres applications dans l'industrie.

Publications

- Pham, Q. H., Gagnon, M., Antoni, J., Tahan, A. S., & Monette, C. (2019, July). Interpolation of periodic hidden signal measured at steady-operating conditions on hydroelectric turbine runners. In *Surveillance, Vishno and AVE conferences*.
- Pham, Q. H., Gagnon, M., Antoni, J., Tahan, A., & Monette, C. (2021, June). A proposal for the dynamic strain interpolation on hydroelectric turbine runner. In *IOP Conference Series: Earth and Environmental Science* (Vol. 774, No. 1, p. 012083). IOP Publishing.
- Pham, Q. H., Gagnon, M., Antoni, J., Tahan, A., & Monette, C. (2021). Rainflow-counting matrix interpolation over different operating conditions for hydroelectric turbine fatigue assessment. *Renewable Energy*, 172, 465-476.
- Pham, Q. H., Gagnon, M., Antoni, J., Tahan, A., & Monette, C. (2022). Prediction of hydroelectric turbine runner strain signal via cyclostationary decomposition and kriging interpolation. *Renewable Energy*, 182, 998-101.

RECOMMANDATIONS ET PERSPECTIVES

Il est malheureux que des contraintes logistiques et temporelles nous obligent à mettre un terme à notre recherche. Néanmoins, en nous basant sur notre expérience, nous suggérons certaines recommandations pour une campagne de mesure :

- Il faut d'abord identifier approximativement de différentes zones de fonctionnements (via simulation numérique ou via une approximation par rapport au régime optimal, ex., zone de torche est souvent de 50%-85% du débit de BEP). Ensuite, nous pouvons mesurer plus dans des zones complexes (ex., zone de torche). Cette stratégie pourrait améliorer la qualité d'interpolation dans ces zones de fonctionnements.
- Il faut assurer d'obtenir des mesures dans certains régimes à la transition entre des zones identifiées. D'après l'expérience dans cette thèse, les interpolations de ces régimes sont souvent instables à cause de l'influence de deux zones de fonctionnement différentes.

De plus, nous aussi formulons quelques perspectives pour de futurs travaux.

- Dans la méthodologie de l'interpolation des matrices de comptage rainflow (Chapitre 3), nous utilisons la distribution Gaussienne pour générer des cycles d'amplitude. Cependant, cette hypothèse simplificatrice pourra être améliorée en utilisant l'estimation par noyau EPN (avec des coefficients de lissage variés par rapport aux localisations de la matrice de comptage rainflow). Cette idée de génération a été proposée dans la recherche de D. Socie (2001) pour extrapoler temporellement la mesure de chargement. Pourtant, l'utilisation de l'EPN améliorera la génération des cycles, mais elle augmentera la complexité de l'interpolation parce qu'il y aura des coefficients de lissage à interpoler en plus. Nous sommes convaincus qu'il y a une matière à investiguer dans une recherche future.

- La méthode du CoKrigage possède un potentiel certain pour améliorer les interpolations dans le domaine hydroélectrique. Dans ce projet, nous avons utilisé le CoKrigage pour interpoler les phénomènes de torche du signal de déformation en utilisant le signal de flexion de l'arbre comme un signal exogène. Nous conjecturons que la qualité de cette interpolation pourra encore être améliorée si nous pouvons trouver un autre type de signal exogène qui est mieux corrélé avec les mesures de déformation de la roue. Un candidat potentiel pour se procurer ce signal est les simulations numériques (par CFD + EF). Ces simulations ne représentent pas exactement les déformations mesurées sur le prototype à cause des contraintes limitées de calculs, mais elles peuvent fournir une relation entre des régimes. Du côté de l'industrie, le Cokrigage pourrait être appliqué pour interpoler les contraintes de la roue d'une turbine donnée en utilisant les mesures disponibles depuis une autre turbine différente, mais similaire (ex., même type, dimensionnement similaire, etc.). Cette perspective permet de réduire le coût des campagnes de mesures sur de nouvelles turbines, en profitant d'anciennes mesures jugées comme fiables.
- Les décompositions cyclostationnaires utilisées dans ce projet sont limitées par certaines méthodes comme la moyenne synchrone et l'interpolation des phases estimées depuis une bande de fréquences d'intérêts (Poirier et al., 2017). Il y a d'autres méthodes de décompositions cyclostationnaires (surtout sur l'extraction des CS2, par exemple la régression cyclique de rang réduit, etc.) qui pourront être appliquées afin d'améliorer l'extraction des composantes périodiques.
- La PSD est une information importante que nous pouvons interpoler entre des régimes. Dans le CHAPITRE 5, la PSD de la partie stochastique a été directement interpolée entre des régimes en utilisant le krigage spatio-temporel. Cependant, cette méthode d'interpolation est un peu instable dans des régimes de petit chargement à cause de leur comportement stochastique complexe. Pour améliorer cette interpolation, nous pourrions prétraiter les PSD avant d'interpoler. Une des méthodes recommandées est par exemple le modèle de mélange gaussien (MMG). L'utilisation de MMG peut permettre de modéliser la forme de PSD en forme comme un assemblage de plusieurs distributions

- gaussiennes. De plus, il y a aussi d'autres méthodes d'analyse pour séparer les PSD en petites composantes à investiguer (ex., Liftrage dans le domaine Cepstrale) (Oppenheim & Schafer, 2004).
- La technique des valeurs supérieures à un seuil (VSS) est aussi une méthode de prétraitement potentielle pour l'interpolation. Les paramètres du modèle VSS comme les seuils, les paramètres de la distribution extrême (le moyen, la variance, le coefficient de forme...) peuvent être interpolés entre des régimes de fonctionnements. Cette interpolation vise à estimer seulement les informations extrêmes, qui sont importantes pour l'analyse de fatigue. Elle deviendrait potentiellement une méthode d'interpolation la plus rapide à réaliser (comparée aux travaux développés dans ce projet). Cependant, la performance de l'interpolation dépend de la qualité des seuils extrêmes choisis (ou estimés). Pär Johannesson (2006) et Poirier (2013) ont proposé des techniques graphiques et statistiques pour estimer ces seuils.
 - La chute est aussi un facteur important pour identifier les régimes permanents de la turbine. Donc, l'interpolation entre de différentes chutes (ensuite la combinaison de différentes chutes et ouvertures de directrices) est une recherche potentielle au futur pour augmenter maximale la flexibilité de l'outil d'interpolation inter-régimes.
 - Trois limitations de mesures expérimentales sont résolues indépendamment par des recherches différentes. Par conséquent, un grand projet, qui fusionne toutes les méthodologies comme l'extrapolation temporelle, l'interpolation spatiale sur l'aube et l'interpolation inter-régimes, est potentiel pour régler complètement les limitations expérimentales. Ce pourra considérablement contribuer à planifier une campagne de mesure.

ANNEXE I

THE CROSS-VALIDATION PERFORMANCE FOR SEVERAL NEIGHBOURHOOD SETS

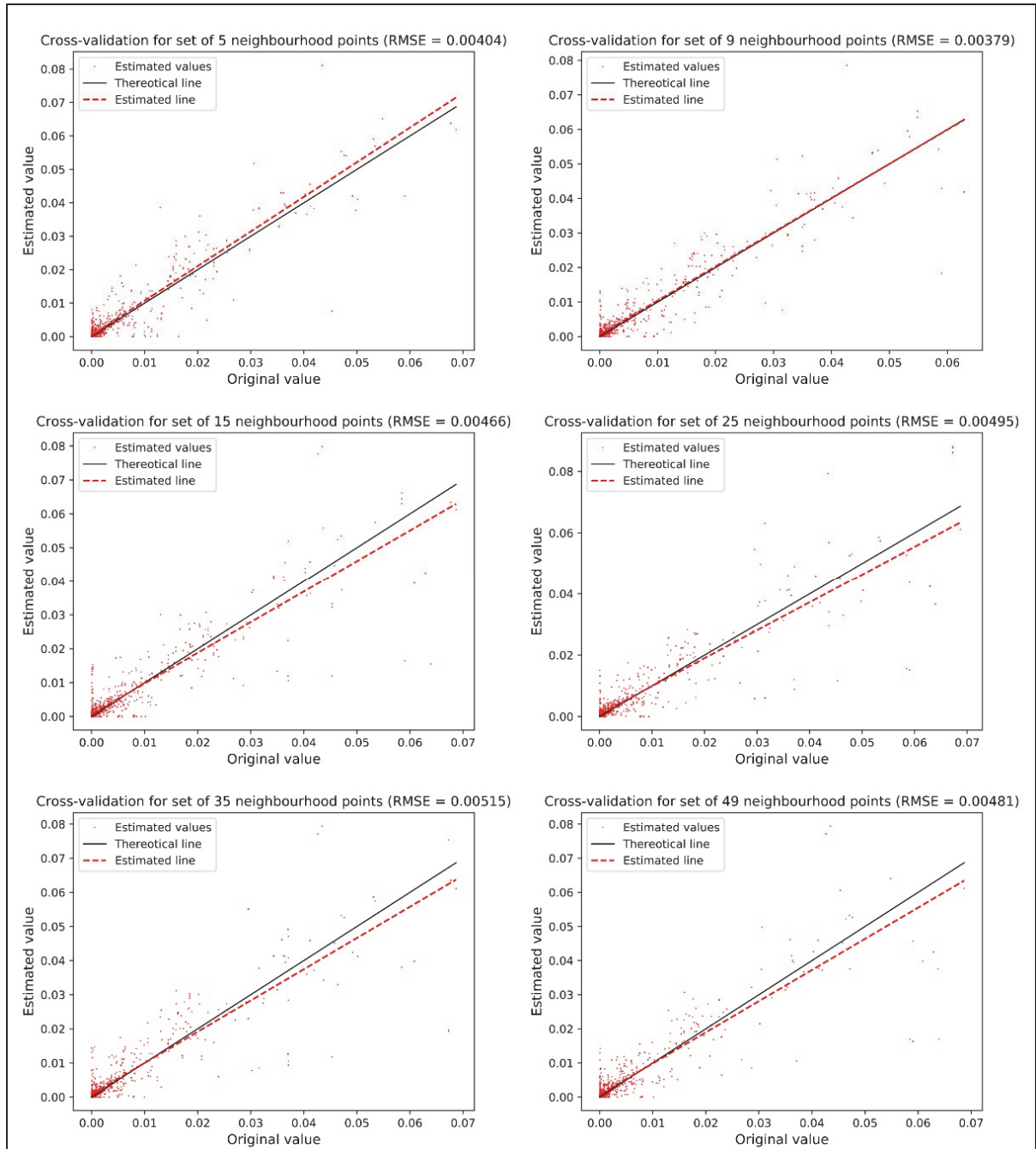


Figure-A I-1 Cross-validation performances for several neighbourhood sets

The cross-validation technique was performed for several neighbourhood sets (with the same semivariogram model). The optimized neighbourhood set for the interpolation is chosen based on the smallest RSME and based on the similarity between theoretical (black) and estimated line (red) in the cross-validation. In observing Figure-A I-1, if one chooses a small set (the case of 5 points) or a big set (the cases of 35 or 49 points), the variation and the RMSE values are larger. In this case study, the set of 9 neighbourhood points is chosen (row 1 and column 2 in Figure-A I-1).

LISTE DE RÉFÉRENCES BIBLIOGRAPHIQUES

- Antoni, J. (2005). Blind separation of vibration components: Principles and demonstrations. *Mechanical Systems and Signal Processing*, 19(6), 1166-1180.
- Antoni, J. (2009). Cyclostationarity by examples. *Mechanical Systems and Signal Processing*, 23(4), 987-1036.
- Antoni, J., Bonnardot, F., Raad, A., & El Badaoui, M. (2004). Cyclostationary modelling of rotating machine vibration signals. *Mechanical Systems and Signal Processing*, 18(6), 1285-1314.
- Antoni, J., & Randall, R. (2002). Differential diagnosis of gear and bearing faults. *J. Vib. Acoust.*, 124(2), 165-171.
- Antunes, I., & Albuquerque, M. (2013). Using indicator kriging for the evaluation of arsenic potential contamination in an abandoned mining area (Portugal). *Science of the Total Environment*, 442, 545-552.
- Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics surveys*, 4, 40-79.
- Aryaputera, A. W., Yang, D., Zhao, L., & Walsh, W. M. (2015). Very short-term irradiance forecasting at unobserved locations using spatio-temporal kriging. *Solar Energy*, 122, 1266-1278.
- Aschenbrenner, T., Otto, A., & Moser, W. (2006). Classification of vortex and cavitation phenomena and assessment of CFD prediction capabilities. *23rd IAHR Symposium, Yokohama, Japan*.
- Beirlant, J., Goegebeur, Y., Segers, J., & Teugels, J. L. (2006). *Statistics of extremes: theory and applications*. John Wiley & Sons.
- Bélisle, E., Huang, Z., Le Digabel, S., & Gheribi, A. E. (2015). Evaluation of machine learning interpolation techniques for prediction of physical properties. *Computational Materials Science*, 98, 170-177.
- Ben Salah, F. (2014). *Modélisation de la propagation des incertitudes des mesures sur l'aube d'une turbine hydraulique par Krigeage et simulations stochastiques* (École de technologie supérieure).

- Biernacki, C., Celeux, G., & Govaert, G. (2003). Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate Gaussian mixture models. *Computational Statistics & Data Analysis*, 41(3-4), 561-575.
- Bishop, N., & Sherratt, F. (1989). Fatigue life prediction from power spectral density data. *Environmental Engineering*, 2(1), 2-11.
- Bocchini, P., & Deodatis, G. (2008). Critical review and latest developments of a class of simulation algorithms for strongly non-Gaussian random fields. *Probabilistic Engineering Mechanics*, 23(4), 393-407.
- Boissonnat, J.-D., & Cazals, F. (2002). Smooth surface reconstruction via natural neighbour interpolation of distance functions. *Computational Geometry*, 22(1-3), 185-203.
- Bonnardot, F. (2004). *Comparaison entre les analyses angulaire et temporelle des signaux vibratoires de machines tournantes. Etude du concept de cyclostationnarité floue* (Institut National Polytechnique de Grenoble-INPG).
- Bonnardot, F., El Badaoui, M., Randall, R., Daniere, J., & Guillet, F. (2005). Use of the acceleration signal of a gearbox in order to perform angular resampling (with limited speed fluctuation). *Mechanical Systems and Signal Processing*, 19(4), 766-785.
- Bonnardot, F., Randall, R. B., & Guillet, F. (2005). Extraction of second-order cyclostationary sources—Application to vibration analysis. *Mechanical Systems and Signal Processing*, 19(6), 1230-1244. doi: 10.1016/j.ymssp.2005.07.008
- Botev, Z. I., Grotowski, J. F., & Kroese, D. P. (2010). Kernel density estimation via diffusion. *The annals of Statistics*, 38(5), 2916-2957.
- Boustany, R., & Antoni, J. (2005). A subspace method for the blind extraction of a cyclostationary source: Application to rolling element bearing diagnostics. *Mechanical Systems and Signal Processing*, 19(6), 1245-1259. doi: 10.1016/j.ymssp.2005.08.016
- Boustany, R., & Antoni, J. (2008). Blind extraction of a cyclostationary signal using reduced-rank cyclic regression—a unifying approach. *Mechanical Systems and Signal Processing*, 22(3), 520-541.
- Breeze, P. (2018). *Hydropower*. Academic Press.
- Brezovec, M., Kuzle, I., & Tomisa, T. (2006). Nonlinear digital simulation model of hydroelectric power unit with Kaplan turbine. *IEEE Transactions on Energy Conversion*, 21(1), 235-241.

- Browne, M. W. (2000). Cross-validation methods. *Journal of mathematical psychology*, 44(1), 108-132.
- Cai, G., & Lin, Y. (1996). Generation of non-Gaussian stationary stochastic processes. *Physical Review E*, 54(1), 299.
- Cao, Y., Xiang, H., & Zhou, Y. (2000). Simulation of stochastic wind velocity field on long-span bridges. *Journal of Engineering Mechanics*, 126(1), 1-6.
- Cetin, A., Naess, A., & Härkegård, G. (2013). A physically based extreme value characterization of material fatigue. *International journal of fatigue*, 47, 216-221.
- Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature. *Geoscientific model development*, 7(3), 1247-1250.
- Chatenet, Q., Tahan, A., Gagnon, M., & Chamberland-Lauzon, J. (2016). Numerical model validation using experimental data: Application of the area metric on a Francis runner. *IOP Conference Series: Earth and Environmental Science*, 49(6), 062015.
- Chiles, J.-P., & Delfiner, P. (2009). *Geostatistics: modeling spatial uncertainty* (Vol. 497). John Wiley & Sons.
- Ciocan, G. D., & Iliescu, M. S. (2007). Vortex rope investigation by 3D-PIV method. *Proc. 2nd Int. Meeting of Workgroup on Cavitation and Dynamic Problems in Hydraulic Machinery and Systems, Timisoara, Romania*, 159-172.
- Coles, S., Bawa, J., Trenner, L., & Dorazio, P. (2001). *An introduction to statistical modeling of extreme values* (Vol. 208). Springer.
- Combet, F., & Zimroz, R. (2009). A new method for the estimation of the instantaneous speed relative fluctuation in a vibration signal based on the short time scale transform. *Mechanical Systems and Signal Processing*, 23(4), 1382-1397.
- Cressie, N. (1990). The origins of kriging. *Mathematical Geology*, 22(3), 239-252.
- Deodatis, G., & Micaletti, R. C. (2001). Simulation of highly skewed non-Gaussian stochastic processes. *Journal of Engineering Mechanics*, 127(12), 1284-1295.
- Deutsch, C. V., & Journel, A. G. (1992). Geostatistical software library and user's guide. *New York*, 119(147).
- Diagne, I., Gagnon, M., & Tahan, A. (2016). Modeling the dynamic behavior of turbine runner blades during transients using indirect measurements. *IOP Conference Series: Earth and Environmental Science*, 49(7), 072014.

- Dollon, Q., Antoni, J., Tahan, A., Gagnon, M., & Monette, C. (2021). Operational Modal Analysis of hydroelectric turbines using an order based likelihood approach. *Renewable Energy*, 165, 799-811.
- Dollon, Q., Tahan, A. S., Antoni, J., Gagnon, M., & Monette, C. (2019). Dynamic Characterization Of Hydroelectric Turbine In Transient Using OBMA And Phase-Shift Analysis. *Surveillance, Vishno and AVE conferences*.
- Dörfler, P., Sick, M., & Coutu, A. (2012). *Flow-induced pulsation and vibration in hydroelectric machinery: engineer's guidebook for planning, design and troubleshooting*. Springer Science & Business Media.
- Dreßler, K., Hack, M., & Krüger, W. (1997). Stochastic reconstruction of loading histories from a rainflow matrix. *ZAMM-Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik*, 77(3), 217-226.
- Dubrule, O. (1983). Two methods with different objectives: splines and kriging. *Journal of the International Association for Mathematical Geology*, 15(2), 245-257.
- Duong, T., & Hazelton, M. L. (2005). Cross-validation bandwidth matrices for multivariate kernel density estimation. *Scandinavian Journal of Statistics*, 32(3), 485-506.
- Eckert-Gallup, A., & Martin, N. (2016). Kernel density estimation (KDE) with adaptive bandwidth selection for environmental contours of extreme sea states. Dans *OCEANS 2016 MTS/IEEE Monterey* (pp. 1-5). IEEE.
- El Haddad, M., Topper, T., & Smith, K. (1979). Prediction of non propagating cracks. *Engineering fracture mechanics*, 11(3), 573-584.
- Erdogan, K. (2013). Spline interpolation techniques. *Journal of Technical Science and Technologies*, 47-52.
- Fanelli, M. (1989). The vortex rope in the draft tube of Francis turbines operating at partial load: a proposal for a mathematical model. *Journal of hydraulic research*, 27(6), 769-807.
- Favrel, A., Gomes Pereira Junior, J., Landry, C., Müller, A., Nicolet, C., & Avellan, F. (2018). New insight in Francis turbine cavitation vortex rope: role of the runner outlet flow swirl number. *Journal of hydraulic research*, 56(3), 367-379.
- Ferreira, A., & de Haan, L. (2013). On the block maxima method in extreme value theory. *arXiv preprint arXiv:1310.3222*.

- Ferson, S., Oberkampf, W. L., & Ginzburg, L. (2008). Model validation and predictive capability for the thermal challenge problem. *Computer Methods in Applied Mechanics and Engineering*, 197(29-32), 2408-2430.
- Foulley, J.-L. (2002). Algorithme EM: théorie et application au modèle mixte. *Journal de la Société française de statistique*, 143(3-4), 57-109.
- Fu, R., Wang, H., & Zhao, W. (2016). Dynamic driver fatigue detection using hidden Markov model in real driving condition. *Expert Systems with Applications*, 63, 397-411.
- Fyfe, K., & Munck, E. (1997). Analysis of computed order tracking. *Mechanical Systems and Signal Processing*, 11(2), 187-205.
- Gagnon, M., & Léonard, F. (2013). Transient response and life assessment: Case studies on the load rejection of two hydroelectric turbines. *Proceedings of the International Conference Surveillance, Chartres, France*, 1-11.
- Gagnon, M., Tahan, A., Bocher, P., & Thibault, D. (2012). The role of high cycle fatigue (HCF) onset in Francis runner reliability. *IOP Conference Series: Earth and Environmental Science*, 15(2), 022005.
- Gagnon, M., Tahan, A., Bocher, P., & Thibault, D. (2013). A probabilistic model for the onset of High Cycle Fatigue (HCF) crack propagation: Application to hydroelectric turbine runner. *International journal of fatigue*, 47, 300-307.
- Gagnon, M., Tahan, A., Bocher, P., & Thibault, D. (2014). Influence of load spectrum assumptions on the expected reliability of hydroelectric turbines: A case study. *Structural Safety*, 50, 1-8.
- Gardner, W. A., Napolitano, A., & Paura, L. (2006). Cyclostationarity: Half a century of research. *Signal processing*, 86(4), 639-697.
- Goovaerts, P. (1997). *Geostatistics for natural resources evaluation*. Oxford University Press on Demand.
- Goovaerts, P. (1998). Ordinary cokriging revisited. *Mathematical Geology*, 30(1), 21-42.
- Goulard, M., & Voltz, M. (1992). Linear coregionalization model: tools for estimation and choice of cross-variogram matrix. *Mathematical Geology*, 24(3), 269-286.
- Griffin, D., & Lim, J. (1984). Signal estimation from modified short-time Fourier transform. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 32(2), 236-243.

- Grigoriu, M. (1995). Applied non-gaussian processes: Examples, theory, simulation, linear random vibration, and MATLAB solutions(Book). *Englewood Cliffs, NJ: Prentice Hall, Inc, 1995*.
- Grigoriu, M. (1998). Simulation of stationary non-Gaussian translation processes. *Journal of Engineering Mechanics*, 124(2), 121-126.
- Hastie, T. J., & Tibshirani, R. J. (1990). *Generalized additive models* (Vol. 43). CRC press.
- Haylock, M., Hofstra, N., Klein Tank, A., Klok, E., Jones, P., & New, M. (2008). A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. *Journal of Geophysical Research: Atmospheres*, 113(D20).
- Heidenreich, N.-B., Schindler, A., & Sperlich, S. (2013). Bandwidth selection for kernel density estimation: a review of fully automatic selectors. *AStA Advances in Statistical Analysis*, 97(4), 403-433.
- Heinzel, G., Rüdiger, A., & Schilling, R. (2002). Spectrum and spectral density estimation by the Discrete Fourier transform (DFT), including a comprehensive list of window functions and some new at-top windows.
- Hu, D., Shu, H., Hu, H., & Xu, J. (2017). Spatiotemporal regression Kriging to predict precipitation using time-series MODIS data. *Cluster Computing*, 20(1), 347-357.
- Hudson, G., & Wackernagel, H. (1994). Mapping temperature using kriging with external drift: theory and an example from Scotland. *International journal of Climatology*, 14(1), 77-91.
- Hughes, A. W., King, M. L., & Kwek, K. T. (2004). Selecting the order of an ARCH model. *Economics Letters*, 83(2), 269-275.
- Hutchinson, M. F. (1995). Interpolating mean rainfall using thin plate smoothing splines. *International journal of geographical information systems*, 9(4), 385-403.
- Johannesson, P. (1999). Rainflow analysis of switching Markov loads. *Lund Institute of Technology, Centre for Mathematical Sciences, Mathematical Statistics, Lund Institute of Technology, Sweden*, 1-27.
- Johannesson, P. (2006). Extrapolation of load histories and spectra. *Fatigue & Fracture of Engineering Materials & Structures*, 29(3), 209-217.
- Johannesson, P., & Thomas, J.-J. (2001). Extrapolation of rainflow matrices. *Extremes*, 4(3), 241-262.

- Jost, G., Heuvelink, G., & Papritz, A. (2005). Analysing the space–time distribution of soil water storage of a forest ecosystem using spatio-temporal kriging. *Geoderma*, 128(3-4), 258-273.
- Journal, A. G. (1983). Nonparametric estimation of spatial distributions. *Journal of the International Association for Mathematical Geology*, 15(3), 445-468.
- Journal, A. G., & Huijbregts, C. J. (1976). Mining geostatistics.
- Jung, B. C., Park, J., Oh, H., Kim, J., & Youn, B. D. (2015). A framework of model validation and virtual product qualification with limited experimental data based on statistical inference. *Structural and multidisciplinary optimization*, 51(3), 573-583.
- Kassim, A. H. M., & Kottegoda, N. (1991). Rainfall network design through comparative kriging methods. *Hydrological Sciences Journal*, 36(3), 223-240.
- Kitagawa, H. (1976). Applicability of fracture mechanics to very small cracks or the cracks in the early stage. *Proc. of 2nd ICM, Cleveland, 1976*, 627-631.
- Krige, D. G. (1951). A statistical approach to some basic mine valuation problems on the Witwatersrand. *Journal of the Southern African Institute of Mining and Metallurgy*, 52(6), 119-139.
- Kuibin, P., Okulov, V., Susan-Resiga, R., & Muntean, S. (2010). Validation of mathematical models for predicting the swirling flow and the vortex rope in a Francis turbine operated at partial discharge. *IOP Conference Series: Earth and Environmental Science*, 12(1), 012051.
- Lark, R. (2002). Robust estimation of the pseudo cross-variogram for cokriging soil properties. *European Journal of Soil Science*, 53(2), 253-270.
- Lark, R., & Papritz, A. (2003). Fitting a linear model of coregionalization for soil properties using simulated annealing. *Geoderma*, 115(3-4), 245-260.
- Larsen, C. E., & Irvine, T. (2015). A review of spectral methods for variable amplitude fatigue prediction and new results. *Procedia Engineering*, 101, 243-250.
- Layer, E., & Tomczyk, K. (2015). *Signal transforms in dynamic measurements*. Springer.
- Lekkas, A. M., & Fossen, T. I. (2014). Integral LOS path following for curved paths based on a monotone cubic Hermite spline parametrization. *IEEE Transactions on Control Systems Technology*, 22(6), 2287-2301.
- Léonard, F. (2015). Noisy data clusters are hollow. *arXiv preprint arXiv:1506.03318*.

- Li, H., & Zhang, D. (2013). Stochastic representation and dimension reduction for non-Gaussian random fields: review and reflection. *Stochastic environmental research and risk assessment*, 27(7), 1621-1635.
- Li, J., & Heap, A. D. (2014). Spatial interpolation methods applied in the environmental sciences: A review. *Environmental Modelling & Software*, 53, 173-189.
- Li, J., Heap, A. D., Potter, A., & Daniell, J. J. (2011). Application of machine learning methods to spatial interpolation of environmental variables. *Environmental Modelling & Software*, 26(12), 1647-1659.
- Li, Z., Wang, K., Ma, H., & Wu, Y. (2018). An adjusted inverse distance weighted spatial interpolation method. Dans *2018 3rd International Conference on Communications, Information Management and Network Security (CIMNS 2018)*. Atlantis Press.
- Liu, C.-W., Jang, C.-S., & Liao, C.-M. (2004). Evaluation of arsenic contamination potential using indicator kriging in the Yun-Lin aquifer (Taiwan). *Science of the Total Environment*, 321(1-3), 173-188.
- Liu, X., Luo, Y., & Wang, Z. (2016). A review on fatigue damage mechanism in hydro turbines. *Renewable and Sustainable Energy Reviews*, 54, 1-14.
- Liu, Y., Chen, W., Arendt, P., & Huang, H.-Z. (2011). Toward a Better Understanding of Model Validation Metrics. *Journal of Mechanical Design*, 133(7). doi: 10.1115/1.4004223
- Lu, G. Y., & Wong, D. W. (2008). An adaptive inverse-distance weighting spatial interpolation technique. *Computers & geosciences*, 34(9), 1044-1055.
- Lucifredi, A., Mazzieri, C., & Rossi, M. (2000). Application of multiregressive linear models, dynamic kriging models and neural network models to predictive maintenance of hydroelectric power systems. *Mechanical Systems and Signal Processing*, 14(3), 471-494.
- Marinoni, O. (2003). Improving geological models using a combined ordinary–indicator kriging approach. *Engineering geology*, 69(1-2), 37-45.
- Marsh, G., Wignall, C., Thies, P. R., Barltrop, N., Incecik, A., Venugopal, V., & Johanning, L. (2016). Review and application of Rainflow residue processing techniques for accurate fatigue damage estimation. *International journal of fatigue*, 82, 757-765.
- Masters, F., & Gurley, K. R. (2003). Non-Gaussian simulation: cumulative distribution function map-based spectral correction. *Journal of Engineering Mechanics*, 129(12), 1418-1428.

- Matsuishi, M., & Endo, T. (1968). Fatigue of metals subjected to varying stress. *Japan Society of Mechanical Engineers, Fukuoka, Japan*, 68(2), 37-40.
- McFadden, P. (1989). Interpolation techniques for time domain averaging of gear vibration. *Mechanical Systems and Signal Processing*, 3(1), 87-97.
- Mende, C., Weber, W., & Seidel, U. (2016). Progress in load prediction for speed-no-load operation in Francis turbines. *IOP Conference Series: Earth and Environmental Science*, 49(6), 062017.
- Monette, C., Marmont, H., Chamberland-Lauzon, J., Skagerstrand, A., Coutu, A., & Carlevi, J. (2016). Cost of enlarged operating zone for an existing Francis runner. *IOP Conference Series: Earth and Environmental Science*, 49(7), 072018.
- Montero, J.-M., Fernández-Avilés, G., & Mateu, J. (2015). *Spatial and spatio-temporal geostatistical modeling and kriging* (Vol. 998). John Wiley & Sons.
- Moreau, V., Bage, G., Marcotte, D., & Samson, R. (2012). Statistical estimation of missing data in life cycle inventory: an application to hydroelectric power plants. *Journal of Cleaner Production*, 37, 335-341.
- Morissette, J., Chamberland-Lauzon, J., Nennemann, B., Monette, C., Giroux, A., Coutu, A., & Nicolle, J. (2016). Stress predictions in a Francis turbine at no-load operating regime. *IOP Conference Series: Earth and Environmental Science*, 49(7), 072016.
- Mosier, C. I. (1951). I. Problems and designs of cross-validation 1. *Educational and Psychological Measurement*, 11(1), 5-11.
- Musallam, M., & Johnson, C. M. (2012). An efficient implementation of the rainflow counting algorithm for life consumption estimation. *IEEE Transactions on reliability*, 61(4), 978-986.
- Myers, D. (1993). Cross-validation and variogram estimation. *Theory of Probability & Its Applications*, 37(2), 345-347.
- Myers, D. E. (1982). Matrix formulation of co-kriging. *Journal of the International Association for Mathematical Geology*, 14(3), 249-257.
- Myers, D. E. (1991). Pseudo-cross variograms, positive-definiteness, and cokriging. *Mathematical Geology*, 23(6), 805-816.
- Narendra, K., & Gallman, P. (1966). An iterative method for the identification of nonlinear systems using a Hammerstein model. *IEEE Transactions on Automatic control*, 11(3), 546-550.

- Nawab, S., Quatieri, T., & Lim, J. (1983). Signal reconstruction from short-time Fourier transform magnitude. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 31(4), 986-998.
- Nennemann, B., Morissette, J., Chamberland-Lauzon, J., Monette, C., Braun, O., Melot, M., . . . Giroux, A. (2014). Challenges in dynamic pressure and stress predictions at no-load operation in hydraulic turbines. *IOP conference series: earth and environmental science*, 22(3), 032055.
- Nowak, M., & Verly, G. (2005). The practice of sequential Gaussian simulation. Dans *Geostatistics Banff 2004* (pp. 387-398). Springer.
- Oliver, M. A., & Webster, R. (1990). Kriging: a method of interpolation for geographical information systems. *International Journal of Geographical Information System*, 4(3), 313-332.
- Oppenheim, A. V., & Schaffer, R. W. (2004). From frequency to quefrency: A history of the cepstrum. *IEEE signal processing Magazine*, 21(5), 95-106.
- Papritz, A., Künsch, H., & Webster, R. (1993). On the pseudo cross-variogram. *Mathematical Geology*, 25(8), 1015-1026.
- Pebesma, E., & Heuvelink, G. (2016). Spatio-temporal interpolation using gstat. *RFID Journal*, 8(1), 204-218.
- Peel, D., & McLachlan, G. J. (2000). Robust mixture modelling using the t distribution. *Statistics and computing*, 10(4), 339-348.
- Pennacchi, P., Borghesani, P., & Chatterton, S. (2015). A cyclostationary multi-domain analysis of fluid instability in Kaplan turbines. *Mechanical Systems and Signal Processing*, 60-61, 375-390. doi: 10.1016/j.ymssp.2014.08.026
- Petrucci, G., & Zuccarello, B. (1999). On the estimation of the fatigue cycle distribution from spectral density data. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 213(8), 819-831.
- Pham, Q. H., Gagnon, M., Antoni, J., Tahan, A., & Monette, C. (2021). Rainflow-counting matrix interpolation over different operating conditions for hydroelectric turbine fatigue assessment. *Renewable Energy*.
- Pham, Q. H., Gagnon, M., Antoni, J., Tahan, A. S., & Monette, C. (2019). Interpolation of periodic hidden signal measured at steady-operating conditions on hydroelectric turbine runners. *Surveillance, Vishno and AVE conferences*.

- Poirier, M. (2013). *Modélisation et simulation du comportement dynamique des aubes de turbines hydroélectriques* (École de technologie supérieure).
- Poirier, M., Gagnon, M., Tahan, A., Coutu, A., & Chamberland-lauzon, J. (2017). Extrapolation of dynamic load behaviour on hydroelectric turbine blades with cyclostationary modelling. *Mechanical Systems and Signal Processing*, 82, 193-205.
- Presas, A., Luo, Y., Wang, Z., & Guo, B. (2019). Fatigue life estimation of Francis turbines based on experimental strain measurements: Review of the actual data and future trends. *Renewable and Sustainable Energy Reviews*, 102, 96-110.
- Presas, A., Valentin, D., Zhao, W., Egusquiza, M., Valero, C., & Egusquiza, E. (2021). On the use of neural networks for dynamic stress prediction in Francis turbines by means of stationary sensors. *Renewable Energy*, 170, 652-660.
- Reynolds, D. A. (2009). Gaussian Mixture Models. *Encyclopedia of biometrics*, 741, 659-663.
- Rychlik, I. (1987). A new definition of the rainflow cycle counting method. *International journal of fatigue*, 9(2), 119-121.
- Sakia, R. M. (1992). The Box-Cox transformation technique: a review. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 41(2), 169-178.
- Schafer, R., & Rabiner, L. (1973). Design and simulation of a speech analysis-synthesis system based on short-time Fourier analysis. *IEEE Transactions on Audio and Electroacoustics*, 21(3), 165-174.
- Schetzen, M. (1980). The Volterra and Wiener theories of nonlinear systems.
- Seidel, U., Mende, C., Hübner, B., Weber, W., & Otto, A. (2014). Dynamic loads in Francis runners and their impact on fatigue life. *IOP conference series: earth and environmental science*, 22(3), 032054.
- Serpedin, E., Panduru, F., Sari, I., & Giannakis, G. B. (2005). Bibliography on cyclostationarity. *Signal processing*, 85(12), 2233-2303.
- Shcherbakov, M. V., Brebels, A., Shcherbakova, N. L., Tyukov, A. P., Janovsky, T. A., & Kamaev, V. A. e. (2013). A survey of forecast error measures. *World Applied Sciences Journal*, 24(24), 171-176.
- Sheather, S. J., & Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society: Series B (Methodological)*, 53(3), 683-690.

- Sherman, M. (2011). *Spatial statistics and spatio-temporal data: covariance functions and directional properties*. John Wiley & Sons.
- Shinozuka, M., & Deodatis, G. (1991). Simulation of stochastic processes by spectral representation.
- Shinozuka, M., & Jan, C.-M. (1972). Digital simulation of random processes and its applications. *Journal of sound and vibration*, 25(1), 111-128.
- Sibson, R. (1981). A brief description of natural neighbour interpolation. *Interpreting multivariate data*.
- Skjelbred, H. I., & Kong, J. (2019). A comparison of linear interpolation and spline interpolation for turbine efficiency curves in short-term hydropower scheduling problems. Dans *IOP Conference Series: Earth and Environmental Science* (Vol. 240, pp. 042011). IOP Publishing.
- Smith, J. L., Halvorson, J. J., & Papendick, R. I. (1993). Using multiple-variable indicator kriging for evaluating soil quality. *Soil Science Society of America Journal*, 57(3), 743-749.
- Snepvangers, J., Heuvelink, G., & Huisman, J. (2003). Soil water content interpolation using spatio-temporal kriging with external drift. *Geoderma*, 112(3-4), 253-271.
- Socie, D. (2001). Modelling expected service usage from short-term loading measurements. *International Journal of Materials and Product Technology*, 16(4-5), 295-303.
- Socie, D. F., & Pompetzki, M. A. (2004). Modeling variability in service loading spectra. *Journal of ASTM International*, 1(2), 1-12.
- Späth, H. (1995). *One dimensional spline interpolation algorithms*. AK Peters/CRC Press.
- Sturm, N., & Daudet, L. (2011). Signal reconstruction from STFT magnitude: A state of the art. *International conference on digital audio effects (DAFx)*, 375-386.
- Suro-Perez, V., & Journel, A. (1991). Indicator principal component kriging. *Mathematical Geology*, 23(5), 759-788.
- Thibault, D., Gagnon, M., & Godin, S. (2014). Bridging the gap between metallurgy and fatigue reliability of hydraulic turbine runners. *IOP Conference Series: Earth and Environmental Science*, 22(1), 012019.
- Trivedi, C., & Cervantes, M. J. (2017). Fluid-structure interactions in Francis turbines: A perspective review. *Renewable and Sustainable Energy Reviews*, 68, 87-101.

- Trudel, A., & Sabourin, M. (2014). Metallurgical and fatigue assessments of welds in cast welded hydraulic turbine runners. *IOP Conference Series: Earth and Environmental Science*, 22(1), 012015.
- Vaidyanathan, P. P. (2007). The theory of linear prediction. *Synthesis lectures on signal processing*, 2(1), 1-184.
- Vann, J., Jackson, S., & Bertoli, O. (2003). Quantitative kriging neighbourhood analysis for the mining geologist-a description of the method with worked case examples. Dans *5th international mining geology conference* (Vol. 1, pp. 215-223). Bendigo, Australia. Melbourne: Australian Inst Mining & Metallurgy.
- Voyles, I. T., & Roy, C. J. (2014). Evaluation of model validation techniques in the presence of uncertainty. Dans *16th AIAA Non-Deterministic Approaches Conference* (pp. 0120).
- Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, K. (1993). *Probability and statistics for engineers and scientists* (Vol. 5). Macmillan New York.
- Wand, M. P., & Jones, M. C. (1994). *Kernel smoothing*. CRC press.
- Wang, N., Yao, W., Zhao, Y., Chen, X., Zhang, X., & Li, L. (2018). A new interval area metric for model validation with limited experimental data. *Journal of Mechanical Design*, 140(6).
- Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate research*, 30(1), 79-82.
- Xiao, R., Wang, Z., & Luo, Y. (2008). Dynamic stresses in a Francis turbine runner based on fluid-structure interaction analysis. *Tsinghua Science and Technology*, 13(5), 587-592.
- Yamazaki, F., & Shinozuka, M. (1988). Digital generation of non-Gaussian stochastic fields. *Journal of Engineering Mechanics*, 114(7), 1183-1197.
- Zhu, X., Zhang, Z., Gao, J., & Li, W. (2019). Two robust approaches to multicomponent signal reconstruction from STFT ridges. *Mechanical Systems and Signal Processing*, 115, 720-735.
- Zougab, N., Adjabi, S., & Kokonendji, C. C. (2014). Bayesian estimation of adaptive bandwidth matrices in multivariate kernel density estimation. *Computational Statistics & Data Analysis*, 75, 28-38.