Interpolation of periodic hidden signal measured at steady-operating conditions on hydroelectric turbine runners

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

Interpolation of periodic hidden signal measured at steady-operating conditions on hydroelectric turbine runners

Q H Pham[1], M Gagnon[2], J Antoni[3], S A Tahan[4], C Monette[5]

[1]Phd Student at ÉTS Montréal and INSA de Lyon
[2]Institut de recherche d'Hydro-Québec, Varenne, Québec, Canada
[3]Univ Lyon, INSA-Lyon, Laboratoire Vibrations Acoustique, Villeurbanne, France
[4] École de technologie supérieure (ÉTS), Montréal, Québec, Canada
[5]Andritz Hydro Canada LTD, Point-Claire, Québec, Canada

Email: quang-hung.pham.1@ens.etsmtl.ca

Abstract

The lack of experimental information can lead to an inaccurate prediction of hydroelectric turbine runners fatigue damage. Therefore, to recover this information, this research aim is the use of existing data measured by strain gauge to interpolate the unknown or not observed information about runner strain over the complete range of the steady-operating conditions for hydroelectric turbine. At steady-operating conditions, a strain signal, measured on the runner, can be separated into three principal components: static, periodic and stochastic. This paper presents the first step of our research that extracts and interpolates the periodic part at steady-operating states. A case study is used to compare two different kriging interpolation methods: the Spatial Kriging Method (based on 2D semivariogram) and the Spatio-Temporal Kriging Method (based on 3D semivariogram). The interpolation results are compared and validated with the experimental values.

1 Introduction

Strain levels play an important role in the fatigue reliability evaluation of hydroelectric turbine runners. Due to high costs and downtime required for the instrumentation, the experimental strain measurements on turbine runners are often limited (short measurement length, limited number of operation conditions, limited number of measured locations...), which could lead to an inaccurate evaluation of fatigue damage. Over the past years, many researches have tried to improve the limitation of experimental measurements. Poirier et al. (2016) regenerated strain signal over a long period by extrapolating a few minutes collected from the experimental measures [1]. Firas (2012) realized spatial interpolations between different locations on turbine runners to evaluate the distribution of damage and uncertainty propagation [4]. Gagnon et al. (2012) developed a model of turbine runner reliability by considering the High Cycle Fatigue (HCF) onset as the limit for the fatigue evaluation [2]. This model highlights the role of HCF, which is mainly linked to hydroelectric turbine Steady-Operating Conditions (noted as SOC). Currently, we are unable experimentally to obtain the dynamic strain in runner blades over all the possible operating conditions. Therefore, our research aims to estimate the missing experimental data by interpolating existing measured data. The idea is to develop an interpolation tool between different SOC available experimental data (Figure 1). This tool should help to make more accurate maintenance plan, leading to time and cost reduction.

The initial parameters used for interpolation are from measurements obtained by the strain gauges installed on a Francis turbine runner blade. Strain signals are complex and contain several physical phenomena. Some of these phenomena are hidden by others. Therefore, the interpolation of the complete signal is difficult. The proposition is to interpolate each phenomenon independently. We propose that the strain signals at SOC are separated into three principal components: static, periodic and stochastic. This paper presents a method that extracts and interpolates the periodic part. The periodic phenomenon hidden in the signal is linked with the

synchronous rotation speed of the turbine and is extracted by using the synchronous average operator to obtain the first order cyclostationary components. In this study, the interpolation method chosen is kriging. Kriging is a well-known spatial interpolation method that is commonly used in many domains, especially geostatistics. In general, the strain signal on turbine runner varies based on time and operating conditions of the turbine, but the influences of these two "dimensions" during interpolation phase are unclear. Thus, two kinds of kriging process were studied, spatial kriging and spatio-temporal one, for the interpolations of the periodic component. This initial interpolation between different operating conditions provides the basis for interpolations of the other components of the signal.

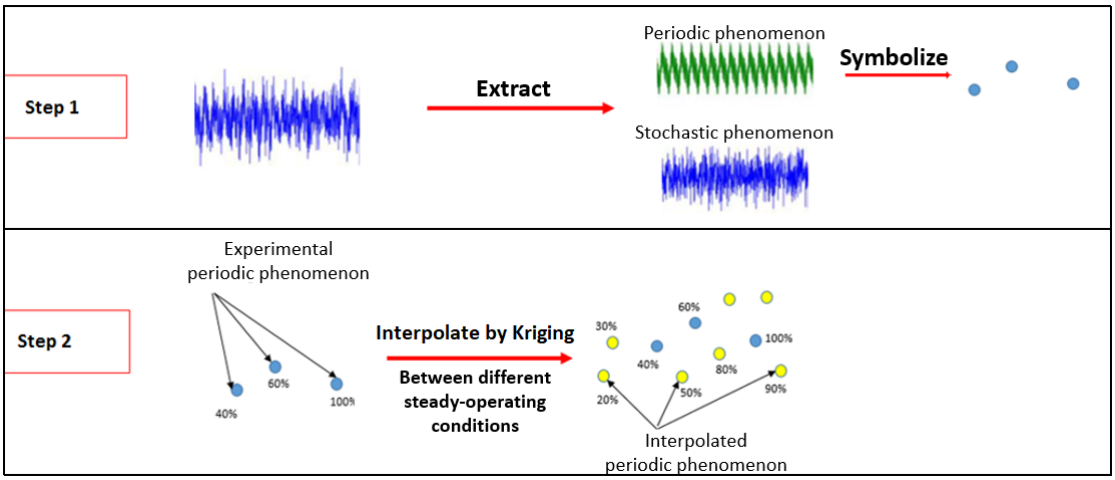


Figure 1 – Proposed interpolation tool

The paper is organized as follow: First, in Section 2, a description of the synchronous averaging, which extracts the periodic phenomenon in the signal, is presented. Then, in Section 3, the interpolation process is presented with spatial and spatio-temporal kriging. Finally, the interpolation results are shown and discussed in Section 4 and 5.

2 Synchronous Average - First order of cyclostationary

The strain signal in the case of the turbine consists of static, periodic and stochastic parts. The periodic parts in the strain signal measured on runner are typical phenomena related to hydroelectric turbine runners rotation (especially in the steady states where the speed of rotation is synchronized to the electrical network) while the stochastic phenomena correspond to the asynchronous parts (residue parts, stochastic vibrations, noises...). The separation will facilitate the interpolation because the relation between the phenomenon assigned to different operating conditions is less complex than the relation between whole signal where some phenomena are hidden by others. This separation leads to cyclostationnarity process, which will make possible the formalization of principles to extract hidden periodic signals. For the cyclostationary domain, the dynamic part of an experimental strain signal $S[n]$ measured on runner can be decomposed into three components: the cyclostationary moment of order 1 (noted as $CS1[n]$) linking to the rotation speed synchronous, the order 2 (noted as $CS2[n]$) referring to the periodic fluctuation of energy, and the residue part $R[n]$ [1].

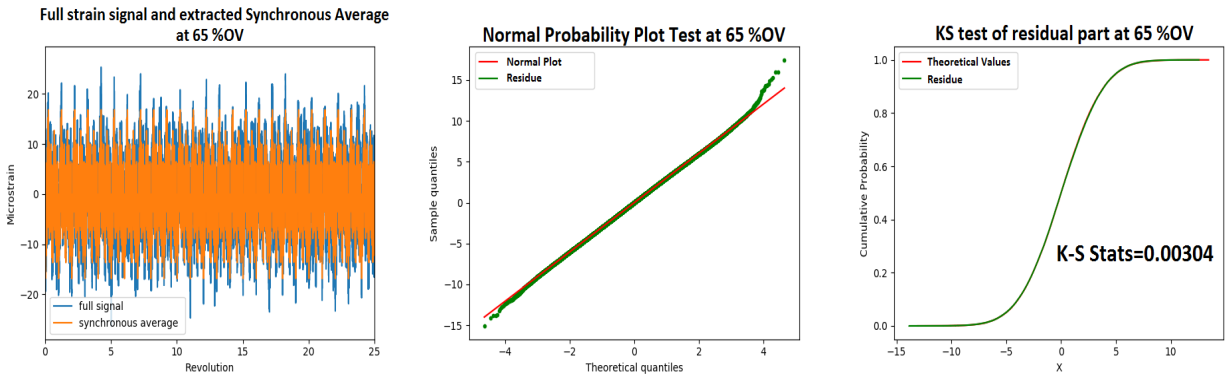
$$S[n] = CS1[n] + CS2[n] + R[n] \tag{1}$$

Synchronous Averaging is one of the extraction tools which allow the extraction of the $CS1$ (periodic part) hidden in the signal. The synchronous average (noted as SA), hidden in a signal $S[n]$ of finite cycles K, is extracted by using the equation (2).

$$CS1[n] = \frac{1}{K}\sum_{k=0}^{K-1} S[m+kN] \tag{2}$$

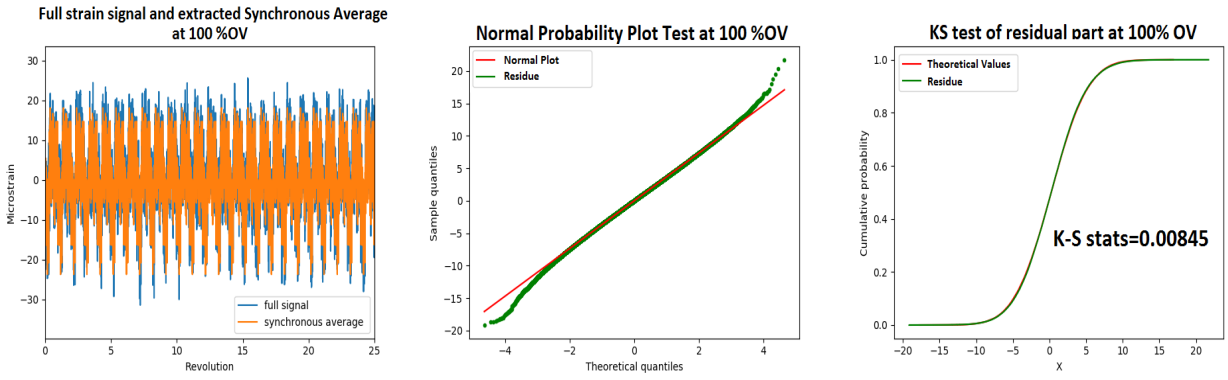
where K is number of cycles, N=L/K is length of cycle (L is full length of signal), m= mod[n,N]= n-[n/N]N.

This formula is applied for experimental strain signal to extract information at a cyclic frequency (including its harmonics), which relates to the synchronous rotation of the runner at steady-operating conditions. The periodic part can be directly extracted in the time domain. However, in the case of the rotating machinery like the hydroelectric turbine, the cyclostationnarity model is often based on the study of the angular domain rather than time. Therefore, an angular sampling of time signal (proposed by Bonnardot [5]) before the extraction is required in order to improve the quality of the synchronous average extraction. Figures 2 and 3 show the synchronous average at some operating conditions of the turbine and the statistic tests of residual part (after subtracting the average). Two statistic tests: the normal probability plot and the Kolmogorov-Smirnov test (noted as KS test) have been applied to verify the quality of the synchronous average extraction. Small KS statistic value and good fit between sample and theoretical quantiles verified that the residual part is almost aleatory. These results mean that the periodic part was almost fully extracted from the strain signal. The word "almost" is mentioned because there are slight divergences at extremity observed in Normal Probability graph. Higher order cyclostationary extraction could be applied to obtain a more aleatory residue. However, by these statistic test results, the synchronous average operator is an appropriate method for the initial extraction of periodic phenomena hidden in runner strain signal.



(a) (b) (c)

Figure 2 – Results of SA extraction process at 65 %OV. (a) Extracted Synchronous Average ; (b) Normal probability test of the residual part; (c) Kolmogorov-Smirnov test of the residual part.



(a) (b) (c)

Figure 3 – Results of SA extraction process at 100 %OV. (a) Extracted Synchronous Average; (b) Normal probability test of the residual part; (c) Kolmogorov-Smirnov test of the residual part.

3 Interpolation process

In this case study, the interpolation of periodic phenomenon is performed between different levels of guide vane opening (noted as %OV). The guide vanes is a part of the turbine that controls the flow rate as a function

of vanes opening. We consider that the interpolation space in this study is different steady-operating conditions (different %OVs) of the hydroelectric turbine. The proposed interpolation process consists of three steps as shown in Figure 4.

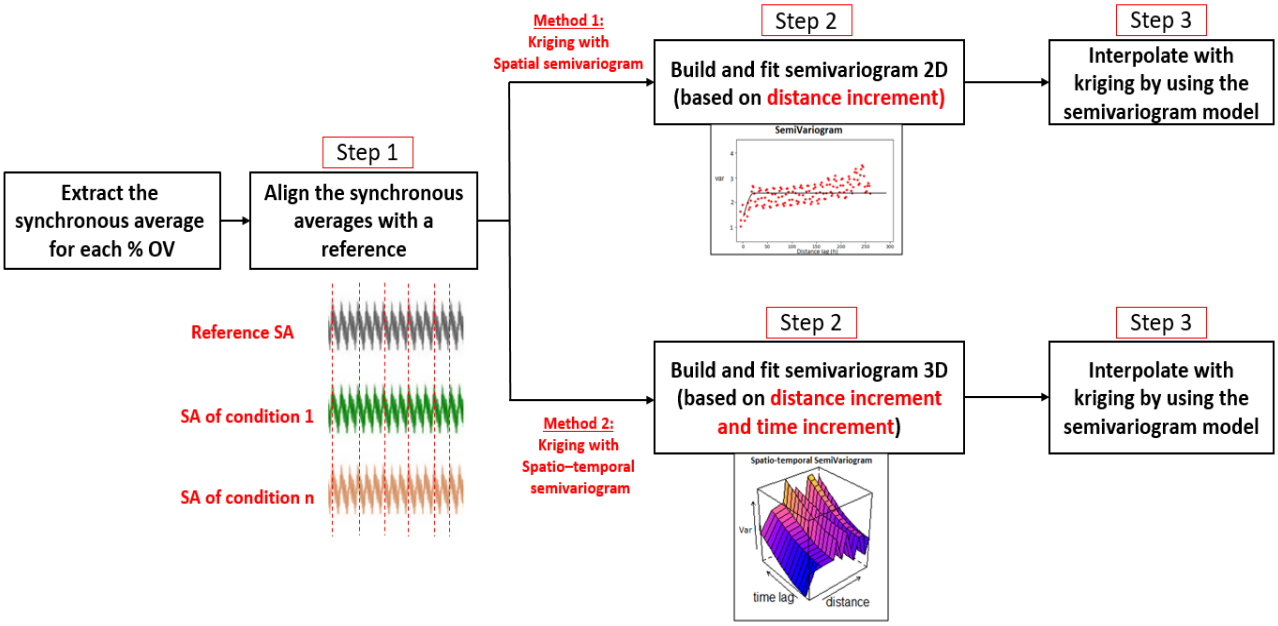


Figure 4 – Proposed interpolation process

Before applying kriging interpolation, the first step is to align the synchronous averages (noted as SA) of measured and known operating conditions (considered as known locations in the interpolation process). The SAs figure in the polar domain (figure 5) shows that there are small lags between values at 0° runner rotation of different %OVs. These lags do not affect the measurement results, they can cause however a systematic error for the interpolation process. For aligning these SAs, the first step is to choose a SA reference from which the other SAs can be aligned (e.g., the reference selected in this case study is the SA at 50 %OV (See figure 5)). The correlation coefficients, between value at 0° rotation of the reference with values around 0° rotation of other SA (from other conditions), is then determined until the best correlation for each condition is found. The final step is to re-arrange all SA by following the new values at 0° rotation corresponding to the best correlation found in the previous step (figure 5).

Next steps of interpolation process relate to the application of kriging interpolation. Kriging allows the estimations of missing values at given locations (from known observed data) by determining the "linear regression weights" which minimize the error variance. The error variance (also called the estimation variance [6]) is theoretically defined as a variance of the difference between the experimental strain value S and the interpolated strain value S^* at the same location u (Equation 3). The location u can be a spatial location, time location or both depending on the choice of kriging method.

$$\sigma_e(u)^2 = Var[S(u) - S^*(u)] = Var[S(u)] + Var[S^*(u)] - 2Cov[S(u), S^*(u)] \tag{3}$$

Each type of kriging has a different way for minimizing this estimation variance. In this study, the Ordinary Kriging (noted as OK) is applied for the interpolation process. According to interpolation results between different points on runner blade in [4], the OK gave the smallest estimation variances between three traditional linear kriging methods: Simple Kriging, Ordinary Kriging and Universal Kriging. The estimation function and the modeling equations for error variance for OK (detail in [4] and [6]) are expressed in equations (4) and (5).

$$S^*_{OK}(u) = \sum_{i=1}^{n} \lambda_i S(u_i) \tag{4}$$

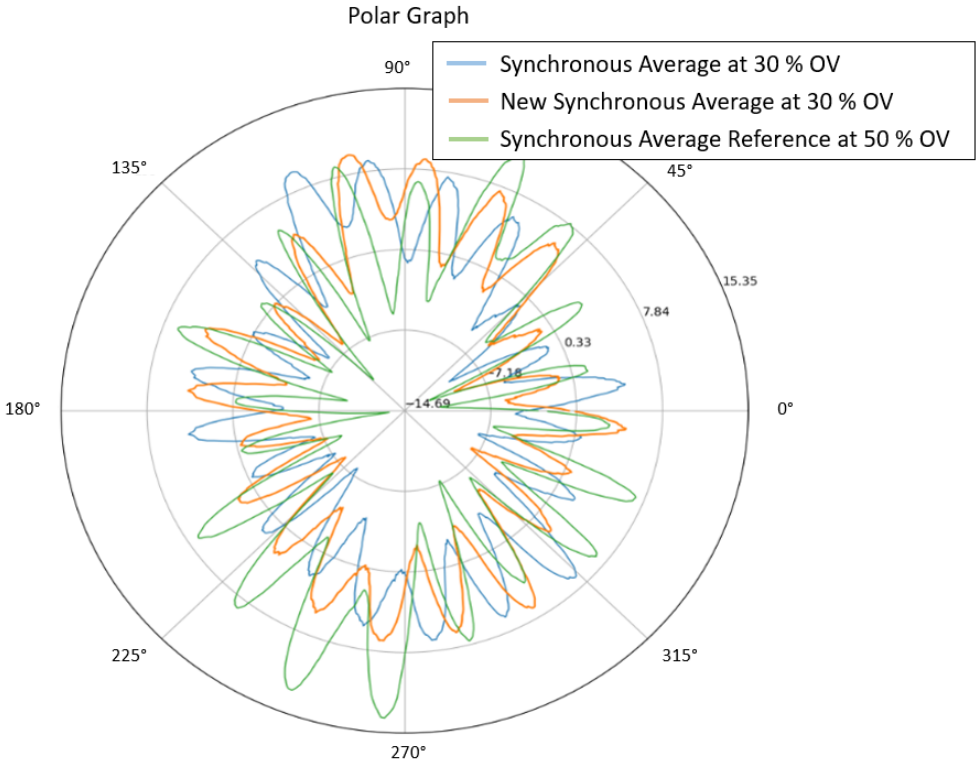


Figure 5 – Synchronous Average in polar field

$$\begin{cases} L_{OK}(\lambda_i,\mu) &= \sigma_e(u)^2 + 2\mu(\sum_{i=1}^{n}\lambda_i - 1) \\ \sigma_{OK}^2 &= \sum_{i=1}^{n}\lambda_i\gamma(u,u_i) - \mu \end{cases} \tag{5}$$

where the first equation of (5) is Lagrange function applied in OK to optimize the mean squared error, the Lagrange multiplier μ associates with the constraint $\sum_{i=1}^{n}\lambda_i = 1$, which is the OK constraint [6]. The second equation of (5) is the estimation variance obtained after minimizing the Lagrange function. Therefore, in OK, to minimize the estimation variance σ_{OK}^2, it is necessary to choose not only the weights λ_i, but also the Lagrange multiplier μ. γ is the semivariogram value (detail in the following paragraphs).

The minimization of estimation variance requires a model for the covariance between value of different locations (both interpolation and known locations). According to equation (5), these covariances are replaced by using the semivariogram values γ. In this paper, two types of semivariogram are applied: the Spatial Semivariogram and the Spatio-temporal Semivariogram. These two methods show the influences of two different types of information (operating conditions and temporal/angular information) on the kriging interpolation process.

Kriging interpolation with spatial semivariogram

To verify the intrinsic hypothesis that the finite variance of increment [$S(u)$ - $S(u+h)$] does not depend on location u, the variogram can be defined by the following function:

$$Var[S(u+h)-S(u)] = 2\gamma(h) = E[S(u+h)-S(u)]^2 \tag{6}$$

where h is the distance increment, γ is the semivariogram.

According to the equation (6), the spatial semivariogram can be defined as the half-variance between pairs of values and depends only on the distance increment h of these two locations. The spatial semivariogram is used for modeling spatial variability in kriging [6]. In this research, the spatial semivariogram is calculated by determining the dispersion of two different observations of a strain data set.

$$\hat{\gamma}(h) = \frac{1}{2N(h)}\sum_{(i,j)\in N(h)}[S(u_i)-S(u_j)]^2 \tag{7}$$

where $N(h)$ is number of paired points separated by a distance increment h, $S(u_i)$ and $S(u_j)$ are respective experimental strain values at positions u_i et u_j. The positions u_i or u_j are defined for each kriging method. The $\hat{\gamma}$ means the experimental semivariogram.

To use the semivariogram, a scatter plot (called as experimental semivariogram) has to be built using equation (7) with the measured data and then fitted with a numerical model by using the least square method. The experimental semivariogram cannot be directly used for kriging interpolation because it sometimes consists of "negative variance". So, the kriging interpolation is performed by using the numerical model of semivariogram for the covariances. The numerical model of semivariogram can be defined by sill, range and nugget effect parameter (if it is necessary). The range parameter is the distance from which there are no longer correlations (null covariance) in the data set. The sill parameter defines the average variance where the experimental variogram stabilizes and it is reached at the range level. The nugget effect parameter is sometimes added to the semivariogram model to represent a very short range variability in the data set and also the error of the experimental measurement. This nugget parameter is added into the sill parameter to obtain the total sill of the model. Some semivariogram models are presented in the following equations:

$$\begin{cases} \gamma_{exponential}(h) &= \text{sill}.[1 - exp(-\frac{3h}{\text{range}})] \\ \gamma_{Gaussian}(h) &= \text{sill}.[1 - exp(-\frac{(3h)^2}{\text{range}^2})] \\ \gamma_{hole-effect}(h) &= \text{sill}.[1 - cos(\frac{h}{\text{range}}.\pi)] \end{cases} \tag{8}$$

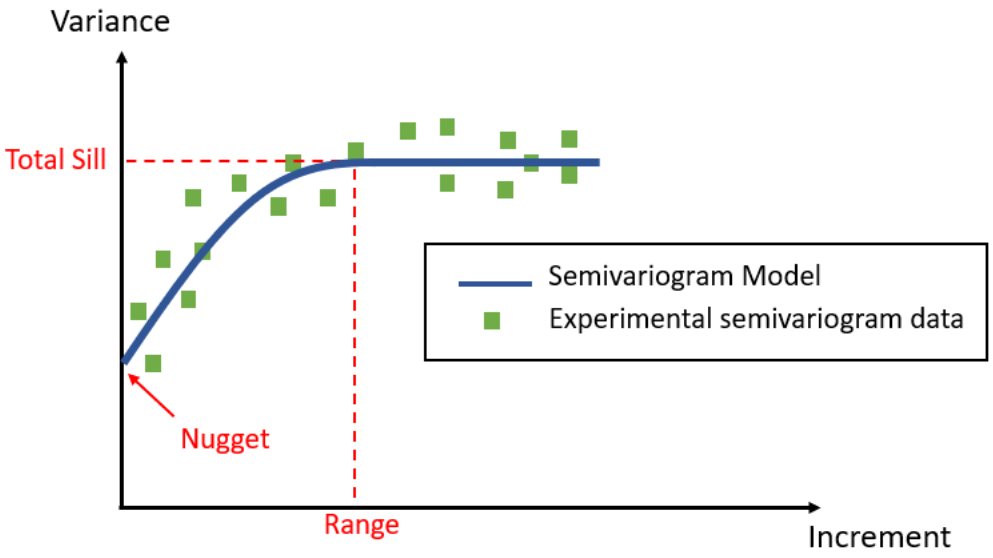


Figure 6 – Semivariogram model

For the case of spatial semivariogram, two different approaches are compared. For the first, an interpolation location u is defined by only one coordinate %OV. The influence of angular/temporal information is ignored. Thus, the interpolation is accomplished between different %OV for each degree of rotation. An extracted synchronous average of each %OV contains points values of a complete rotation of runner (0° to 360°). Therefore, this approach requires many semivariograms (number of semivariograms is equal to number of SA point values that is about 1000 points), leading to many interpolations to execute. To reduce the calculation time, only three simple models are used to fit the experimental semivariogram: Spherical, Exponential and Gaussian. Moreover, because of experimental limitations, the number of operating-conditions measured is also limited, which can affect interpolation results.

For the second approach of spatial semivariogram, the angular information θ (or time information) of measurement is taken into account as a "second spatial coordinate". An interpolation location is now defined by two coordinates, %OV and rotation angle of runner ($u_i = (\%\text{OV}_i, \theta_i)$). The distance between two locations in

this interpolation space is calculated as the Euclidian distance. However, these two coordinates do not have the same range to the spatial dimension, which can lead to an error of kriging interpolation. Thus, a normalization is required before building the semivariogram. Contrary to the first approach, this second one has more points in the scatter plot of semivariogram and requires only one semivariogram for interpolation. Therefore, to fit the experimental semivariogram, some nested variogram models are used in order to improve the performance of kriging interpolation. The nested model is a linear combination of several simple models like a combo presented in the result part: nugget model combines with a long-range Gaussian model and a short-range hole effect model (figure 7b). Since kriging uses the semivariogram model for covariances, the nested model helps to avoid the lack of information during the interpolation process.

Kriging interpolation with spatio-temporal semivariogram

In the second interpolation method, the spatio-temporal semivariogram is built with two independent dimensions: spatial and temporal/angular (respectively %OV and rotation angle of runner in this research). The half variance between two "locations" depends not only on distance increment h, but also on time/angle increment a (see equation 9). This type of semivariogram is often used in geostatistics for the case that the temporal dimension does not have the same range to the spatial one. Therefore, using this spatio-temporal semivariogram helps to independently observe the influences and the contributions of time (or angular field) to the interpolation process.

$$\hat{\gamma}(h,a) = \frac{1}{2N(h,a)} \sum_{(i,j)\in N(h,a)} [S(\%\mathrm{OV}_i, \theta_i) - S(\%\mathrm{OV}_j, \theta_j)]^2 \tag{9}$$

The experimental semivariogram is now a 3D scatter plot with an axis of half variance and two axes of distance and time increment (figure 10). For modeling covariance, semivariogram model is now presented by combining the spatial semivariogram and the temporal/angular semivariogram (there is also the joint semivariogram in some models[7].) with different mathematical operators. To fit the 3D experimental semivariogram, each semivariogram is first independently fitted by a simple model (the definition of model parameters, like sill and range, for the temporal/angular semivariogram model is similar to the spatial one presented in the first method) and they are then combined by using the spatio-temporal semivariogram model function. Separable model function [7], which is used in this case study, are presented in equation (10). This Separable model and the other spatio-temporal semivariogram models are available in code library Gstat [7]. This library is used to execute the calculation and the graphic observation of the spatio-temporal semivariogram in this study.

$$\begin{cases} C_{separable}(h,a) &= C_s(h).C_t(a) \\ \gamma_{separable}(h,a) &= \mathrm{sill}.(\bar{\gamma}_s(h) + \bar{\gamma}_t(a) - \bar{\gamma}_s(h)\bar{\gamma}_t(a)) \end{cases} \tag{10}$$

where $C_s(h)$ and $C_t(a)$ are respectively the spatial and temporal covariance, sill is the overall sill parameter, $\bar{\gamma}_s$ and $\bar{\gamma}_t$ are respectively spatial and temporal semivariogram [7].

4 Interpolation results

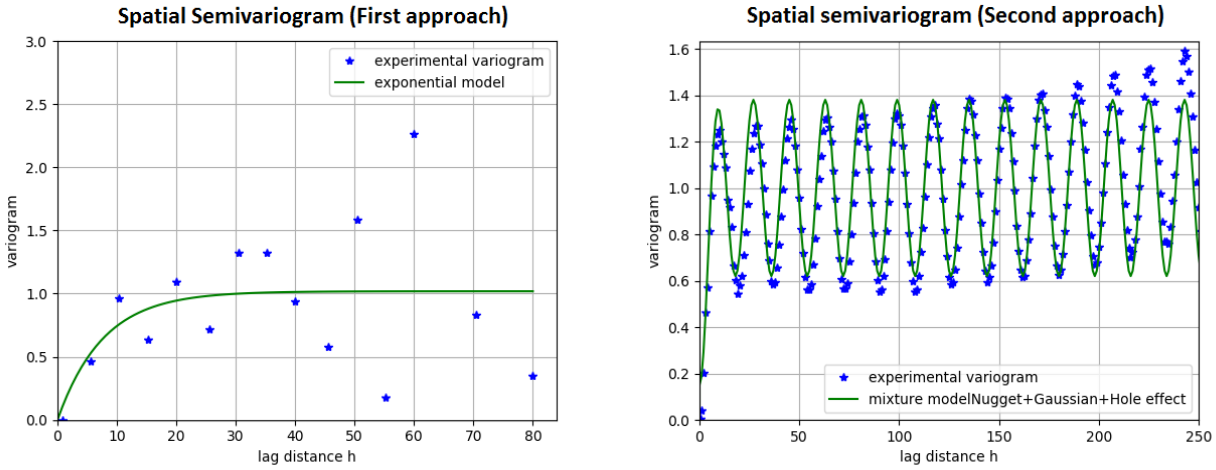
To deal with the limited experimental data, the interpolation is performed from a part of known observed operating conditions and the remaining ones are used for comparison. Input and Output data is shown in the Table 1. As mentioned in the previous section, the semivariogram model directly influences on the kriging interpolation quality. Therefore, only results of the best model are shown for each method. Using cross-validation technique helped to find out these best models. The principle of cross validation technique is to temporary remove one or several data and then to re-interpolate these data from the remaining data by kriging, this process is repeated several times (100-1000 times) to find at the end a set of correlation coefficients between real and interpolated data. Moreover, in the case of spatio-temporal semivariogram, the Mean Absolute Error (MSE) has been also calculated to verify the interpolation quality of model [7]. In the following, only two interpolation results at 20 %OV and 55 %OV are presented.

For the Spatial Semivariogram method, the interpolation results of the first approach are presented in figures 8a and 9a. These results show a low quality of interpolation because of many disturbances throughout

Input: SA(θ)	Semivariogram model		Output: SA(θ)
	Spatial Semivariogram	Spatio-temporal Semivariogram	
At 30 %OV, 40 %OV, 50 %OV, 60 %OV, 70 %OV, 90 %OV	First Approach: Exponential model Second Approach: Nugget effect + long-range gaussian model + short-range hole effect model	Separable model: +) Gaussian model (with Nugget effect) for the **spatial semivariogram** +) Gaussian model (with Nugget effect) for the **temporal semivariogram**	At 20 %OV, 45 %OV, 55 %OV, 80 %OV, 100 %OV

Table 1 – Information of input, ouput and semivariogram model used in this study paper

the angular axis (temporal axis). This observation highlights the importance of angular/temporal information in the kriging interpolation process. The high frequency noises observed in these figures confirm that the angular/temporal information must be considered in this interpolation study. Moreover, the limited experimental data for the semivariogram might also contribute to this. However, the interpolated signals are smoother with the second approach (figures 8b and 9b). This can be explained via a large number of information points taken by the nested semivariogram model (figure 7b).



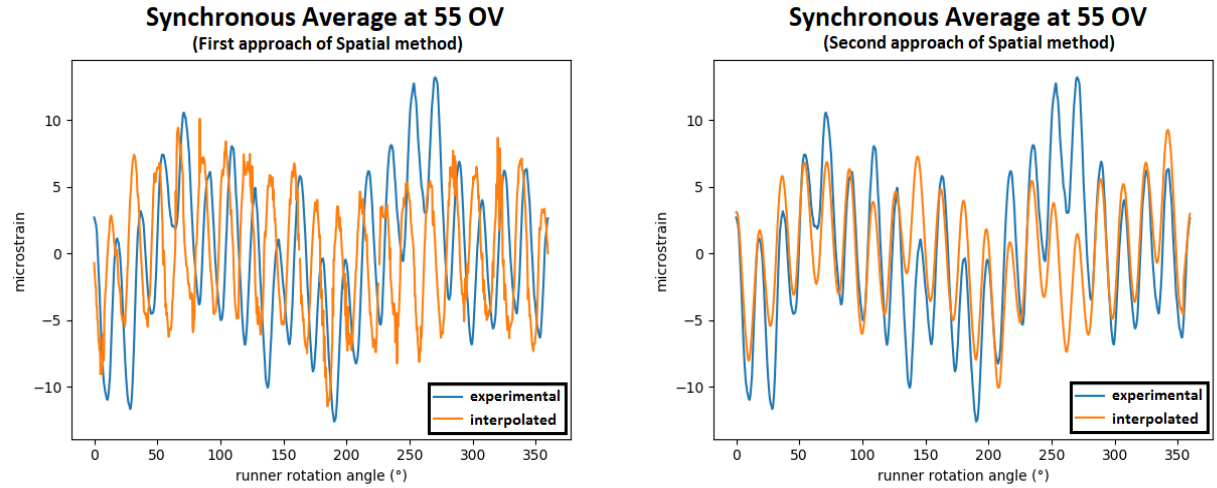
(a) (b)

Figure 7 – Experimental semivariogram and fitted model in the case of spatial method: (a) First approach; (b) Second approach

The interpolation results of the spatio-temporal semivariogram method are presented in figure 11. By graphically comparing to the spatial semivariogram method, the form and the amplitude level of the interpolated signal in this case have better correlations with the experimental signal. However, these results are not enough to confirm which process is better for the synchronous average interpolation, a proper validation is required.

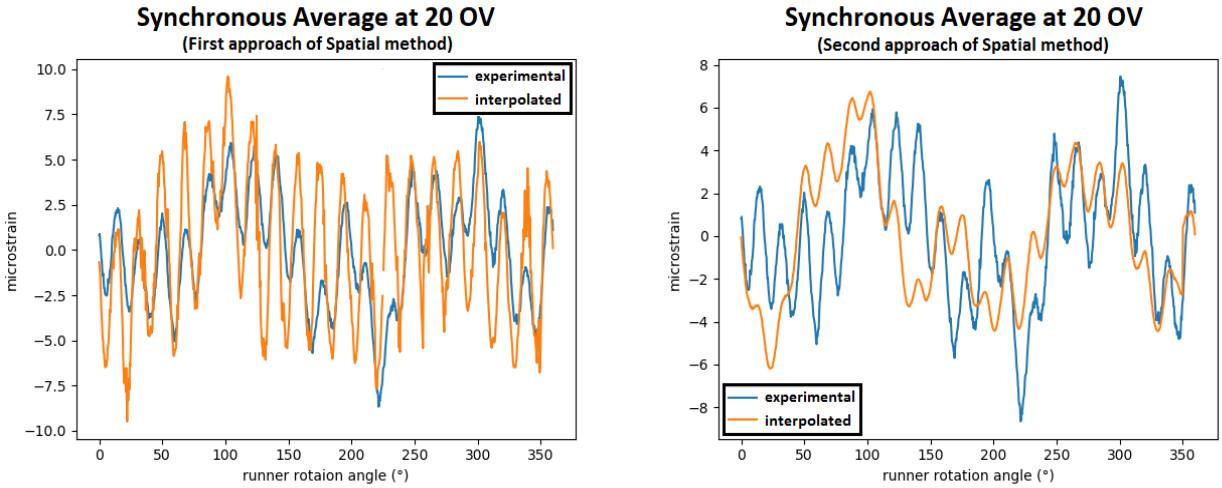
5 Discussion

In this paper, using the interpolation tool, a representation of the periodic phenomenon, varying through time and operating conditions, are generated. It is unable to foresee all variations of this phenomenon. With this kind of data, the generation of an interpolated signal that perfectly matches the experimental signal is not a requirement of this study. To verify the quality of interpolation process, the load spectrum, which is mainly used in the fatigue evaluation process, is presented. This graph, based on the Rainflow algorithm, allows the representation of the strain cycles contained in the signal as a function of cumulative number of cycles for fatigue. Figure 12 presents load spectrum curves of two kriging methods. The load spectrum curve found



(a) (b)

Figure 8 – Experimental and interpolated data at 55 %OV in the case of spatial method: (a) First approach; (b) Second approach

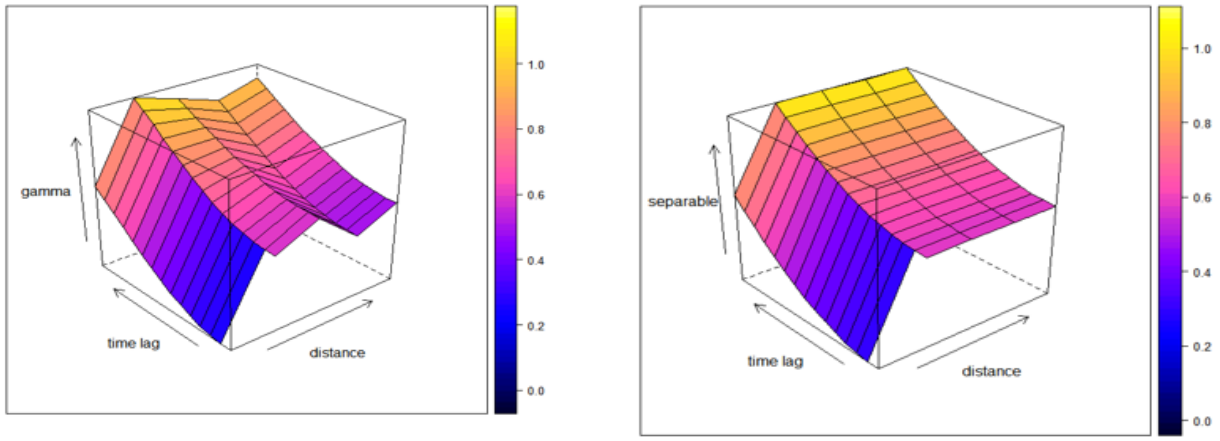


(a) (b)

Figure 9 – Experimental and interpolated data at 20 %OV in the case of spatial method: (a) First approach; (b) Second approach

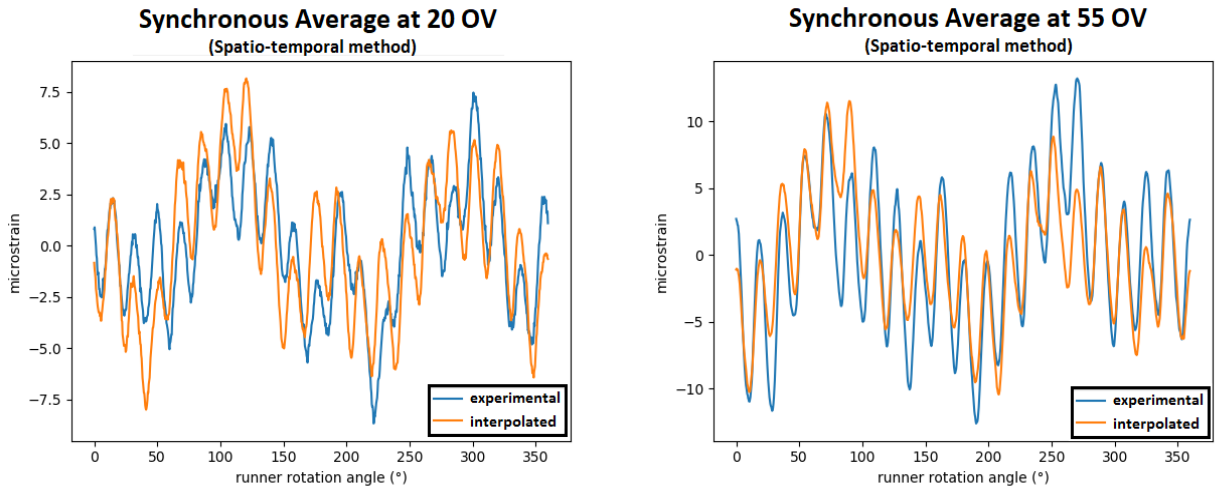
for the spatio-temporal method is clearly closer to the experimental value than for the spatial one. It can be observed from these load spectrum performances that the spatio-temporal semivariogram is better adapted this type of data.

The spatio-temporal semivariogram seems better but it is unable to conclude that this method satisfies our requirements. We need not only a unique interpolation result, but also a confidence interval. In this paper, the conditional stochastic simulation is used to generate a first assessment of the confidence interval. The conditional stochastic simulation is a statistic simulation process based on kriging and experimental data. The simulations are commonly used to correct the smoothing effect of kriging method [4][6]. The figure 13 presents 1200 conditional stochastic simulations at 20 %OV based on the Separable spatio-temporal model (see Table 1). The interpolation value and the experimental value mostly fall into the set of conditional simulations.



(a) (b)

Figure 10 – (a) Experimental spatio-temporal semivariogram; (b) Separable model used to fit the experimental semivariogram

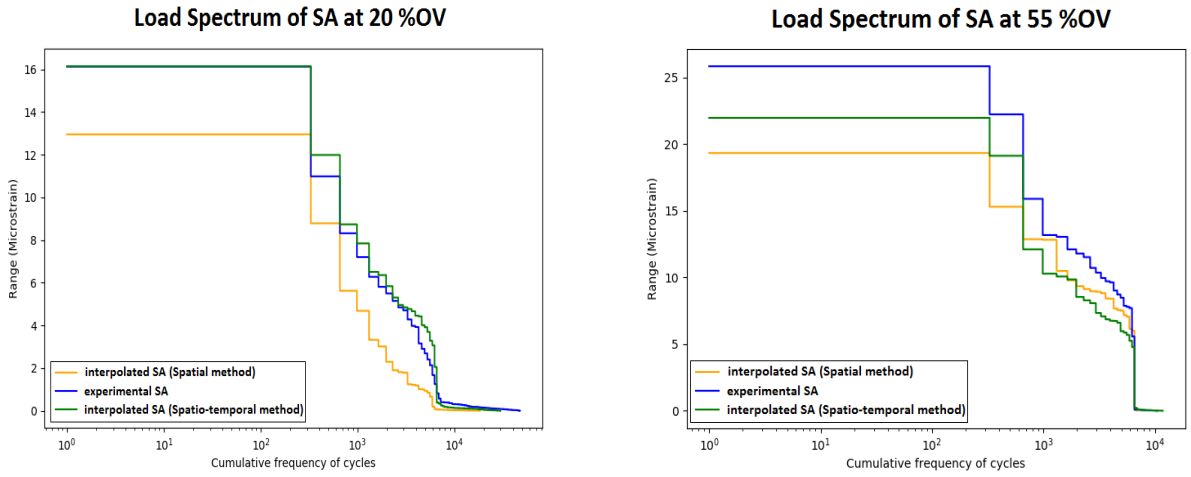


(a) (b)

Figure 11 – Experimental and interpolated data in the case spatio-temporal method: (a) 20 %OV; (b) 55 %OV

6 Conclusion

The periodic phenomenon has been extracted as the first order of cyclostationarity from the runner strain signal by applying the synchronous average operator. Even if the periodic part does not give all the needed information to assess the runner fatigue, the results in this paper set the first bases to build an interpolation tool for the whole signal. The synchronous average signals are considered as a function depending on the different %OV and the runner rotation angle in the interpolation process. The spatio-temporal kriging has a good interpolation performance for the synchronous average. Although the spatio-temporal kriging process might not be the most appropriate interpolator for other phenomena in signal, it can be a good start nonetheless.



(a) (b)

Figure 12 – Load spectrum comparison between original SA and interpolated SA of two kriging methods: (a) 20 %OV; (b) 55 %OV

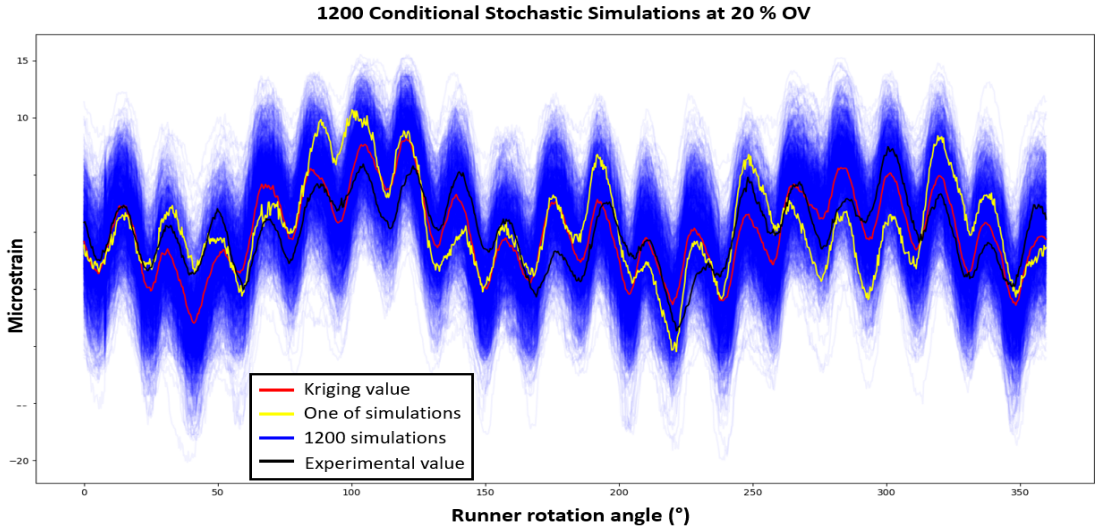


Figure 13 – Confidence interval by 1200 conditional stochastic simulations at 20 %OV

Acknowledgements

The author would like to thank the Hydro-Québec Research Institute (IREQ), the Andritz Hydro Canada LTD, the Insa de Lyon, the École de technologie supérieure Montréal (ÉTS) and the Mitacs-Accélération program for their support and financial contribution.

References

[1] M. Poirier , M. Gagnon, A. Tahan, A. Coutu, J. Chamberland-Lauzon, *Extrapolation of dynamic load behaviour on hydroelectric turbine blades with cyclostationary modelling*, Mechanical Systems and Signal Processing, 2017, Vol. 82, p. 193-205.

[2] M. Gagnon, A. Tahan, P. Bocher, D. Thibault, *The role of high cycle fatigue (HCF) onset in Francis runner reliability*, In IOP Conference Series: Earth and Environmental Science, 2012, IOP Publishing,p. 22005.

[3] J. Antoni, F. Bonnardot, A. Raad, M. El Badaoui, *Cyclostationary modelling of rotating machine vibration signals*, Mechanical Systems and Signal Processing, 2004, Vol. 18, p. 1285-1314.

[4] B. Firas, *Modélisation de la propagation des incertitudes des mesures sur l'aube d'une turbine hydraulique par Krigeage et simulations stochastiques*, Research paper for École de technologie supérieure Montréal,Québec,Canada, 2014.

[5] F. Bonnardot, *Comparaison entre les analyses angulaire et temporelle des signaux vibratoires de machines tournantes. Etude du concept de cyclostationnarité floue*, Traitement du signal et de l'image. Institut National Polytechnique de Grenoble - INPG, 2004. Français. (tel-00103067)

[6] C.V. Deutsch, A.G. Journel, *GSLIB: Geostatistical Software Library and User's guide* , Oxford University Press, 1995, 8e edition.

[7] B. Gräler, E. Pebesma, G. Heuvelink, *Spatio-temporal interpolation using Gstat*, The R Journal, 2016, Vol. 8:1, p. 204 - 218.

[8] J. Antoni, *Cyclostationarity by examples*, Mechanical Systems and Signal Processing 23, 2009, p. 987 - 1036.