Correlation between numerical simulations and measurements to assess uncertainties: a case study on a hydroelectric runner

M Gagnon¹, A Immarigeon², J Nicolle¹ and J-F Morissette¹

¹ Institut de recherche d’Hydro-Québec (IREQ), Varennes, QC, J3X 1S1, Canada
² Hydro-Québec, Montréal, QC, H2Z 1A4, Canada

gagnon.martin@ireq.ca

Abstract. For fatigue reliability analysis, we need an estimate of the structure response and uncertainty at critical locations where often direct observation from a measurement campaign cannot be obtained. At these critical locations, two types of information are required: static amplitudes and dynamic ranges. In this work, we assume a linear correlation between simulated and measured values to estimate the information needed at critical locations on the runner blades. The proposed methodology is illustrated using the data from a six-blade propeller runner. The methodology enables one to easily transpose information from observed values to a critical location and assess uncertainty for fatigue reliability analysis.

1. Introduction

One of the goals of numerical simulation is to accurately predict the response of a system prior to empirical evidence from direct observations or representative experiments. To use these results, decision makers need to be confident that the results obtained have enough precision to support their actions. This confidence can often only be gained a posteriori through observations; hence the need for validation using observed data [1]. However, even if the numerical model is valid within the precision needed at the measured locations, for fatigue analysis [2], we still need to estimate the response and uncertainty levels at critical locations where we do not have experimental data.

This paper presents how to perform such estimation for static and dynamic strains in a six-blade hydraulic turbine propeller by combining Computational Fluid Dynamic (CFD), Finite Element Analysis (FEA) and experimental results. The fatigue analysis requires that the static and dynamic responses be estimated for every steady operating condition of interest. Figure 1 illustrates how the response is decomposed in static amplitude and dynamic range. We define the static amplitude as the mean value of the response and the dynamic range corresponds to the maximum peak-to-peak interval over the considered time interval in steady operating condition.
These values are estimated using numerical models during the design phase of the turbine and need to be validated during commissioning through field measurements. To validate the stress amplitudes used in fatigue analysis, a first difficulty resides in the choice of measurement locations. It is often impossible to position a strain gauge exactly where the maximal values should be observed. Hence, a transposition method is needed to estimate expected values and uncertainty levels at these desired critical locations. In this study, to transpose the information at the critical location, we assume that simulation results and measurements are linearly correlated which enables one to estimate both the expected value and its uncertainty.

The study is structured as follows: first, the proposed correlation methodology is presented and then our case study is detailed by reviewing numerical simulation results followed by observed data from the measurement campaign. Finally, the transposition results and their implications are discussed. The unit studied is a six-blade propeller turbine that has previously been referred to by the authors in others studies [3-5].

2. Correlation methodology

Numerical simulations are an approximation of the reality of interest. In our case, this approximation needs to answer the following engineering question: are the maximum values at critical locations (i.e. hotspots) acceptable to warrant reliable operation during the expected life of the runner? This should hold true if numerical simulation and measurements are well correlated. In our case, a linear correlation between numerical results and measurements is assumed. This statement implies the following model function:

$$y = \beta_0 + \beta_1 x$$

where $x$ is the simulation result and $y$ is the measured value which are the two variables of interest. Ideally, one would expect an intercept $\beta_0 = 0$ and a slope $\beta_1 = 1$. However, because we expect that the model might have bias either constant $\beta_0 \neq 0$ or non-constant $\beta_1 \neq 1$, the regression problem becomes:

$$y_i = f_i(x_i + \delta_i; \beta) - \epsilon_i$$

where the $y_i$ are a function of the $x_i$ and the unknown parameters $\beta$ subjected to error $\delta_i$ and $\epsilon_i$. This means that we expect error $\epsilon_i$ in measured values and $\delta_i$ in the simulation results. Such problem can be solved using Orthogonal Distance Regression (ODR) using the following minimization:

$$\min_{\beta, \delta, \epsilon} \sum_{i=1}^{n}(\epsilon_i^2 + \delta_i^2)$$

where $n$ is the number of $x_i$ and $y_i$ available. Furthermore, to account for unequal uncertainty across the $x_i$ and $y_i$, weights in the form of $w_{\epsilon_i}$ and $w_{\delta_i}$ are introduced as follows:
This enables us to account simultaneously for error and uncertainty in both the numerical results and measured values. In this study, we used the software package ODRPACK [6] to estimate the model parameters $\beta_i$ using weighted orthogonal distance regression. Afterwards, the confidence interval was constructed using the $\beta_i$, the covariance matrix $C$ and the appropriate Student’s distribution values [7]. The confidence interval was obtained as follows:

$$\hat{y} \pm t_{\alpha/2,\nu} \sqrt{\chi^2_{0} \sum_{j=0}^{n} \sum_{k=0}^{n} \frac{\partial f}{\partial \beta_j} \frac{\partial f}{\partial \beta_k} C_{jk}}$$

(5)

Where $\chi^2_{0}$ is the reduced Chi square and $t_{\alpha/2,\nu}$ the Student’s t-distribution with $\nu$ degree of freedom at $\alpha$ significance level. For any given $x$ value, this interval is interpreted as the probability that the true value of the function should be within its limits $(1 - \alpha)\%$ of the time.

An example using randomly generated values is presented in Figure 2. In this example, randomly generated error and uncertainty for values between 0 and 5 have been used. Using the confidence interval around the estimated linear model shown in blue area, we can obtain, for any interval representing the uncertainty interval for the corresponding $y$ value. Notice, the black line representing the relation $y = x$, which was included to help visualize the biases in the estimated linear model. In this example, we observe that for $x = 6 \pm 0.25$, we obtain the interval $[5.18, 7.32]$ in red given a 95% confidence interval. This approach can be used to estimate the expected uncertainty interval for any values as long as we accept the assumptions made initially by using a linear model to correlate both datasets while neglecting the spatial relation between values. Due to the limited number of sensors, spatial information is difficult to take into account using spatial correlation. Accounting for spatial correlation would give a better assessment of uncertainty by using distances between measured locations and the location of interest.

3. Numerical simulation results

Four different operating conditions for this propeller turbine have been simulated with the use of commercial CFD solver Ansys CFX. They consist in four stable regimes, namely: speed-no-load (SNL), part load (60% opening), best efficiency (85% opening), and full load (100% opening).

Two configuration setups were used for the different simulations (Figure 3). A complete machine setup with the entire spiral casing, distributor, runner and draft tube was used for three operating points at stable regime with the advanced SAS (Scale-Adaptive Simulation) turbulence model to get as much dynamic content as possible. The speed-no-load condition was modelled without the spiral casing but with the entire distributor, air injection, and also used the SAS turbulence model.
Those simulations enabled us to get the pressure loads. FEA structural analyses were then conducted in Ansys Mechanical with the full-runner geometry meshed with 1.34M nodes, refined at the strain gauges locations shown in Figure 4. Quasi-static analyses were used with two pressure loads per operating point to determine the dynamic range. Those loads correspond to the instants with maximum and minimal recorded torque on a specific blade. Dynamic amplification was not simulated because it was not detected during measurements. Static values were simply the arithmetic mean of the maximum and minimum values for each location and operating point.

![Figure 3. CFD setup for pressure loading. (a) Full machine setup for steady operating condition, (b) 360 model used for speed-no-load.](image)

![Figure 4. FEA model for structural analysis, the trailing edge measurements location E, F and G are visible.](image)

In order to account for uncertainties, an arbitrary 10% of the maximum simulated value for a given operating condition was used and applied at every location on the blade. The same methodology was used for both the static and dynamic estimates.

4. Field measurement data
The runner was instrumented with strain gauges at seven locations replicated on two different blades. Figure 5a presents the location of the strain gauges and an example of the measured data at location A for a complete start-to-stop cycle is presented in Figure 5b. The locations A, B, C and D were
instrumented with strain gauges rosettes while E, F and G were monitored with unidirectional strain gauges.

![Figure 5. Strain gauge measurement. (a) Location of strain gauges. (b) Measured data at location A.](image)

The static amplitude and dynamic range for maximum wicket gate opening (full load), 85% wicket gate opening (best efficiency), 60% wicket gate opening (part-load) and speed-no-load (SNL) conditions were obtained from the measured data. The uncertainty intervals used to weight the static amplitudes include:
- Differences between the zeros before and after the considered operating condition
- Differences between similar measured operating conditions
- Unaccounted uncertainty sources with 10% of the maximum static stress

For dynamic ranges, the uncertainty intervals used to weight the data include:
- Differences between similar measured operating conditions
- Unaccounted uncertainty sources with 10% of the maximum dynamic range

The zeros were verified before and after each measurement record while the unit was at rest. Notice that the unaccounted uncertainty sources include more than only variations in parameters like geometry, material properties, loading, etc. They also include epistemic uncertainty due to lack of knowledge and modeling assumptions.

5. Correlation and expected maximum values

The correlation results and uncertainty intervals expected at the location of the maximum deformation on the model (which is in the vicinity of measurement locations A and B in Figure 5a) are presented in Figure 6 and 7. In each figure, 95% confidence bands are presented in blue and the interval at the maximum location is shown in red. The static amplitude results showed in Figure 6 exhibit good correlations with small confidence bands for all operating conditions except SNL. However, it can be noticed that for the SNL condition all the values are grouped together and have relatively small values which might explain the wide confidence band observed when projected to higher value. This also highlights the difficulty in choosing proper measurement locations on a runner in order to obtain useful data for a wide range of operating conditions.
Figure 6. Correlation results for static amplitude.

For the dynamic range, Figure 7 shows that none of the conditions studied exhibit good linear correlation. For every studied condition, we observe large constant bias $\beta_0$ and non-constant bias $\beta_1$. Furthermore, there is a large dispersion that translates into a large 95% confidence band. Hence, larger uncertainty intervals are estimated at the maximum stress location. These large biases and dispersion indicate that our numerical model is not well representative of the measured values which might warrant for larger unaccounted uncertainty sources. Nonetheless, as such, this generates large uncertainty intervals for maximum dynamic stresses that could be decreased by improving the numerical model.
6. Discussions
The principal advantage of the proposed methodology is that it can bring added value to numerical analysis and measurements by accounting for errors and sources of uncertainty in the various inputs to quantify overall uncertainties. The outcomes of the regression are estimates of the intercept $\beta_0$, slope $\beta_1$, and confidence bands to account for the dispersion. Using these, we can obtain estimates for the expected value and uncertainty at any location. In this case, the value obtained shows good overall behavior as it properly rises with uncertainty and dispersion. However, since we neglect spatial correlation between the locations of interest, uncertainty might not be properly estimated. Also, we have observed that additional care is required if we want to use the correlation results for validation because rather than having just one metric to consider, we now have three (constant bias $\beta_0$, non-constant bias $\beta_1$, and dispersion). Furthermore, linear correlation results for different runners cannot be directly compared because the number of measured locations and their locations are often different. Nonetheless, the proposed methodology proves to be a good qualitative tool to assess the validity of numerical model.

7. Conclusions
The objective of this paper was to propose a methodology to correlate the measurement data and simulation results in order to estimate the critical values on runner blades. With the proposed methodology, we have shown that it is possible to account for errors in measured data and in simulation results while simultaneously weighting the data to account for uncertainty using weighted orthogonal distance regression. The results of our study case show better correlations and lower
uncertainty for static amplitude than for dynamic fluctuations. These results were to be expected since dynamic ranges are more difficult to obtain using numerical simulation. The methodology proves to be a good qualitative tool to assess the validity of these simulation results and helps highlight the model deficiencies. Notice that having three metrics to take into account at the same time (constant bias $\beta_0$, non-constant bias $\beta_1$ and dispersion) makes the correlation results not well suited for quantitative model comparison [8]. Nonetheless, the proposed methodology, with its simplicity, enables the analyst to easily transpose the observed value to the critical location and assess uncertainty for fatigue reliability analysis. Other studies are needed to obtain a methodology that could at the same time be used for quantitative model comparison, account for spatial correlation and estimate expected values with uncertainties at critical locations.

References

[1] ASME 2006 Guide for Verification and Validation in Computational Solid Mechanics vol PTC 60/V&V 10
[2] Gagnon M, Tahan A, Bocher P and 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
[3] Gagnon M, Jobidon N, Lawrence M and Larouche D 2014 Optimization of turbine startup: Some experimental results from a propeller runner. 27th IAHR symposium on hydraulic machinery and systems, 22-26 of September 2014, Montreal, Canada
[4] Diagne I, Gagnon M and Tahan A 2016 Modeling the dynamic behavior of turbine runner blades during transients using indirect measurements 28th IAHR symposium on hydraulic machinery and systems, 4-8 of July 2016, Grenoble, France
[5] J Nicolle and J-F Morissette 2016 Simulation of air admission in a propeller hydroturbine during transient events 28th IAHR symposium on hydraulic machinery and systems, 4-8 of July 2016, Grenoble, France
[6] Boggs P T, Byrd R H, Rogers J E and Schnabel R B 1992 User’s Reference Guide for ODRPACK Version 2.01 Software for Weighted Orthogonal Distance Regression National Institute of Standards and Technology Gaithersburg, MD 20899
[7] Wolberg J 2006 Data Analysis Using the Method of Least Squares: Extracting the Most Information from Experiments Springer-Verlag Berlin Heidelberg
[8] Chatenet Q, Tahan A, Gagnon M and Chamberland-Lauzon J 2016 Numerical model validation using experimental data: Application of the area metric on a Francis runner 28th IAHR symposium on hydraulic machinery and systems, 4-8 of July 2016, Grenoble, France