



HAL
open science

Rainflow-counting matrix interpolation over different operating conditions for hydroelectric turbine fatigue assessment

Quang Hung Pham, Martin Gagnon, Jérôme Antoni, Antoine Tahan,
Christine Monette

► **To cite this version:**

Quang Hung Pham, Martin Gagnon, Jérôme Antoni, Antoine Tahan, Christine Monette. Rainflow-counting matrix interpolation over different operating conditions for hydroelectric turbine fatigue assessment. *Renewable Energy*, 2021, 172, pp.465-476. 10.1016/j.renene.2021.03.036 . hal-03211931

HAL Id: hal-03211931

<https://hal.science/hal-03211931v1>

Submitted on 24 Apr 2023

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

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



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

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), H3C 1K3, Montreal, QC, Canada

^b University of Lyon, INSA Lyon, Laboratoire Vibration Acoustique, F69621, Villeurbanne, France

^c Institut de recherche d'Hydro-Québec (IREQ), J3X1S1 Varennes, QC, Canada

^d Andritz Hydro Canada Inc, H9R 1B9, Point-Claire, QC, Canada

Contact : quang-hung.pham.1@ens.etsmtl.ca or quang-hung.pham@insa-lyon.fr

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.

Keywords: Hydroelectric turbine; Kriging interpolation; Runner fatigue assessment; Rainflow counting

Nomenclature

B	Bin value of ARM histogram [probability]	ARM	Asymmetric Rainflow Matrix
f	From value of a strain cycle [μ strain]	BEP	Best Efficiency Point
t	To value of a strain cycle [μ strain]	BGD	Bivariate Gaussian Distribution
o	Operating condition [%OV]	CFD	Computational Fluid Dynamic
p	Bin position ($p(f, v)$)	DC	Decreasing Cycle
n_o	Number of experimental operating conditions	DPL	Deep Part-Load
ϑ_i	Set of bin position p_i and its neighbours	ES	Experimental Semivariogram
λ_{kl}	Regression weight	FEA	Finite Element Analysis
σ_e^2	Error variance	FPC	Four-Points Counting
γ	Semivariogram value	HCF	High Cycle Fatigue
$\hat{\gamma}$	Experimental Semivariogram	IC	Increasing Cycle
h	Histogram Space Increment	MAE	Mean Absolute Error
u	Operating Conditions Increment	RMSE	Root Mean Square Error
$v(h, u)$	Set of paired points	SNL	Speed No-Load
$N(h, u)$	Cardinal of $v(h, u)$	STK	Spatio-Temporal Kriging
$L(\cdot)$	Lagrange function	%OV	Guide Vane Opening Level
μ	Lagrange multiplier	κ	Anisotropy correction
$\gamma_{Smetric}$	Sum-metric semivariogram model	γ_o	Semivariogram model for variable o
γ_{joint}	Joint semivariogram	γ_p	Semivariogram model for variable p

1. 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 [1]. These changes mean that the turbine must

34 operate under conditions that deviate significantly from the Best Efficiency Point (BEP), which leads
35 to a higher risk of fatigue crack. Monette et al. (2016) highlighted the reduced fatigue life of turbine
36 runner and the high repair costs generated by this enlarged operating range [1]. A general overview of
37 conditions and dynamic phenomena that significantly influence runner fatigue is presented by Liu et
38 al. (2016), Seidel et al. (2014) and Gagnon et al. (2014) [2-4]. Steady state operating conditions
39 containing a high stochastic component such as Speed No-Load (SNL) or Deep Part-Load (DPL) or
40 Part load containing vortex rope phenomenon, and transient conditions such as turbine start-up can
41 have major impact on runner fatigue assessment if not properly accounted for.

42 Fatigue is assessed based on several factors (runner material, stress level, initial crack), and
43 requires a representative strain/stress history. Many approaches are available for improving runner
44 fatigue life estimation. To limit fatigue failure, Gagnon et al. (2013) proposed a runner reliability
45 model using the High Cycle Fatigue (HCF) onset as the limit [5]. Fatigue damage is linked to the
46 propagation of cracks on the runner, and thus, an investigation of the metallurgical or mechanical
47 behaviour of the material is also needed [6-9]. However, fatigue assessment cannot be carried out if
48 the dynamic load information is missing. Computational Fluid Dynamic (CFD) combined with Finite
49 Element Analysis (FEA) can be used to estimate the static and dynamic behaviour of the structure
50 under such as SNL or DPL [10-12]. Nevertheless, numerical prediction is still challenging when
51 turbulence models are not able to properly describe the behaviour of the turbine [13]. Moreover, the
52 validation of numerical simulation depends on the availability of the experimental information.
53 However, to recover this missing experimental information, other approaches, such as the
54 interpolation and extrapolation of experimental data, can be used [14-16]. The use of indirect method
55 can also help minimize the measurement requirement, Diagne et al. (2016) utilised the ARMAX
56 (Autoregressive Moving Average with eXogenous input) model to predict the runner dynamic stress
57 using shaft torsion information, which is more readily available [17]. For a detailed review of the
58 fatigue life estimation methodologies used for Francis turbines, please refer to Presas et al. (2019)
59 [13]. Our state-of-art summary shows that in situ field strains/stress during or after commissioning
60 provide important and essential information for the study of the hydroelectric turbine runner fatigue.

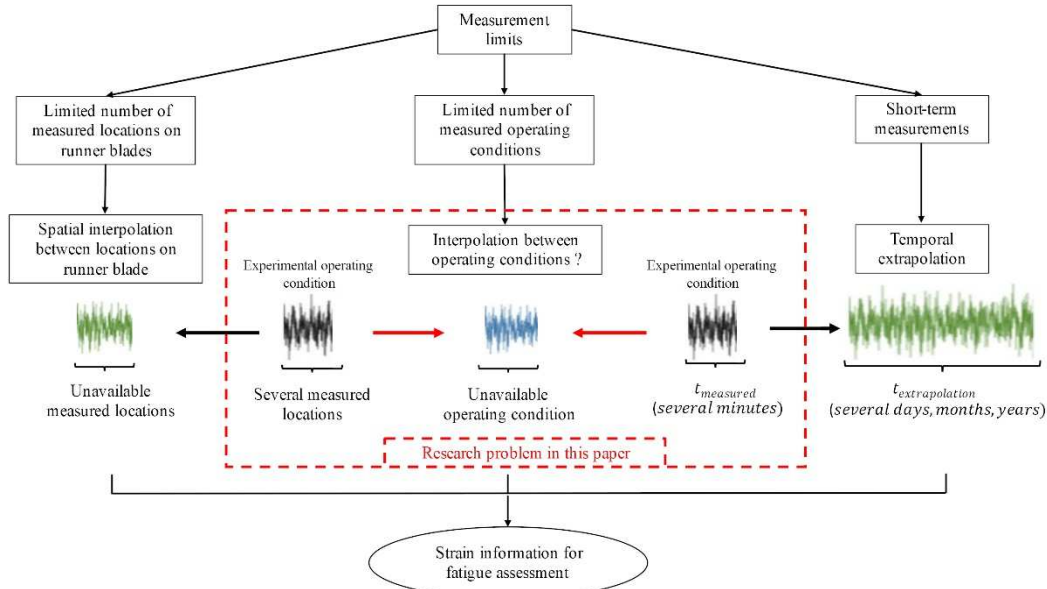
61 Finding a solution that can provide missing experimental information and improve knowledge on
62 the dynamic behaviour remains a challenge. Obtaining experimental strains over all the possible states
63 of the hydroelectric turbine operation would be overly costly and time-consuming; hence,
64 experimental measurements are usually limited to a limited set of operating conditions, which leads to
65 difficulties during runner fatigue assessment. The goal of this paper is to propose a methodology to
66 minimize the requirements for experimental strain measurements. There are three difficulties
67 associated with the measurement of strain for fatigue evaluation:

- 68 • Limited number of measured locations on runner blades
- 69 • Limited measurement length
- 70 • Limited number of measured operating conditions

71 The first was studied by Salah where the kriging method was used for the interpolation between
72 different locations of the runner blade [14]. For the second, several approaches can be used for
73 temporal extrapolation to obtain long-term fatigue assessments. Poirier et al. (2016) [15] extrapolated
74 information on the cyclostationary decompositions of the short-term records of strain signals to
75 evaluate runner fatigue cycles over a longer period of usage. Other researchers extrapolated the
76 rainflow matrices directly over longer periods [19-22]. However, such strain interpolation between
77 blade locations or a temporal extrapolation can be performed if and only if an estimate of the signal is
78 available (generally in the form of measurements) at a given turbine operating condition. Therefore,
79 the lack of measurements over the complete operating range should be studied as shown in *Figure 1*.

80 In a previous work the authors developed an interpolation process for a subcomponent of the strain
81 signal between different turbine operating conditions [16]. Only the periodic component of the signal
82 was interpolated while the complete strain signal was overlooked as it is more complex and contains
83 other hidden structures. The direct interpolation of complete strain signal (measurement form) is a
84 complicated problem because of these hidden structures. However, in cases where only fatigue
85 information is needed for further evaluation, the strain signal could be transformed into a cycle
86 counting form (level crossing, peak counting, rainflow matrices, etc.) to simplify the interpolation.
87 Among cycle counting methods, the rainflow algorithm is commonly used for fatigue assessment of

88 hydroelectric turbines [1, 10-11, 15-16]. Therefore, our goal is to interpolate the rainflow counting
 89 matrices (instead of the signal itself), which contain information on the fatigue cycles, over the
 90 complete range of operating conditions. Notice that there are two important variables that define the
 91 operating conditions of turbine: opening vane level and head level. However, in this study, the aim is
 92 to focus on the differences of stress level between different opening vane levels for the same head (the
 93 data in the case study also comes from the measurements strain at relatively constant head). The head
 94 level might be used in the future research if observed data become available for multiple head levels.
 95



96
 97 *Figure 1. Measurement limits and interpolation/extrapolation solutions for the fatigue assessment of hydroelectric turbine*
 98 *runner*

99 The kriging method was chosen for this study to interpolate rainflow counting matrices (histogram
 100 form), Then, the fatigue cycle reconstruction was generated via stochastic simulations using bivariate
 101 Gaussian distributions. This reconstruction step helps recover the long-term information of fatigue,
 102 more cycles are generated, more fatigue information is evaluated. The obtained results were compared
 103 with the experimental fatigue cycles to evaluate the correctness of the proposed interpolation
 104 methodology. In addition, the signal was reconstructed from the simulated cycles by using the inverse
 105 method proposed in [19-20] to compare with the measured signal.

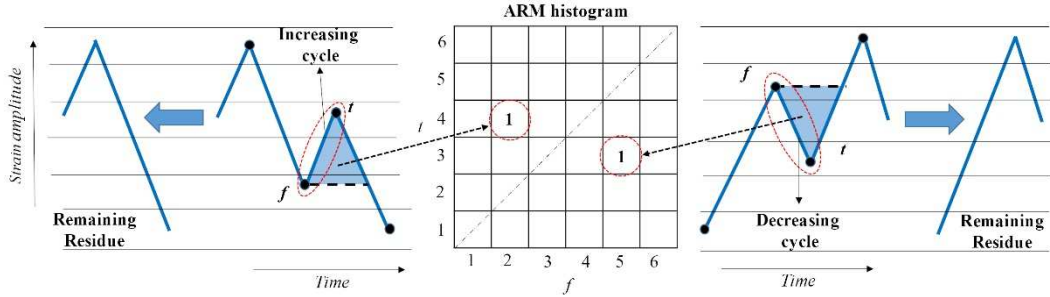
106 The paper is organized as follows: First, the methodology is presented with detailed explanations
 107 of the principal methods (rainflow-counting, kriging interpolation, reconstruction simulation). Then, a
 108 hydroelectric turbine runner case study is presented followed by interpolation results and a discussion.
 109

110 2. Proposed Methodology

111 2.1 Asymmetric Rainflow Counting Matrices

112 The rainflow algorithm allows the analysis of signal amplitudes (or range) by assigning extreme (max,
 113 min) values of closed loading cycles. To evaluate fatigue, only the range/amplitude information of
 114 each cycle is often needed. In this study, the strain signal is transformed into an Asymmetric
 115 Rainflow Matrix (ARM), which separates the cycles into two types: increasing and decreasing cycles
 116 (or standing and hanging cycles in some references) [24]. These two types of cycles are defined based
 117 on the order of occurrence of maximum and minimum values in the signal. To construct the ARM, a
 118 rainflow-counting procedure called the Four-Point Counting (FPC) algorithm is used [25].
 119 Considering four consecutive points (three consecutive cycles) in the loading signal, if the cycle range
 120 created by the second and the third points is smaller than the range of the two other points, this cycle
 121 is reduced and is added to the ARM. An ARM is constructed by two variables: from value (noted as f)
 122 corresponding to the amplitude of the first point of the cycle (the first point of the cycle arrives on the
 123 axis of time) and to value (noted as t) corresponding to the amplitude of the final point of the cycle
 124 (*Figure 2*). In our study, the created ARM is represented under histogram density form where each
 125 bin contains the number of cycles or the probability of occurrence of cycles. The ARM is a

126 composition of max-min cycles (above the matrix diagonal) and min-max cycles (below the matrix diagonal).
 127 Furthermore, the choice of the asymmetric rainflow matrix allows a distinction to be made
 128 between increasing and decreasing cycles during the signal reconstruction. To properly apply the
 129 FPC, the turning points, which represent the peak and the valley of the signal, are extracted before the
 130 ARM is constructed. This extraction deletes variations that are neither peaks nor valleys (these small
 131 variations in the rainflow approach are considered to have minor impact on the fatigue assessment).



132
133

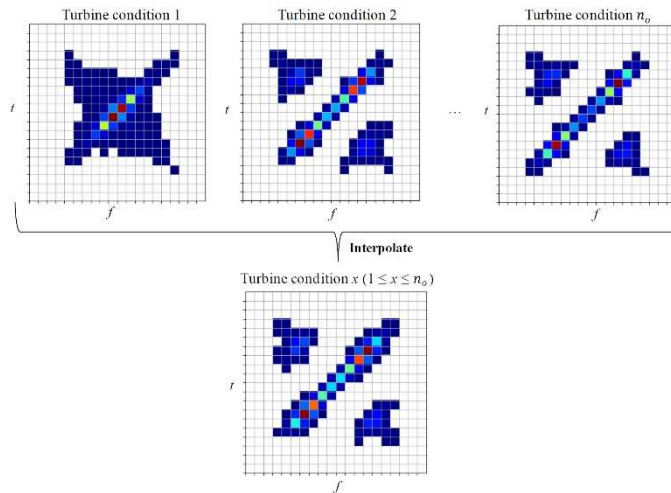
Figure 2. Four Point Counting algorithm applied for strain signal.

134 2.2 Kriging Interpolation Method

135 Before presenting the interpolation method, the context must be clarified. The interpolation input
 136 is a set of ARM histograms, which are spread across the different turbine operating conditions. The
 137 histogram bins are fixed for all the histograms. The goal is to estimate new ARM histograms at any
 138 given operating conditions (see Figure 3).

139 The relation between the bins of each ARM matrix is an important bit of information that cannot
 140 be ignored in the interpolation phase. Hence, the kriging interpolation method is a suitable method to
 141 preserve the spatial covariance within the ARM histogram. However, there is also a covariance to
 142 preserve between each operating condition. Thus, a multivariate kriging method, inspired by Spatio-
 143 Temporal Kriging (STK), where the temporal and spatial coordinates are treated separately, is
 144 applied, and developed for this study. Readers can consult the general definition of the kriging
 145 principle in [27-30].

146



147
148
149

Figure 3. Interpolation objective: new ARM histogram (at a given turbine operating condition) is interpolated from all experimental ARM histograms

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

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

152 where p_i is a given bin position in the ARM histogram (p depends on a 'from value' f_i and a 'to
 153 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

154 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
 155 the regression weight to optimize during the interpolation process (details below), n_o is the number of
 156 experimental turbine conditions, ϑ_i is a set of neighbourhood positions of p_i (including p_i). It must
 157 be noted that the spatial covariance map is obtained from the whole histograms of all the experimental
 158 turbine conditions, while Eq. 1 searches for the input in the neighbourhood ϑ_i of the target bin. This
 159 reduces the calculation time and avoids the interpolation smoothing effect.

160 Kriging allows the unbiased estimation of values at given locations by determining the weights
 161 λ_{kl} which minimize the error variance $\sigma_e^2(p_i, o_x)$ between the real value B and the interpolated value
 162 B^* at the same location, $\sigma_e^2(p_i, o_x) = Var(B(p_i, o_x) - B^*(p_i, o_x))$. The condition for non-bias
 163 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
 164 in every location. Therefore, to impose this non bias constraint, the minimization system is updated by
 165 adding a Lagrange parameter μ :

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

166 where $L(\lambda_i, \mu)$ is the Lagrange function. Making use of Eq.1, the minimization of Eq.2 leads to the
 167 system that allows to obtain the λ_{kl} and μ :

$$\begin{cases} \sum_{k \in \vartheta_i} \sum_{l=1}^{n_o} \lambda_{kl}(p_i, o_x) \gamma(h_{\|p_k - p_j\|}, u_{|o_l - o_m|}) - \mu = \gamma(h_{\|p_i - p_j\|}, u_{|o_x - o_m|}), j \in \vartheta_i, m = 1, \dots, n_o \\ \sum_{k \in \vartheta_i} \sum_{l=1}^{n_o} \lambda_{kl}(p_i, o_x) = 1 \end{cases} \quad (3)$$

168 where

$$\gamma(h_{\|p_a - p_c\|}, u_{|o_b - o_d|}) = 0.5 \times Var[B(p_a, o_b) - B(p_c, o_d)] \quad (4)$$

169 is the semivariogram, which depends on the increments h and u , p_a and p_c are two positions of the
 170 bin in the histogram, o_b and o_d are two turbine operating conditions such that: $h = \|p_a - p_c\|$ and
 171 $u = |o_b - o_d|$. Based on Eq. 4, the semivariogram is defined as the half variance between two values
 172 separated by the increments h and u . This semivariogram can be considered as a simpler
 173 representative of the covariance map that does not require information on the mean value of B (that is
 174 often unknown in practice [29]). To obtain the semivariogram at any possible increment (h, u) , some
 175 numerical models can be used to fit to the semivariogram using analysis estimations (e.g., Least
 176 Squares method) [27][30]. However, to choose the best numerical model, one needs to create a
 177 semivariogram from the experimental data. In practice, the quantity of data is not enough to provide
 178 values exactly separated by $h = \|p_a - p_c\|$ and $u = |o_b - o_d|$, as in Eq. 4. Consequently, to create
 179 the Experimental Semivariogram (ES), Eq. 5 is used. This equation allows an estimation of the
 180 semivariogram value associated with a set of (h, u) values, where h and u are midpoints of imposed
 181 spatial increment intervals:

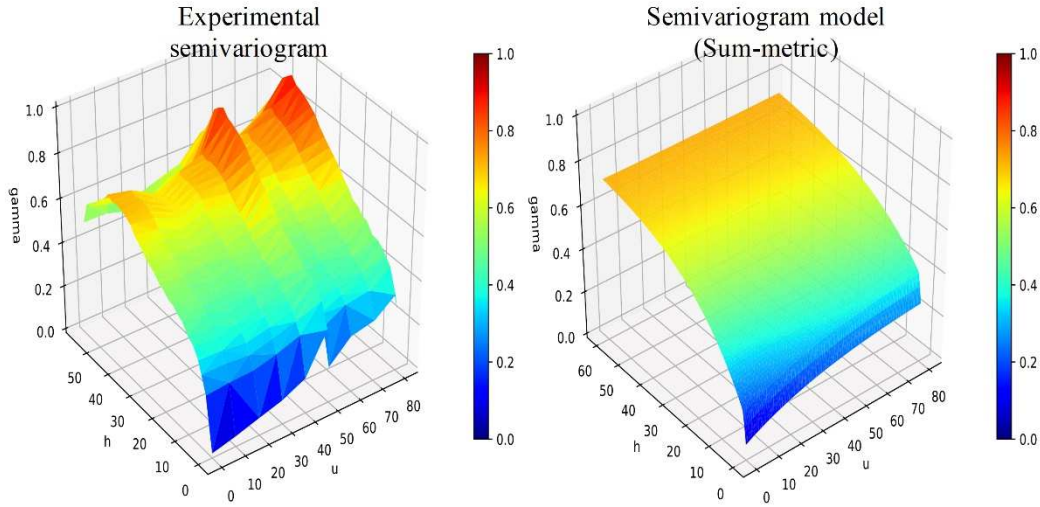
$$\hat{\gamma}(h, u) = \frac{1}{2N(h, u)} \sum_{(a,b,c,d) \in v(h,u)} [B(p_a, o_b) - B(p_c, o_d)]^2 \quad (5)$$

182 where $\hat{\gamma}(h, u)$ is the experimental semivariogram, $v(h, u)$ is a set of paired points separated such that
 183 $\|p_a - p_c\| \in [h - dh, h + dh]$ and $|o_b - o_d| \in [u - du, u + du]$ with $dh > 0$, $du > 0$, $N(h, u)$ is
 184 cardinal of $v(h, u)$. The next step is to fit the ES with a numerical model that will make a continuity
 185 throughout the increments for the semivariogram. Classes of semivariogram models are presented in
 186 [29-30]. For example, the sum-metric model (Figure 4), which turns out as one of the best fitting
 187 models in our study, is expressed as:

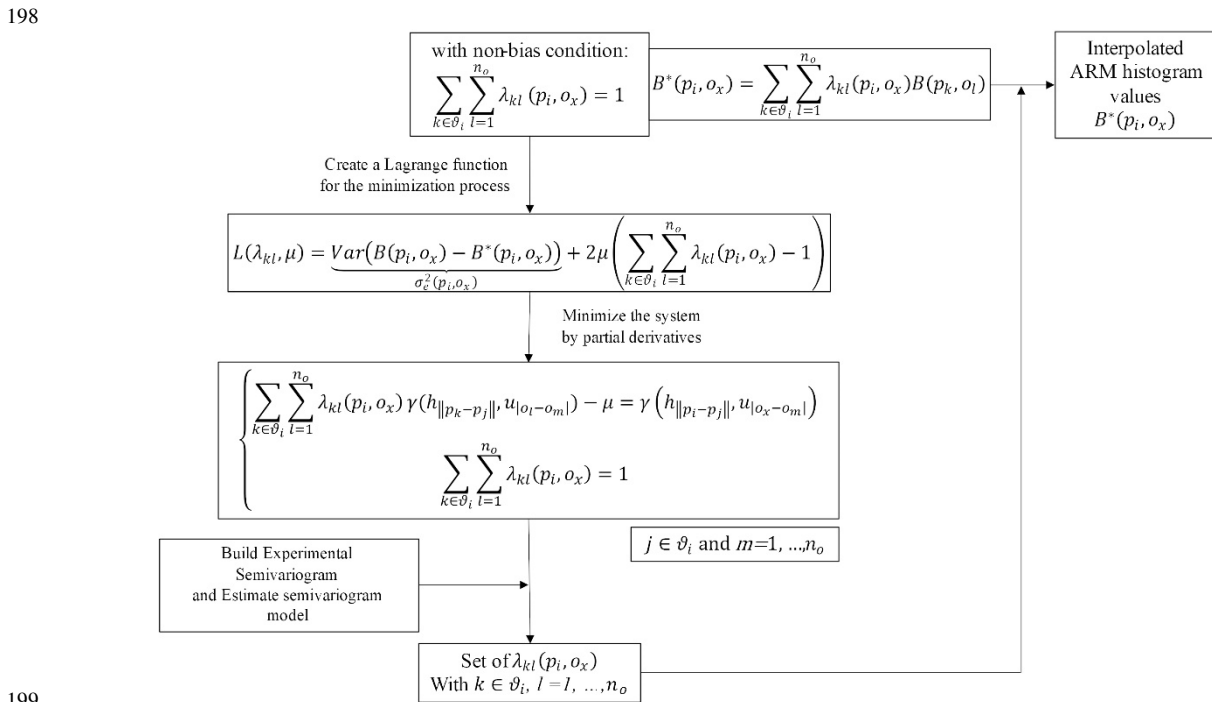
$$\gamma_{Smetric}(h, u) = \gamma_p(h) + \gamma_o(u) + \gamma_{joint}(\sqrt{h^2 + (\kappa \cdot u)^2}) \quad (6)$$

188 where $\gamma_p(h)$ is the univariate semivariogram model for the variable p , $\gamma_o(u)$ is the univariate
 189 semivariogram for the variable o , and $\gamma_{joint}(\sqrt{h^2 + (\kappa \cdot u)^2})$ is a metric joint variogram model
 190 including an anisotropy correction κ [30]. Many basic analytical models exist for the univariate
 191 semivariograms such as spherical model, the exponential model, the Gaussian model, power model,

192 etc. [14][27]. The schema in *Figure 5* summarizes the kriging interpolation approach used in our case
 193 study.
 194



195
 196 *Figure 4. Experimental semivariogram and example of semivariogram model (Sum-metric). Semivariogram values in this*
 197 *figure are normalized.*

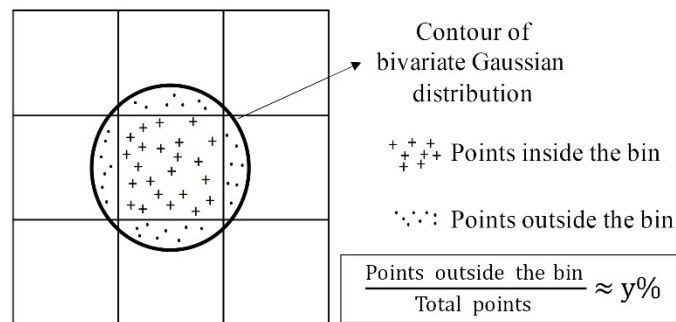


199
 200 *Figure 5. Schema of the Kriging process in this paper applied for the studied variables*

201 2.3 Simulation Reconstruction

202 To generate strain cycles from an interpolated ARM histogram, it must be converted to a
 203 continuous distribution. A smoothing process for the interpolated histogram is proposed by
 204 considering that the cycles inside each bin follow a Bivariate Gaussian Distribution (BGD). The BGD
 205 mean corresponds to the coordinates of the center point of the bin and the proposed covariance matrix
 206 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
 207 denominator is a coefficient proposed to ensure that the simulated cycle distribution does not create
 208 too much variability outside its bin. A variability test is proposed to select this variability parameter.
 209 In this study, since the interpolation input comes from the experimental strains measured by gauges,
 210 we impose the strain gauge uncertainty as a limited level of variability, which limits the variability of

211 generated point outside each bin. For example, if the relative error of the strain gauge is $\pm y\%$, the
 212 imposed limited level for the variability will be $y\%$. It must be noted that the experimental
 213 uncertainties are not the focused problem of this study. The test of variability is then launched for
 214 several variability parameters. For each ARM, Monte Carlo simulation is generated from the bivariate
 215 normal distribution with the mean $[0,0]$ and the covariance $\sigma^2 \mathbf{I}_2$. If the percentage of the sample
 216 values outside the bin dimension is approximately equals to the limited level $y\%$ (Figure 6), this
 217 variability parameter is chosen for the ARM histogram smoothing process. The reader can also
 218 optimize the variability parameter by using or adding other uncertainty types (e.g., error of kriging
 219 interpolation) if needed for other study. In addition, other approaches, such as the Kernel Density
 220 Estimation (KDE), can also be useful for smoothing the histogram [19-20]. Finally, the cycles are
 221 randomly generated from the set of BGDs by respecting the probability of occurrence of cycles in
 222 each bin (the sum of all the probability values in the histogram is equal to 1). This generation process
 223 allows an observation of a set of simulations, which is useful for the industry when evaluating the
 224 fatigue. The risk of fatigue will be predicted with a set of possibilities instead of a unique prediction.
 225



226
 227

Figure 6. Percentage of the sample values outside the bin dimension is approximately equal to $y\%$

228 In addition, an inverse method is used to reconstruct the strain signal from the simulated samples
 229 [19-20]. The signal reconstruction consists of two steps: First, the largest cycles (the furthest points
 230 from the ARM diagonal), either a Decreasing Cycle (DC) or an Increasing Cycle (IC), are used to
 231 create an initial reconstructed signal (also called the *remaining residue* in Figure 2). Second, all the
 232 other cycles are randomly inserted into the remaining residue descending order (from the furthest to
 233 the closet point from the diagonal) by respecting the rule proposed by Socie et al. [19-20]. The smaller
 234 IC is added to the position of the larger DC (peak and valley values of the smaller cycle fall within the
 235 range of the larger cycle) and vice versa. Each cycle added to the remaining residue is removed in the
 236 ARM until there are no cycles left. It is not expected that the reconstructed signal will be similar to
 237 the experimental one because only the rainflow information is considered for the interpolation;
 238 nevertheless, the signal should represent a time history that generates equivalent fatigue damage. The
 239 flowchart in Figure 7 illustrates the interpolation scheme used in this study.
 240

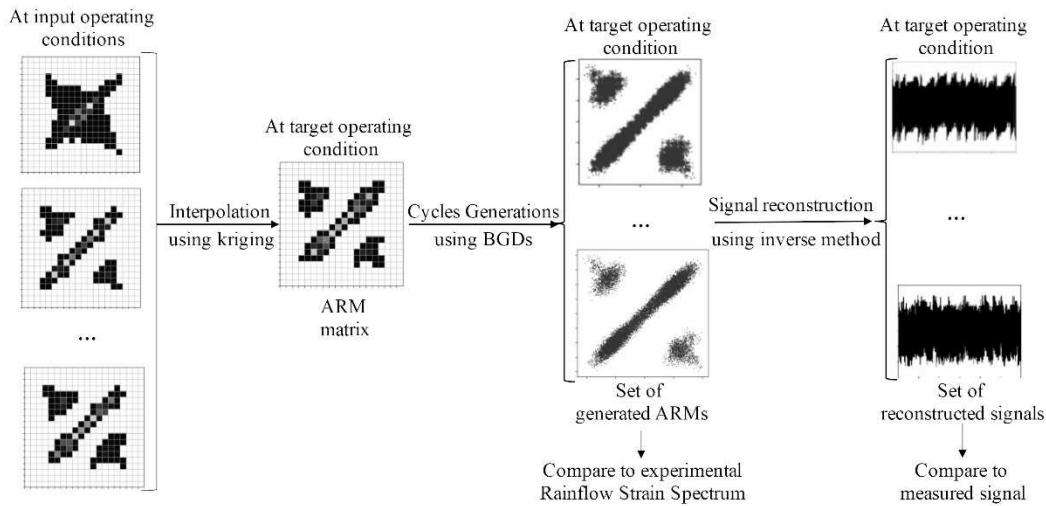


Figure 7. Flowchart displays the interpolation scheme

241
242

3. 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

243
244
245
246
247
248
249

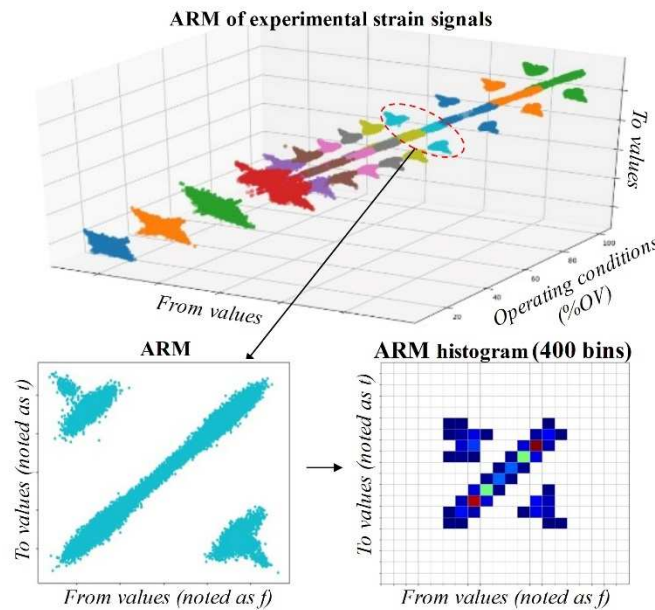


Figure 8. ARMs over different operating conditions for this case study and example of ARM histogram (the white bin corresponds to the null value)

250
251
252

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 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 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

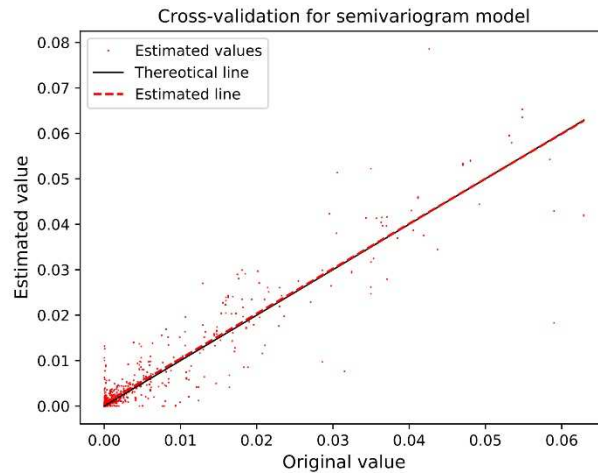
253
254
255
256
257
258
259
260
261
262

263 imprecise results because of the large interval of values covered by each bin.

264 Fitting a model to the experimental semivariogram is an important step of the interpolation
 265 process. We highly recommend using the cross-validation technique for this step [29]. This technique
 266 removes some point values in the input set and then predicts these values from the remaining points
 267 using fitted semivariogram models. This procedure is repeated several times for each semivariogram
 268 model to observe the interpolation performance of this model. Moreover, other error estimations such
 269 as Root Mean Square Error (RMSE) or Mean Absolute Error (MAE) are additionally used to obtain
 270 the best semivariogram model. In our case study, based on a good cross-validation performance
 271 (Figure 9) and on the small error estimation values (see MAE and RMSE in Table 1), the chosen
 272 semivariogram model was sum-metric model (see Eq. 6 and Table 1). For each univariate model,
 273 $\gamma_p(h)$ was modeled by sum of an exponential model and a Gaussian model, $\gamma_o(u)$ and
 274 $\gamma_{joint}(\sqrt{h^2 + (\kappa \cdot u)^2})$ was respectively modeled by a Gaussian and an exponential model (Table 1).

$$\begin{cases} \gamma_{exponential}(h) = (Sill - nu) \cdot \left(1 - e^{-\frac{3 \cdot h}{range}}\right) + nu \\ \gamma_{Gaussian}(u) = (Sill - nu) \cdot \left(1 - e^{-\frac{(3 \cdot u)^2}{range^2}}\right) + nu \end{cases} \quad (7)$$

275 where *Sill* corresponds to the average variance where the experimental semivariogram stabilizes,
 276 *range* presents the distance from which no more correlations exist between the data, and *nu* is a small
 277 value added to the sill parameter to represent a very short-range variability in the dataset (Nugget
 278 effect) [14][27].



279
 280 Figure 9. Cross validation performance of sum-metric model (its parameters are presented in Table 1)

281 The neighbourhood set ϑ_i (Eq. 1) is chosen using the cross-validation technique. Several sets with
 282 different numbers of neighbourhood positions are tested and the optimized set is chosen based on the
 283 smallest RMSE and based on the similarity between theoretical and estimated line (see APPENDIX).
 284 Notice that the set of neighbourhoods is applied only for the histogram position “*p*”, but not for
 285 operating conditions “*o*”. Thus, the points in a neighbourhood set correspond to the closest points
 286 from the target points in ARMs. In this case study, the set of 9 nearest positions is chosen for the
 287 neighbourhood set ϑ_i in Eq. 1 (see APPENDIX).

288 Table 1 shows the optimized parameters used for this case study. To verify the quality of the
 289 interpolated result, a validation is made using known experimental operating conditions (not included
 290 in the set of interpolation inputs).

291

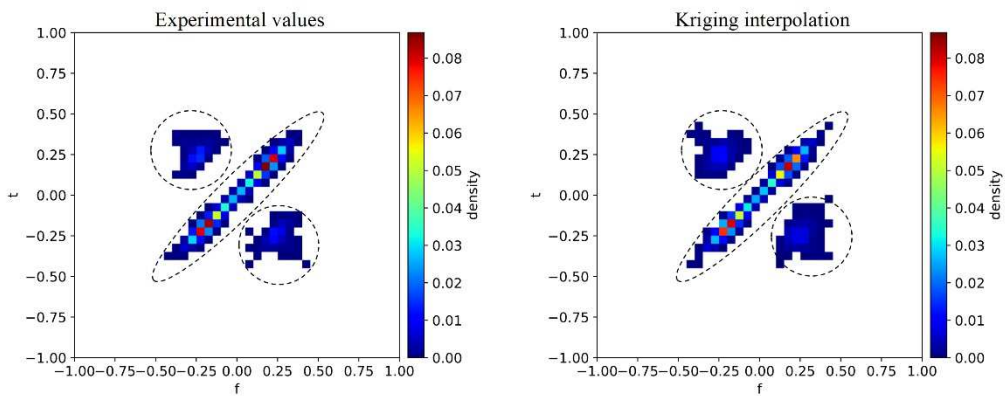
Parameters	Semivariogram model parameters (fitted to normalized ES- see Figure 4)					
Input/Output histogram bins number:	Optimized sum-metric model (RMSE = 0.0993, MAE= 0.0747)					
	Eq. 6: $\gamma_{Smetric}(h, u) = \gamma_p(h) + \gamma_o(u) + \gamma_{joint}(\sqrt{h^2 + (\kappa \cdot u)^2})$					
	Model type	<i>Sill</i>	<i>nu</i>	<i>range</i>	κ	

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	$\gamma_p(h)$	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	

292 *Table 1. Parameters of the proposed interpolation process adapted for the case study dataset*

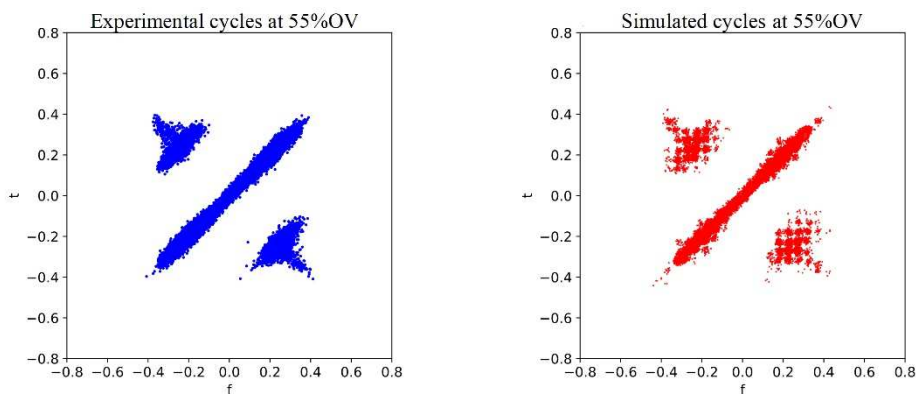
293 **4. Results and Discussions**

294 The interpolated results (corresponding to each step illustrated in *Figure 7*) and their comparison with
 295 experimental ones at 55% OV are presented in the *Figures 10-12*. All the results are normalized to
 296 respect the confidential information.
 297



298 *Figure 10. Comparison between experimental and Interpolated ARM histogram at 55% OV*

300



301 *Figure 11. Experimental and generated cycles (one simulation) at 55 %OV*
 302

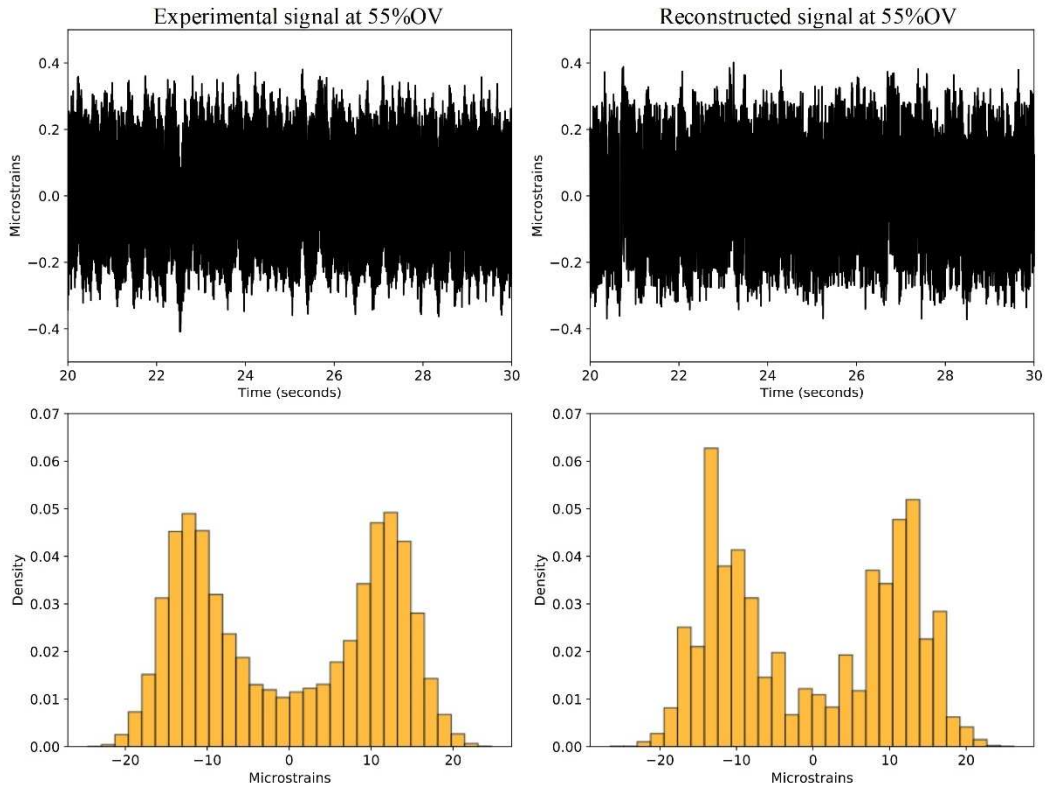


Figure 12. Comparison between experimental signal and one reconstructed signal at 55 %OV.
Above: 10 seconds of signals. Below: Histogram of whole signal

303
304
305

306 The interpolated ARM shows the same number of distribution modes (dashed line in *Figure 10*) as
307 in the experimental ones; this means that this interpolation did not generate large overestimation in
308 positions the null bins. Nonetheless, some differences are observed at the frontier positions of each
309 mode in the interpolation result (*Figure 10, 11*). Graphically, the reconstructed signal (*Figure 12*) is
310 not similar to the experimental one like we have foreseen, however its amplitudes have the same level
311 as the experimental one. To verify the quality of interpolation process, we have compared the rainflow
312 strain spectrums (strain range as a function of cumulative number of fatigue cycles) between 500
313 simulations (sampled from the set of BGDs) and the experimental values at some operating conditions
314 (*Figure 13*).

315 For fatigue assessment, the expected goal in this study is to capture the amplitude/range level of
316 fatigue cycle (especially high range cycles which cause more damage). At 55 %OV and 65 %OV, the
317 experimental rainflow (the blue line in *Figure 13*) is mostly within of the set of simulations while
318 there seems to be some overestimations at the high ranges at 80 %OV. This difference might be
319 explained by that there are more operating conditions close to the 55 %OV, 65 %OV in the
320 interpolation input set (see *Table 1*).

321 An unusual variability is observed at 30 %OV. The interpolation result depends on the nature of
322 known data. The more complicated the combination of phenomena hidden in the signal, the less
323 confident we might be in the interpolation result. Therefore, at some partial load conditions of turbine,
324 this unusual variability of the interpolated result occurs when an operating condition with a
325 complicated dynamic behavior is presented next to the interpolation target condition. Some of those
326 conditions contain specific fluctuation coming from the vortex in the draft tube [3], which appears and
327 disappears over limited set of measured operating conditions rendering the interpolation difficult. This
328 problem provides new challenge for further research, in which we would like to recover such
329 fluctuation information.

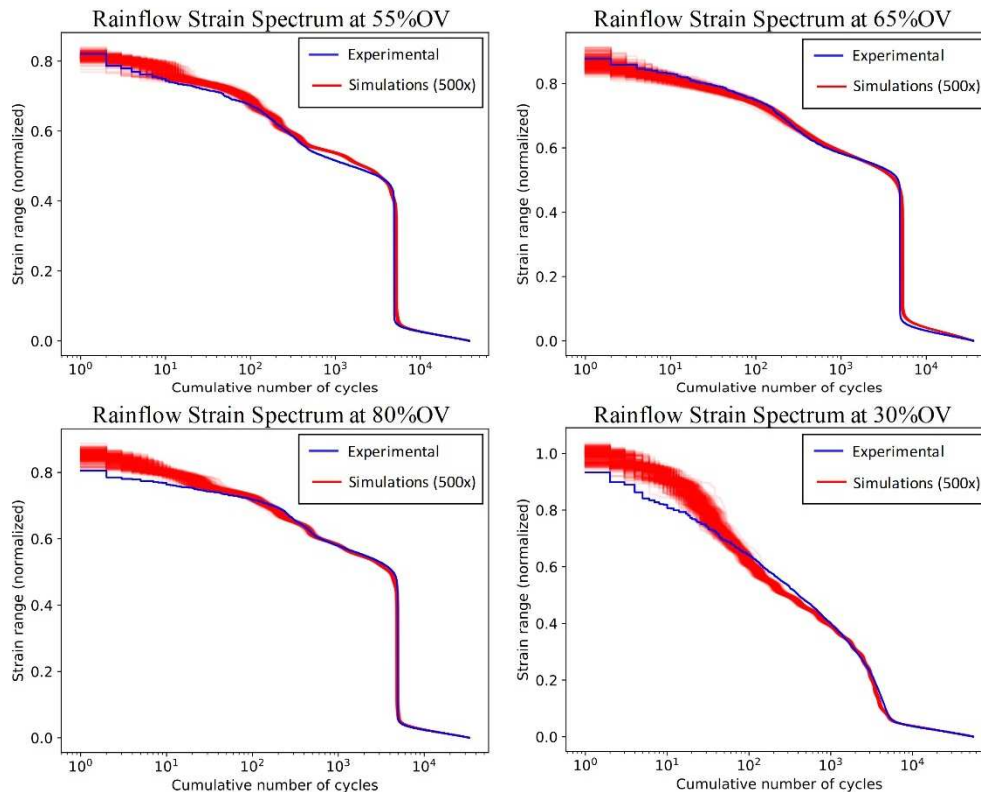


Figure 13. Rainflow strain spectrum between experimental values and simulations

330
331

332 For the runner maintenance strategy, even if the reconstructed signal is not similar to the
333 experimental one (Figure 12), the interpolated strain ranges might be more representative of data
334 variability and uncertainty than reusing experimental measurements from nearby operating conditions.
335 Also, the proposed interpolation can be a solution to reduce the measurement requirements for runner
336 fatigue assessment.

337

338 5. Conclusions

339 In this article, an interpolation process using kriging has been presented to estimate the strain range at
340 any steady state operating condition based on rainflow cycle counting for hydroelectric turbine runner
341 fatigue assessment. A set of bivariate Gaussian distributions were used to smooth the histogram,
342 allowing quicker simulations of the loading cycles. It must be noted that the proposed interpolation
343 process is sensitive to the parameters used (i.e., the semivariogram model, etc.), using the error
344 optimizing methods (such as cross-validation, Mean Absolute Error, etc.) are therefore highly
345 recommended. The results of the case study show that even when the reconstructed signal is not
346 similar to the experimental one, our proposed interpolation might be useful to estimate the dynamic
347 strain level using only the load cycles information of strain signal.

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

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

356

357 Acknowledgments

358 The author would like to thank the Hydro-Québec Research Institute (IREQ), Andritz Hydro
359 Canada LTD, INSA de Lyon (France), École de Technologie Supérieure de Montréal (Quebec,
360 Canada) and the Mitacs Acceleration program for their support and financial contribution.

361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420

About 4410 words (excluding abstract, figure captions and references). Figure 10, 13 must be printed in color

References

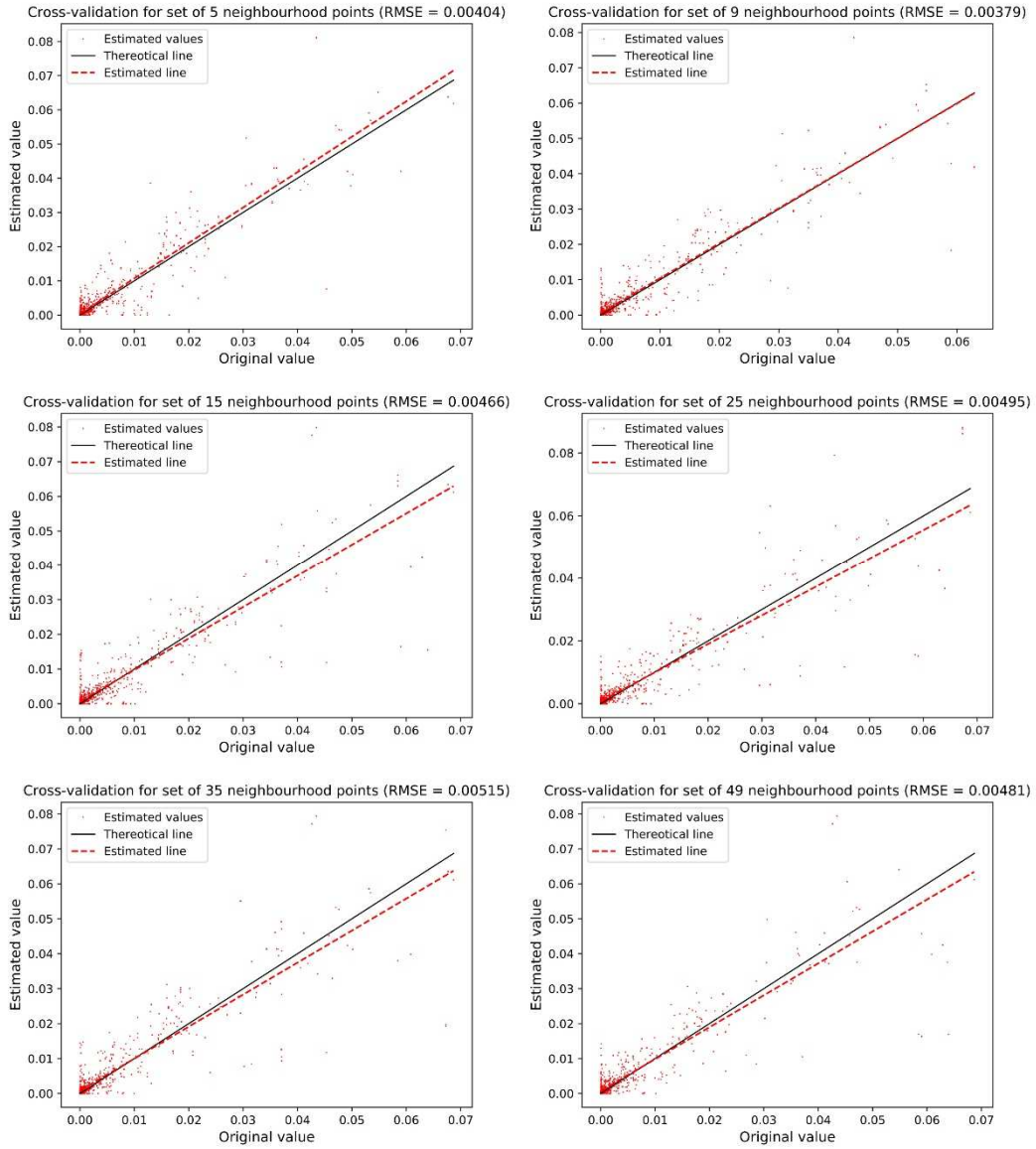
- [1] Monette, C., Marmont, H., Chamberland-Lauzon, J., Skagerstrand, A., Coutu, A., & Carlevi, J. (2016, November). Cost of enlarged operating zone for an existing Francis runner. In *IOP Conference Series: Earth and Environmental Science* (Vol. 49, No. 7, p. 072018). IOP Publishing.
- [2] Liu, X., Luo, Y., & Wang, Z. (2016). A review on fatigue damage mechanism in hydro turbines. *Renewable and Sustainable Energy Reviews*, 54, 1-14.
- [3] Seidel, U., Mende, C., Hübner, B., Weber, W., & Otto, A. (2014). Dynamic loads in Francis runners and their impact on fatigue life. In *IOP conference series: earth and environmental science* (Vol. 22, No. 3, p. 032054). IOP Publishing.
- [4] 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.
- [5] 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.
- [6] Trudel, A., & Sabourin, M. (2014). Metallurgical and fatigue assessments of welds in cast welded hydraulic turbine runners. In *IOP Conference Series: Earth and Environmental Science* (Vol. 22, No. 1, p. 012015). IOP Publishing.
- [7] Thibault, D., Gagnon, M., & Godin, S. (2014). Bridging the gap between metallurgy and fatigue reliability of hydraulic turbine runners. In *IOP Conference Series: Earth and Environmental Science* (Vol. 22, No. 1, p. 012019). IOP Publishing.
- [8] El Haddad, M. H., Topper, T. H., & Smith, K. N. (1979). Prediction of non propagating cracks. *Engineering fracture mechanics*, 11(3), 573-584.
- [9] 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.
- [10] Nennemann, B., Morissette, J. F., Chamberland-Lauzon, J., Monette, C., Braun, O., Melot, M., ... & Giroux, A. M. (2014). Challenges in dynamic pressure and stress predictions at no-load operation in hydraulic turbines. In *IOP conference series: earth and environmental science* (Vol. 22, No. 3, p. 032055). IOP Publishing.
- [11] Morissette, J. F., Chamberland-Lauzon, J., Nennemann, B., Monette, C., Giroux, A. M., Coutu, A., & Nicolle, J. (2016, November). Stress predictions in a Francis turbine at no-load operating regime. In *IOP Conference Series: Earth and Environmental Science* (Vol. 49, No. 7, p. 072016). IOP Publishing.
- [12] 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.
- [13] 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.
- [14] Ben Salah, F. (2014). *Modélisation de la propagation des incertitudes des mesures sur l'aube d'une turbine hydraulique par Krigeage et simulations stochastiques* (Master dissertation, École de technologie supérieure, QC, Canada).
- [15] 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.
- [16] 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 *SURVISHNO conference*, Lyon, France.
- [17] Diagne, I., Gagnon, M., & Tahan, A. (2016, November). Modeling the dynamic behavior of turbine runner blades during transients using indirect measurements. In *IOP Conference Series: Earth and Environmental Science* (Vol. 49, No. 7, p. 072014). IOP Publishing.
- [18] Arpin-Pont, J., Gagnon, M., Tahan, S. A., Coutu, A., & Thibault, D. (2012). Strain gauge measurement uncertainties on hydraulic turbine runner blade. In *IOP Conference Series: Earth and Environmental Science* (Vol. 15, No. 6, p. 062042). IOP Publishing.
- [19] Socie, D. (2001). Modelling expected service usage from short-term loading measurements. *International Journal of Materials and Product Technology*, 16(4-5), 295-303.
- [20] Socie, D. F., & Pompetzki, M. A. (2004). Modeling variability in service loading spectra. *Journal of ASTM International*, 1(2), 1-12.
- [21] Johannesson, P. (2006). Extrapolation of load histories and spectra. *Fatigue & Fracture of Engineering*

- 421 *Materials & Structures*, 29(3), 209-217.
- 422 [22] Johannesson, P., & Thomas, J. J. (2001). Extrapolation of rainflow matrices. *Extremes*, 4(3), 241-262.
- 423 [23] Matsuishi, M., & Endo, T. (1968). Fatigue of metals subjected to varying stress. *Japan Society of*
424 *Mechanical Engineers, Fukuoka, Japan*, 68(2), 37-40.
- 425 [24] Johannesson, P. (1999). Rainflow analysis of switching Markov loads. *Lund Institute of Technology, Centre*
426 *for Mathematical Sciences, Mathematical Statistics, Lund Institute of Technology, Sweden*, 1-27.
- 427 [25] Dreßler, K., Hack, M., & Krüger, W. (1997). Stochastic reconstruction of loading histories from a rainflow
428 matrix. *ZAMM-Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und*
429 *Mechanik*, 77(3), 217-226.
- 430 [26] Rychlik, I. (1996). Simulation of load sequences from rainflow matrices: Markov method. *International*
431 *Journal of Fatigue*, 18(7), 429-438.
- 432 [27] Deutsch, C. V., & Journel, A. G. (1992). Geostatistical software library and user's guide. *New York*, 119,
433 147.
- 434 [28] Goovaerts, P. (1997). *Geostatistics for natural resources evaluation*. Oxford University Press on Demand.
- 435 [29] Montero, J. M., Fernández-Avilés, G., & Mateu, J. (2015). *Spatial and spatio-temporal geostatistical*
436 *modeling and kriging* (Vol. 998). John Wiley & Sons. (page 266 to 273).
- 437 [30] Pebesma, E., & Heuvelink, G. (2016). Spatio-temporal interpolation using gstat. *RFID Journal*, 8(1), 204-
438 218.
- 439 [31] Chiles, J. P., & Delfiner, P. (2009). *Geostatistics: modeling spatial uncertainty* (Vol. 497). John Wiley &
440 Sons, (page 49-54).
- 441 [32] WAFO Group. (2017). WAFO—a Matlab toolbox for analysis of random waves and loads, tutorial for
442 WAFO version 2017. *Mathematical Statistics, Lund University*.
- 443 [33] Sherman, M. (2011). *Spatial statistics and spatio-temporal data: covariance functions and directional*
444 *properties*. John Wiley & Sons.
- 445 [34] Duda, R. O., Hart, P. E., & Stork, D. G. (2001). Pattern classification 2nd ed. *John Willey & Sons Inc*.
- 446

447 **Appendix: The cross-validation performances for several neighbourhood sets**

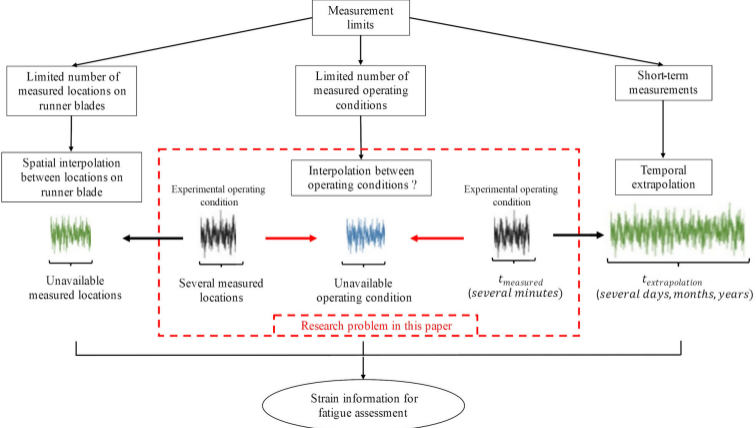
448 The cross-validation technique was performed for several neighbourhood sets (with the same
449 semivariogram model). The optimized neighbourhood set for the interpolation is chosen based on the
450 smallest RSME and based on the similarity between theoretical (black) and estimated line (red) in the
451 cross-validation. In observing *Figure 14*, if one chooses a small set (the case of 5 points) or a big set
452 (the cases of 35 or 49 points), the variation and the RMSE values are larger. In this case study, the set
453 of 9 neighbourhood points is chosen (row 1 and column 2 in *Figure 14*).

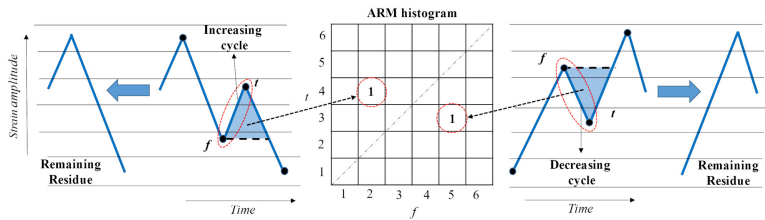
454



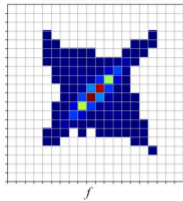
455
456
457

Figure 14. Cross-validation performances for several neighbourhood sets

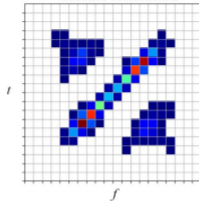
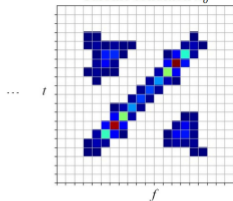




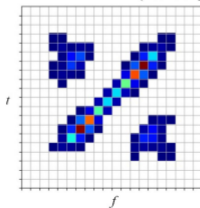
Turbine condition 1



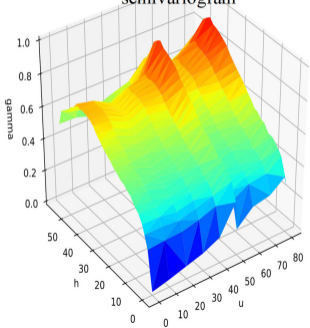
Turbine condition 2

Turbine condition n_o 

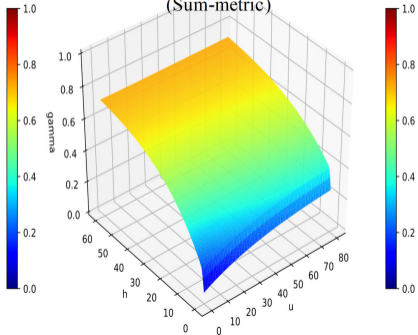
Interpolate

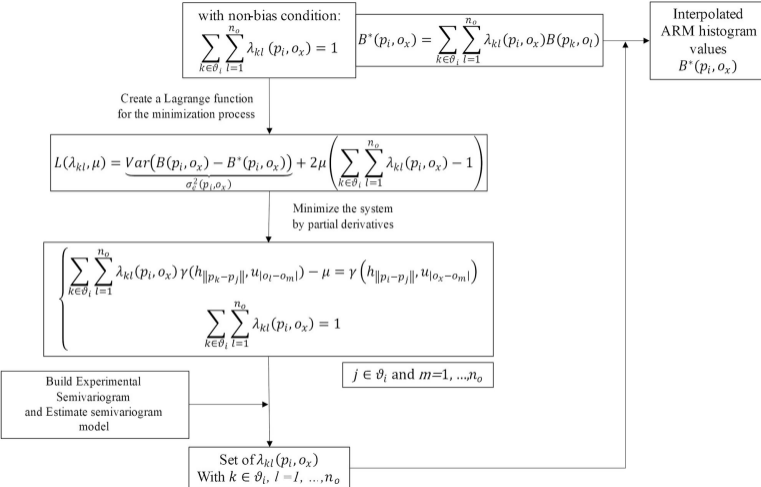
Turbine condition x ($1 \leq x \leq n_o$)

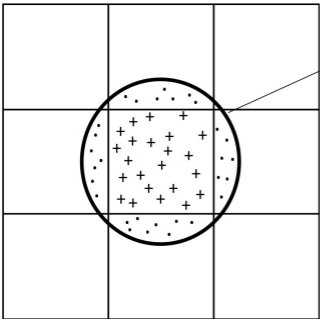
Experimental
semivariogram



Semivariogram model
(Sum-metric)







Contour of
bivariate Gaussian
distribution

+ + +
+ + + Points inside the bin

. . . . Points outside the bin

$$\frac{\text{Points outside the bin}}{\text{Total points}} \approx y\%$$

At input operating conditions



...



Interpolation
using kriging

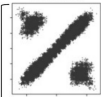
At target operating condition



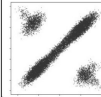
ARM
matrix

Cycles Generations
using BGDs

At target operating condition



...



Set of
generated ARMs

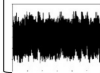
Compare to experimental
Rainflow Strain Spectrum

Signal reconstruction
using inverse method

At target operating condition



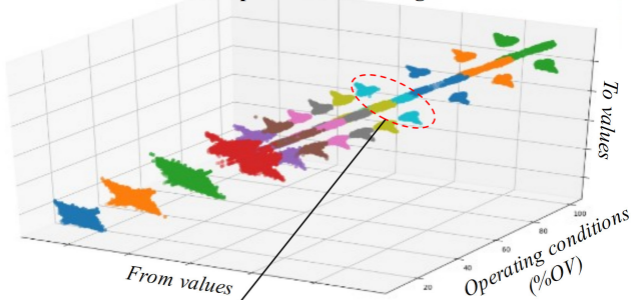
...



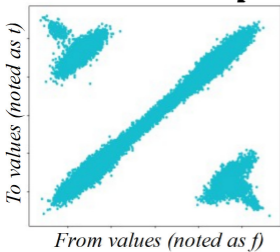
Set of
reconstructed signals

Compare to
measured signal

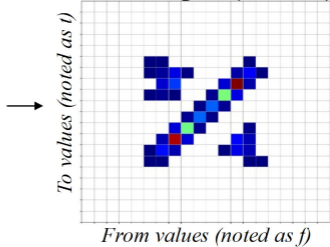
ARM of experimental strain signals



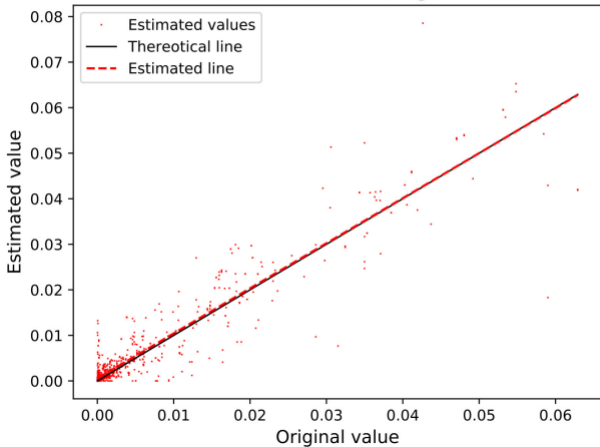
ARM



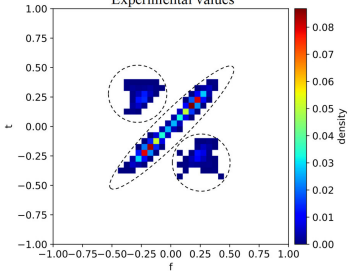
ARM histogram (400 bins)



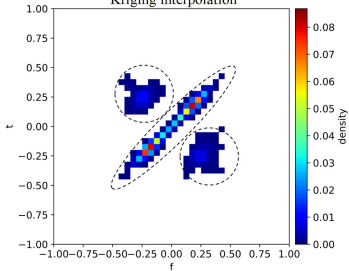
Cross-validation for semivariogram model



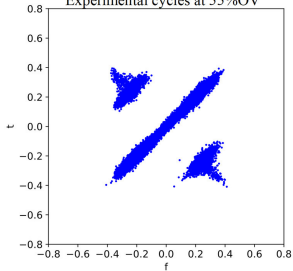
Experimental values



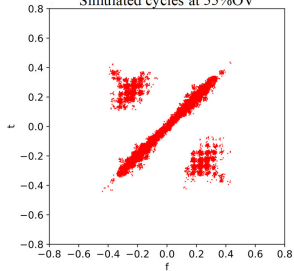
Kriging interpolation



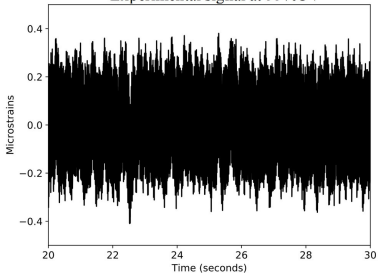
Experimental cycles at 55%OV



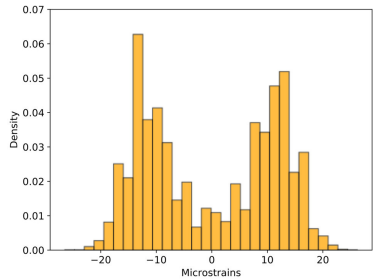
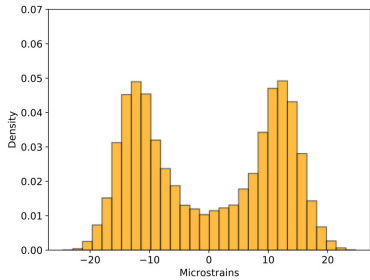
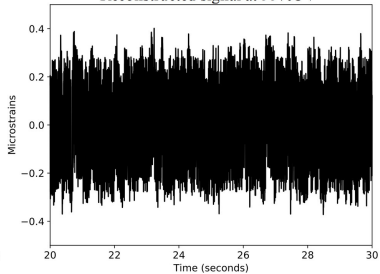
Simulated cycles at 55%OV



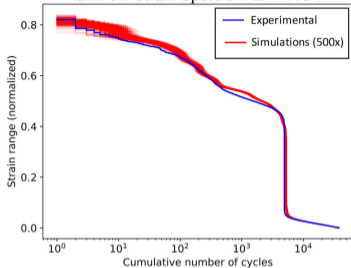
Experimental signal at 55%OV



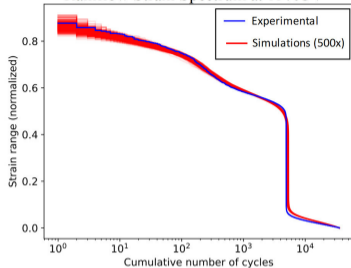
Reconstructed signal at 55%OV



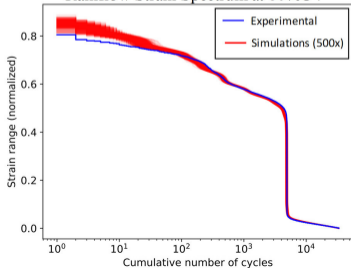
Rainflow Strain Spectrum at 55%OV



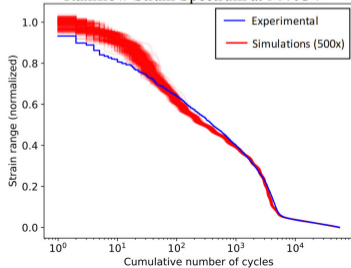
Rainflow Strain Spectrum at 65%OV



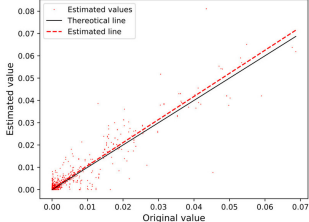
Rainflow Strain Spectrum at 80%OV



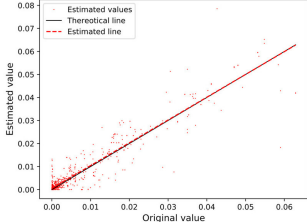
Rainflow Strain Spectrum at 30%OV



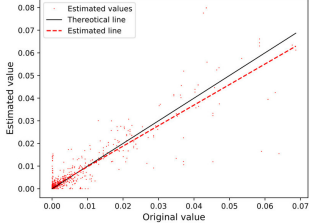
Cross-validation for set of 5 neighbourhood points (RMSE = 0.00404)



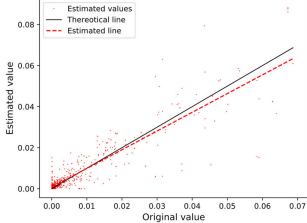
Cross-validation for set of 9 neighbourhood points (RMSE = 0.00379)



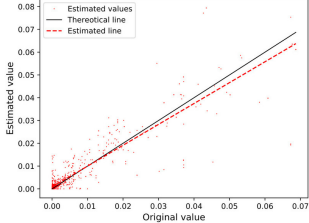
Cross-validation for set of 15 neighbourhood points (RMSE = 0.00466)



Cross-validation for set of 25 neighbourhood points (RMSE = 0.00495)



Cross-validation for set of 35 neighbourhood points (RMSE = 0.00515)



Cross-validation for set of 49 neighbourhood points (RMSE = 0.00481)

