HAL
open science

# Improving the monitoring indicators of a variable speed wind turbine using support vector regression

Hugo André, Flavien Allemand, Ilyes Khelf, Adeline Bourdon, Didier Rémond

## To cite this version:

Hugo André, Flavien Allemand, Ilyes Khelf, Adeline Bourdon, Didier Rémond. Improving the monitoring indicators of a variable speed wind turbine using support vector regression. Applied Acoustics, 2020, 166, pp.107350. 10.1016/j.apacoust.2020.107350 . hal-03178511

**HAL Id: hal-03178511**

**https://hal.science/hal-03178511v1**

Submitted on 23 Mar 2021

**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.

# Improving the monitoring indicators of a variable speed wind turbine using support vector regression.

Hugo ANDRÉ[a,*], Flavien ALLEMAND[b], Ilyes KHELF[c], Adeline BOURDON[d], Didier RÉMOND[d]

[a] *Univ Lyon, UJM St-Etienne, LASPI, EA3059, F-42023 Saint-Etienne, France*
[b] *ENGIE GREEN, France*
[c] *Univ Lyon, INSA Lyon, LVA EA677, F-69621 Villeurbanne, France*
[d] *Univ Lyon, INSA Lyon, LaMCoS, CNRS UMR5259, 18, rue des Sciences, F69621*

**Abstract**

For over a decade, most wind turbines have worked by adapting their rotation speed to that of the wind. This operating method, now widely used, allows optimal tip speed ratio to be achieved whatever the weather conditions, and in fact produces much better output than stall controlled turbines, particularly in calm weather conditions. However, this improvement means that monitoring systems are required to adapt to constant macroscopic variations in load and speed. In addition, these non-stationary operating conditions make it difficult to undertake machine diagnostics over the long term, due to the fact that the operating conditions in which successive indicators are obtained will almost never be the same. The scientific community has, in many respects, proved the usefulness of regression analysis of these indicators in relation to properly selected variables. The focus of this paper is on regression methods based on machine learning tools, which are becoming more and more popular. The difficulty lies in designing a robust self-adaptive method for estimating the statistical behaviour of an indicator in relation to operating conditions. Indeed, the concern is that indicators may obey disparate and unpredictable multivariate laws: there are many complications which make it difficult to use linear regression tools. Kernel machines, used in this paper as a robust and efficient way of normalising indicators, have proved to be capable of greatly improving a monitoring system's diagnostic capabilities. The demonstration is based on a practical example: monitoring a bearing defect by analysing the instantaneous angular speed of the wind turbine shaft line. As this defect can only be detected under certain operating conditions  a priori unknown  the chosen example will be particularly effective in highlighting the usefulness of such an approach.

*Corresponding author
*Email address:* `hugo.andre@univ-st-etienne.fr` (Hugo ANDRÉ)

*Keywords:* parametrization, kernel machine, support vector machine, non-stationary conditions, instantaneous angular speed.

---

**Nomenclature**

| | |
|---|---|
| $CRPS$ | Continuous ranked probability score |
| $IAS$ | Instantaneous angular speed |
| $IRLS$ | Iteratively reweighted least squares |
| $LOOHKRR$ | Leave one out heteroscedastic kernel ridge regression |
| $PRESS$ | Predicted residual error sum of squares |
| $RBF$ | Radial basis function |
| $SVR$ | Support vector regression |
| $SVM$ | Support vector machine |

## 1. Introduction

Instantaneous Angular Speed (IAS) has turned out to be an original and promising tool for monitoring rotating machines. To date, unlike traditional vibration monitoring, there is not a great deal of feedback from those with first-hand experience, and this tool is still largely confined to testing laboratories. IAS obtained through the elapsed-time method has the advantage of being angularly discretised by its very nature. Most cyclostationary phenomena are therefore concentrated in a single frequency channel, regardless of variations in the rotation speed of the mechanism being monitored. This makes it a particularly appropriate tool for machines operating in non-stationary conditions, such as wind turbines. Other specificities of this method, such as the quantification error or the inevitable angle aliasing, are set out in detail in [1]. Certain processing tools have been specially adapted for use with this method in order to observe extremely low-energy cyclical phenomena under conditions with highly variable speeds [2][3][4]. A mechanical model has also been developed which provides a digital analysis of how the beginnings of a bearing defect can be detected by an encoder on the other side of a shaft line [5].

However, one aspect which makes it difficult to monitor a rotating machine is the simple fact that the sensors are activated not only by the components being monitored. In most cases, measurements are disrupted by noise which drowns out the indicator revealing the condition of the component being monitored. Although the noise should remain constant as long as operating conditions remain stable, this is no longer the case as soon as there are variations in the rotational speed of the machine being monitored. Macroscopic load variations have an effect on the rotating machine's modal response, and macroscopic variations in speed will alter the position of characteristic frequencies in relation to the vibrational modes of the rotating machine.

Furthermore, it may not be possible to observe the defect under certain as yet unknown operating conditions. For example, a bearing's characteristic frequencies are known to be dependent on the angle of the ball bearings' contact with the inner and outer rings of the bearing. This angle is notably dependent upon the ratio between the axial and radial loads placing pressure on the bearing. If one of the rings has the beginnings of a localised flaw, the ball bearing will only roll over the imperfection if the contact angle allows it to do so. The fault can therefore only be observed under certain operating conditions. These operating conditions are a priori different from those causing the defect characteristic frequencies to match the vibrational modes of the machine being monitored.

The two previous paragraphs explain why simple order analysis alone cannot solve the issue. Although an angular description of the signal does allow cyclical phenomena to be contained within thin frequency intervals, it does not enable one to differentiate between those changes in the amplitude of the indicator which are linked to the development of a fault and those linked to a match with a vibrational mode or, more generally, those linked to a match with a noise whose statistical characteristics depend on operating conditions.

In structural health monitoring, a method using normalised parameters has been used to reduce the effect of temperature on the modal behaviour of structures [6],[7]. In the more specialised area of rotating machine diagnostics, Stander et al's ground-breaking work uses a straightforward method for load normalisation [8]. As this method can only be applied in cases where there is strictly linear dependence between the indicator being monitored and the load, it cannot be used more widely, and only provides a partial response to the issue previously outlined. McBain and Timusk then continued with Worden et al's [9] work, by subdividing non-stationary signals into stationary sub-signals, then categorising them into bins representing each operating condition [10]. Statistical parameters  means and variances  are calculated for each subdivision so that a statistical model can be assigned to each bin. Finally, these statistical parameters can be interpolated in order to calculate their values for operating conditions, without the need to use training data for every operating conditions.

This last paper tackles the core of the problem: the number of points needed to build a model is directly influencing the computation cost along with the time needed to gather enough training data. Especially in the domain of wind turbine monitoring, where some operating conditions are rarely met. Another solution has been proposed two decades ago, by Saunders et al, and does not require to fill a priori defined bins[11]. Based on fundamental works of Vapnik et al [12], they circumvented this limitation by introducing kernel functions into ridge regression algorithms. The commonly named kernel trick approach takes inspiration from methods based on the use of kernels, first developed to address support vector machine (SVM) classification problems [13] [14].

Whilst the solution put forward in the present paper is influenced by this research, the objective of the first section is to provide a summary of the entire procedure. The second section present the measurement campaign and describe

in detail the influence model configuration parameters have on the efficiency of the monitoring system.

## 2. Material and methods

The method set out in this paper has two objectives: firstly, to estimate the statistical parameters of indicators in relation to operating conditions, and secondly to use these parameters to normalise the indicators. The first stage could be described as regression: overall statistical behaviour is estimated from data constituting a baseline. The second stage could be described as standardisation: the indicators are reduced in such a way as to follow a normal distribution centred on 0 with unit standard deviation.

### *2.1. Machine learning*

Since the advent of data science, which is increasingly attracting attention from across the scientific community, machine learning protocol has been used to address regression problems: in particular regarding our hopes of being able to standardise monitoring indicators, which depends on building a model from indicators obtained during a learning phase. The indicators obtained during this period are compared with the preestablished model, and the only aspect which is of interest from this point onwards is the discrepancy between them. Of course, the learning phase is based on the assumption that the component being monitored is in good working order. The goal is therefore first to estimate a model to define $H_0$ using a supervised learning approach, before testing future data against $H_0$.

Machine Learning is used here to carry out regression analysis rather than classification analysis. This analysis can be carried out using parametric methods such as polynomial coefficients, where hypotheses initially formulated about the form of the model give access to a set of parameters which have yet to be estimated. Non-parametric methods can be used instead in adverse conditions where theories regarding the form of the model cannot be formulated. This second group of methods is also more appropriate in cases where samples are not uniformly distributed, and/or where the number of indicators needed to estimate the model is too great.

The objective of non parametric regression is to estimate the function $f$ describing at best the law followed by the monitored indicator, which is symbolized by the random variable $I$ :

$$I = f(\theta(\mathbf{x})) = f(H(\mathbf{x}), \mu(\mathbf{x}), \sigma(\mathbf{x})) \tag{1}$$

$\theta$ is a vector containing all parameters describing the evolution of the observed indicator $I$. This vector contains three parameters:

- $H$: the actual health status of the monitored element.
- $\mu$: the noise mean value

- $\sigma$: the noise standard deviation

All of them depend on the vector $\mathbf{x}$, whose coordinates describe every operating conditions. $\mathbf{x}$ is defined in the operating condition space $\mathscr{X}$, whose dimension depends on the number of influential operating conditions. The dependency of $H$ illustrates the conditional visibility of the fault, which is not always excited for all operating conditions. Under $H_0$, it has no influence over the indicator value, which then is defined such that: $I = f(\mu(\mathbf{x}), \sigma(\mathbf{x}))$.

The normalized indicator $I^{(n)}$ proposed in this paper is defined such that:

$$I^{(n)}(H(\mathbf{x})) = \frac{I - \mu(\mathbf{x})}{\sigma(\mathbf{x})} \tag{2}$$

### *2.2. Kernel machines*

Of all the available algorithms for establishing a nonparametric regression model, kernel machines appear to have the advantage of being both simple and effective [15][16]. Kernel machines have primarily been used in regression problems [17], nevertheless adjusting them to ridge regression problems has allowed the weaknesses in parametric models to be addressed without making the problems themselves significantly more complex. These tools effectively allow linear models to be set up with non-linear input data. They can also work with sparse estimation: the statistical parameters can be estimated using just enough components to form a basis in the subspace in question, provided that the presumed continuity has been verified.

#### *2.2.1. continuity hypothesis*

The respect of the continuity hypothesis is mandatory, and this constitutes a clear limit to the application of the method put forward in this paper.

A first example is proposed to picture the problem: let's consider a lift operating either with a given weight or without any. The load is, in this case, not defined continuously. Since the continuity hypothesis is not respected, the method presented here cannot estimate the behaviour of the loaded lift using the empty lift measurements. two different models must be built for the loaded and the empty cases respectively using only the loaded and empty case measurements.

The validity of the continuity hypothesis is now discussed in the case of wind turbines. At first, some technical terms must be introduced: the nacelle is the name of the frame carrying the gearbox and the generator studied in section 3.1. The turbine rotor axe is parallel to the main axe of the nacelle which consequently needs to track the wind direction. For this purpose, the yaw system is the name of the geared bearings that ensures reliable and precise tracking and following of the wind direction. This slewing bearing is indirectly

actuated by a wind vane to cope with the variations of wind direction. This yaw motor activates intermittently and therefore is one of the systems invalidating the continuity hypothesis. This hypothesis is maintained as long as every influential operating conditions vary in a continuous manner. Ancillary components operating intermittently, such as the yaw motor, are expected to affect the monitoring sensors and indicators. The method set out in this paper must be combined with a multi-label classification algorithm, in order to produce sub-models for all discontinuous operating conditions.

*2.2.2. Non linear ridge regression*

Assuming that the measurements respect the continuity hypothesis, each indicator can be used to estimate the function $f$, under $H_0$, for any operating conditions experienced by the machine. Of course, the influence of an indicator obtained under similar operating conditions will be larger than if it was obtained under completely different operating conditions. This weighting is operated by the kernel, and gives access to a non linear estimation of the function $f$ for any operating conditions. The objective function minimized in constructing a kernel ridge regression model is given by:

$$L(\mathbf{w}, b) = \frac{1}{2}\gamma||\mathbf{w}||^2 + \frac{1}{n}\sum_{i=1}^{n}\left(I_i - \langle \mathbf{w}, \phi(\mathbf{x}_i)\rangle - b\right)^2 \tag{3}$$

with $\mathbf{x}_i$ the $i^{th}$ indicator operating conditions, $I_i$ the indicator measured value, $\gamma$ the regularization term, $\mathbf{w}$ and $b$ are the objective coefficients of the regression. $\phi$ is a black box function that carries the problem in a high dimensional feature space: $\mathscr{F}(\phi : \mathscr{X} \rightarrow \mathscr{F})$. This function is used to introduce the non linear kernel function $K : \mathscr{X} \cdot \mathscr{X} \rightarrow \mathbb{R} \mid K(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j)\rangle$ but is not precisely developed. It suffices to know $K$ rather than $\phi$ explicitly which allows us to restate the support vector optimization problem. Kernel functions represent dot products in a feature space, which allows the algorithms to be used in a feature space without having to carry out computations within that space.

By introducing Lagrange multipliers $\boldsymbol{\alpha}$, the output of the least-square support vector machine is then given by the kernel expansion:

$$f(\mathbf{x}) = \sum_{i=1}^{n} \alpha_i K(\mathbf{x}, \mathbf{x}_i) + b \tag{4}$$

*2.2.3. Gaussian RBF kernel*

The choice of kernel function is very important and different kernel leads to different Support Vector Regression (SVR) algorithm. Commonly used kernel functions include the linear kernel, the polynomial kernel, the radial basis function (RBF) and the sigmoid kernel function. The gaussian RBF kernel is

commonly used because it has fewer parameters and is applicable in many situations, no matter how small and high dimensions samples. The gaussian RBF kernel is defined as [18]

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\| \mathbf{x}_i - \mathbf{x}_j \|_2^2}{\lambda^2}\right) \tag{5}$$

At this stage, the problem is aiming at determining $\boldsymbol{\alpha}$, $b$, $\lambda$ and $\gamma$; that is to say $\dim(\mathbf{x}) + 3$ parameters.

### *2.3. Heteroscedasticity issue*

Homoscedasticity refers to the circumstance in which the variability of a first variable is equal across values of a second variable that predicts it. Heteroscedasticity refers, on the contrary, to the circumstance in which the variability of a dependent variable is unequal across the range of values of the independent variable that predicts it.

With the kernel ridge regression method, statistical parameters such as the mean can be estimated for one operating condition using points obtained with different operating conditions. The further away the points used are, the more difference there will be between the statistical laws they obey and those followed by the points being investigated, which is not an issue provided that the variance remains the same. If one zone shows higher variance than the zones around it, the points contained within it will be overestimated and will carry more statistical weight when estimating the mean of the surrounding zones. The aforementioned algorithm may therefore only be applied if the distribution variance is not dependent on operating conditions, which at first glance seems unlikely. This issue is tackled through the introduction of a zero mean gaussian error whose variance depends on the operating condition [19]. The minimized objective function therefore becomes:

$$L(\mathbf{w}^\mu, b^\mu, \mathbf{w}^\sigma, b^\sigma) = \frac{1}{2}\gamma^\mu ||\mathbf{w}^\mu||^2 + \frac{1}{2}\gamma^\sigma ||\mathbf{w}^\sigma||^2 + \frac{1}{n}\sum_{i=1}^{n}\left(\log\sigma(\mathbf{x}_i) + \frac{(\mu(\mathbf{x}_i) - I_i)^2}{2\sigma^2(\mathbf{x}_i)}\right) \tag{6}$$

where $\gamma^\mu$ and $\gamma^\sigma$ are regularization parameters, and where:

$$\begin{array}{l} \mathbf{w}^\mu = \sum_{i=1}^{n} \alpha_i^\mu \phi^\mu(\mathbf{x}_i) \\ \mathbf{w}^\sigma = \sum_{i=1}^{n} \alpha_i^\sigma \phi^\sigma(\mathbf{x}_i) \\ \mu(\mathbf{x}) = \sum_{i=1}^{n} \alpha_i^\mu K(\mathbf{x}, \mathbf{x}_i) + b^\mu \\ \log\sigma(\mathbf{x}) = \sum_{i=1}^{n} \alpha_i^\sigma K^\sigma(\mathbf{x}, \mathbf{x}_i) + b^\sigma \end{array} \tag{7}$$

At this stage, the problem is aiming at determining $\boldsymbol{\alpha}^\mu$,$\boldsymbol{\alpha}^\sigma$, $b^\mu$, $b^\sigma$, $\lambda^\mu$,$\lambda^\sigma$, $\gamma^\mu$ and $\gamma^\sigma$; that is to say $(2 \cdot \dim(\mathbf{x}) + 6)$ parameters.
A robust way of estimating the parameters of conditional means and conditional variances is the Iteratively Reweighted Least Squares (IRLS) method [20]. The quality of the mean and variance models is quantified using criteria that are adapted to each specific case. The IRLS method is based on two alternative

criteria: the Predicted Residual Error Sum of Squares (PRESS) criterion when estimating the mean model, and the CRPS criterion when evaluating the variance model [21].

$$\begin{array}{lcl} PRESS & = & \sum_{i=1}^{n} \left(I_i - \mu(\mathbf{x}_i)\right)^2 \\ CRPS & = & \sigma(\mathbf{x}_i) \cdot \left(I_i^{(n)} \cdot \left(2\Phi\left(I_i^{(n)}\right) - 1\right) + 2\phi\left(I_i^{(n)}\right) - \frac{1}{\sqrt{\pi}}\right) \end{array} \tag{8}$$

With $I_i^{(n)} = \frac{I_i - \mu(\mathbf{x}_i)}{\sigma(\mathbf{x}_i)}$ and where $\Phi$ and $\phi$ represent the standard gaussian probability density and cumulative distribution functions, respectively. Amongst the many indicators of the variability estimation quality, the CPRS criterion shows the benefit to penalize prediction which are located too far away from the training data [22].

*2.4. Model hyper parameter selection*

Blind regression induces the danger that the model over-fits reality: the prediction could correspond too closely to a particular set of data, and therefore runs the risk of moving towards a model which, rather than describing the statistical law generated by the learning base, describes the specificities of that base. Overfitting must therefore be limited if useless reproduction of the vagaries of one particular case is to be avoided. Regulating the model in this way through the use of hyperparameters produces smoother models. The hyperparameters are also subject to an iterative method, all of which leads to an optimal solution. Various cross-validation diagrams allow a model's capacity for generalisation to be quantified. Each diagram offers a different way of partitioning the learning data in order to assess how appropriate it is to apply a model based on independent data. Cawley et al demonstrated that the best results were achieved by using leave-one-out cross-validation of an SVM, which produces the best ratio between the quality of the optimisation and the computational cost. [23]. Leave One Out Heteroscedastic Kernel Ridge Regression (LOOHKRR) is therefore the model used in the method outlined below.

The optimization function is realized with the matlab function *fminsearch*, which uses the Nelder-Mead simplex algorithm as described in Lagarias et al. [24] [25]. This algorithm uses a simplex of $n+1$ points for n-dimensional vectors $\mathbf{x}$. The algorithm first makes a simplex around the initial guess $\mathbf{x}_0$ by adding 5% of each component $\mathbf{x}_0(i)$ to $\mathbf{x}_0$. The algorithm uses these $n$ vectors as elements of the simplex in addition to $x_0$. Then, the algorithm modifies the simplex repeatedly according to an iterative procedure minimizing the objective function, which either is the PRESS or the CPRS criterion. The Matlab function is only improved to deal with the constraint imposing positive criterion.

The overall normalization algorithm is resumed here under:

1. Rescale $\mathbf{x}$ to avoid ill conditioned problem.
2. Define the initial condition:
   $\lambda^{\mu}[0,0] = \lambda^{\sigma}[0,0] = \frac{max(\mathbf{x}) - min(\mathbf{x})}{\lambda_0}$
   $\gamma^{\mu}[0,0] = \gamma^{\sigma}[0,0] = 1$

3. Main loop: iterate on $\lambda^{\mu}[\underline{i}, j]$, $\gamma^{\mu}[\underline{i}, j]$, $\lambda^{\sigma}[\underline{i}, j]$ and $\gamma^{\sigma}[\underline{i}, j]$ for $i = 0...N-1$
   (a) First sub loop: iterate on $\lambda^{\mu}[i, \underline{j}]$ and $\gamma^{\mu}[i, \underline{j}]$ for $j = 0..M-1$,
      i. Assuming: $\lambda^{\mu}[i, j]$, $\gamma^{\mu}[i, j]$,$\lambda^{\sigma}[i, 0]$, $\gamma^{\sigma}[i, 0]$
      ii. optimize $\boldsymbol{\alpha}^{\mu}$, $b^{\mu}$, $\boldsymbol{\alpha}^{\sigma}$ and $b^{\sigma}$ using LOOHKRR (through $P$ iterations).
      iii. optimize $\lambda^{\mu}[i, j+1]$ and $\gamma^{\mu}[i, j+1]$ using PRESS criterion
   (b) after $M$ steps of the previous subloop, define:
   $\lambda^{\mu}[i+1, 0] = \lambda^{\mu}[i, M]$
   $\gamma^{\mu}[i+1, 0] = \gamma^{\mu}[i, M]$
   (c) Second sub loop: iterate on $\lambda^{\sigma}[i, \underline{j}]$ and $\gamma^{\sigma}[i, \underline{j}]$ for $j = 0..M-1$,
      i. Assuming: $\lambda^{\mu}[i+1, 0]$, $\gamma^{\mu}[i+1, 0]$,$\lambda^{\sigma}[i, j]$, $\gamma^{\sigma}[i, j]$
      ii. optimize $\boldsymbol{\alpha}^{\mu}$, $b^{\mu}$, $\boldsymbol{\alpha}^{\sigma}$ and $b^{\sigma}$ using LOOHKRR (through $P$ iterations).
      iii. optimize $\lambda^{\sigma}[j+1]$ and $\gamma^{\sigma}[j+1]$ using CRPS criterion,
   (d) after $M$ steps of the previous subloop, define:
   $\lambda_{\sigma}[i+1, 0] = \lambda_{\sigma}[i, M]$
   $\gamma_{\sigma}[i+1, 0] = \gamma_{\sigma}[i, M]$.
4. After $N$ steps of the main loop, define the final estimation of the SVR: $\lambda^{\mu}$, $\lambda^{\sigma}$, $\boldsymbol{\alpha}^{\mu}$,$\boldsymbol{\alpha}^{\sigma}$,$b^{\mu}$ and $b^{\sigma}$.
5. Estimate $\mu(\mathbf{x})$ and $\sigma(\mathbf{x})$ using Eq.7.
6. Normalise the indicator using Eq. 2 after having unscaled $\mathbf{x}$.

## 3. Experimental validation

The approach set out in this paper is a combination of various tools which have already been used successfully in various fields. Whether or not this approach is of interest to those working in rotating machine monitoring remains to be seen. This paper sets out one possible application where automatic detection of a bearing defect would not be possible without using this method. The first step must therefore be to describe the practical example, which is based on spectral analysis of Instantaneous Angular Speed, an alternative to the traditional vibration monitoring method.

### *3.1. Experiment presentation*

The wind turbine monitored in this paper is an onshore machine with an 80m diameter rotor, delivering 2MW of nominal power. The low-speed shaft connects the hub blades to the gearbox, which reduces the torque by a factor of 100. The high-speed shaft connects the gearbox to the generator via a flexible fuse coupling. Using this coupling means that problems relating to misalignment and the radial load on the generator bearings can be kept to a minimum. Figure 1 shows the shaft line of the wind turbine which was analysed. The defect analysed in this case was in one of the generator's two bearings. A peculiarity of this machine is that the generator's two bearings are identical: the same ring geometry, the same number of rolling elements, and therefore the same characteristic frequencies. The characteristic frequencies of these deep groove bearings

depend on the contact angle between the rolling element and the internal and external rings. The contact angle depends on the ratio between the axial and radial loads. As the shaft is tilted to 5, these two loads are partly connected to the weight of the shaft itself. They are also liable to take on torsion torque via the coupling of the helical gears on the gearbox output shaft. The relationship between the axial and radial loads is not necessarily the same for all speed and load conditions, and this may lead to two eventualities: firstly, the characteristic frequency will change from one load condition to another; secondly, the rolling elements can only make contact with the defect under certain load conditions. The operator of the wind farm shared a database containing 400 measurements

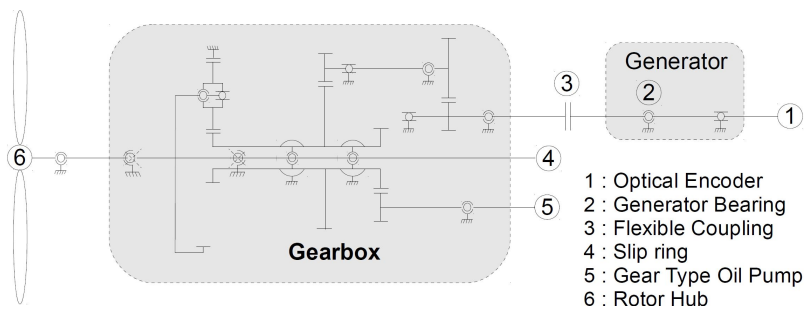


Figure 1: Kinematic Scheme of the wind turbine set-up.

from turbines without defects and 600 measurements where the probable defect emerged at an unknown time. The measurement campaign analysed here comprises signals detected via the elapsed-time method, using an optical encoder with a resolution of 2048 pulses per revolution, via a 120MHz high-frequency clock. Signals length is $3 \cdot 10^6$ points. The measurements from turbines without defects were used to provide learning data, as though they had been taken when the monitoring system was put in place. The unknown measurements are considered to be those taken after the learning period. In fact, the measurements from turbines with no defects were taken in another time period, once the faulty bearing had been replaced.

### *3.2. Results*

The scalar indicators used in this paper to build the SVR model are extracted from the spectral observation of the IAS signal, as explained in [2]: the mean value of the whitened spectrum between the null frequency and the $10^{th}$ harmonic of the bearing characteristic outer cyclic frequency. These indicators are dotted in black on Fig 2 at abscissae corresponding to the average speed of their respective measurements.

The regression law $\mu(\mathbf{x})$ is drawn as a continuous blue curve, while its standard deviation $\sigma(\mathbf{x})$ is drawn in dashed red above and under it. Both of these statistical parameters have been obtained using the algorithm presented in part 2.4, and whose optimization parameters are defined such as: $\lambda_0 = 10$, $M = 4$, $N = 10$, $P = 10$. The influence of these parameters on the SVR quality will be

discussed later on.
The left plot shows the unsteady variability of the indicator: low speed measurements give relatively weak and stable indicators while high speed measurements give relatively great and variable indicators.
The green points plotted on Fig. 2 right plots show the indicator amplitudes obtained in January, just before the defective bearing removal. Top figure shows that in the presence of defect, low speed measurement give higher level indicators while indicators obtained under high speed measurement are not moved.
Top right and bottom right plots shows the same measurements presented either in raw or normalized format. The bottom plot presents the normalized indicators obtained using Eq. 2. The value of 40 means that the indicator is $40\sigma(\mathbf{x})$ away from its mean value for such an operating condition $\mathbf{x}$, assuming that the bearing is healthy. As a consequence of the previous observation, this great scattering is only obtained for low speed measurements.

Although this specific result cannot be generalized since it only corresponds to this machine, this defect and this indicator; this case enlightens the normalisation process proposed in this paper: low speed measurements being here more precise, they are more sensible to a slight evolution of the bearing health status. Moreover, since the raw amplitude of the indicator obtained under $H_1$ are lower than those obtained under $H_0$, a threshold on the raw indicator cannot differentiate $H_0$ from $H_1$.

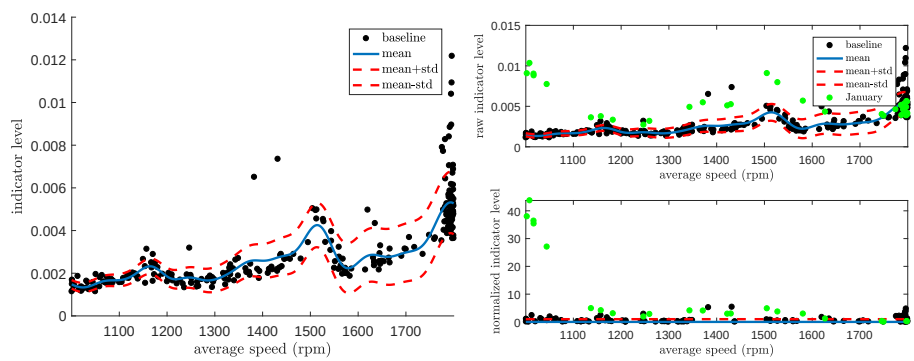


Figure 2: Left: mean and variance regression law for a bearing indicator observed under different speed conditions, under $H_0$: healthy case.
Right: raw indicators (on top) and normalized indicators (on bottom) against their respective acquisition date.

Fig 3 plots every indicator obtained from the 600 undetermined measurements against their acquisition date. On left, the ordinates represent the exact amplitude of the indicator. This amplitude can be seen as the multiplication of the defect harmonic amplitude by the transfer function between the fault and the sensor, which is completely unknown. As a consequence, it is hard to grasp any evolution on the left plot.
On right, ordinates present the normalized indicator level. Moving average,

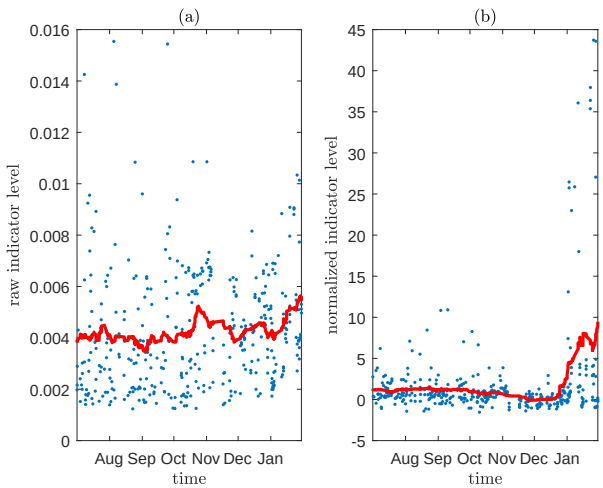


Figure 3: Raw indicators (on left) and normalized indicators (on right) against their respective acquisition date. red lines are moving average.

plotted in red for both chart, takes 20 points into account. The raw indicator average (left plot) is weakly increasing while the normalized indicator goes from 0 to almost 10. As one could expect, indicator normalization appears as an efficient tool to automate condition monitoring: a threshold can now easily be tuned to differentiate $H_0$ from $H_1$.

### *3.3. Discussion*

The paper is not focused on determining the most efficient indicator in term of early detection, but is rather focused on underlining the optimization of any indicator once it is normalized. Although the indicator chosen in this paper does not differentiate $H_0$ from $H_1$ as plainly as many other could, it was chosen since its improvement through normalization appeared plainly necessary to advocate for the presence of a defect.

The left plot presented on Figure 2 shows that this raw indicator does not follow a univariate gaussian law, but a multivariate law which is perfectly fit for this normalization. This specific indicator presents higher values around $1500rpm$ and its distribution is spreading out for higher speeds, but other indicators might behave completely differently: in any cases, the kernel ridge regression proposed in this paper does not require any assumption regarding the shapes of the local mean value and variance except two: they must be describable by continuous functions and the indicator noise needs to be gaussian. An indicator that is affected by a $\chi^2$ noise would not be modified as it should by the normalization, although this indicator is independent from the operating conditions.

The SVR approach proposed in this paper relies on several parameters that need to be tuned beforehand. This limitation is even more problematic if some awkward parameter tuning lead to a wrong diagnostic: a miss or false alarm. Figures 4 and 5 aim to moderate that risk.

The attention is first drawn on the five plots in figure 4, where $\lambda_0$ is set equal to 10, which means that each gaussian kernel is wide as one tenth of the operating condition entire span (see Eq. 5). In this case, measurement mean speed lays within $[1000 - 1800]$ rpm, and $\lambda_0 = 10 \Rightarrow \lambda^\mu[0,0] = 80rpm$, which is consistent with the expected shape of the aimed model along with the number of indicator used to estimate it. As a consequence, every time series but the last one presents similar result: only $P$ needs to be kept equal to 10 for the regression to be efficient. $P$ is the number of iteration used in the LOOHKRR to estimate $\boldsymbol{\alpha}^\mu$,$\boldsymbol{\alpha}^\sigma$,$b^\mu$ and $b^\sigma$. The four right plots show the influence of each step of the PRESS and CRPS sub loop on the hyper parameter values. As a reminder, PRESS sub loop only influence $\lambda^\mu$ and $\gamma^\mu$ while CRPS sub loop only influence $\lambda^\sigma$ and $\gamma^\sigma$. These plot shows the first subloop is already able to reach a stable estimation of $\lambda^\mu$. This implies this hyper parameter is mainly influencing the regression quality, and that the $M$,$N$,$P$ values should not considered as a source of error when $\lambda_0$ is correctly initialised.

In the five plots of figure 5, $\lambda_0$ is intentionally set to 1000, which is obviously too high: every kernel will hardly find more than one training data to estimate a regressed value. As a consequence, the time series present various shape depending on the number of iteration performed by the algorithm. Only $M = P = N = 10$ presents results similar to the case where $\lambda_0 = 10$. The four right plots shows the number of steps needed by the algorithm to converge towards a stable estimation of $\lambda^\mu$ and $\lambda^\sigma$. For that matter, Fig. 6 shows the functions $\mu(\mathbf{x}) \pm \sigma(\mathbf{x})$ estimated after $M = 5$ and after $M = 8$. These plots underline both the importance and the difficulty of correctly estimating $\mu(x)$.

One should finally mention that although only speed was taken into account, there is no strong limitations to adding other operating conditions with the SVR technique, except computation time. The computation is not expected to be run in a real-time approach but only once per indicator. The results will then be instantaneously reused to normalize each new indicator.

## 4. Conclusion

In the wide domain of machine monitoring, although scalar indicators extracted from IAS as well as vibration signals are both numerous and unique, they all are subject to noise. Although the noise should remain constant as long as operating conditions remain stable, this is no longer the case as soon as there are variations in the rotational speed or macroscopic load variations of the machine being monitored. The method set out in this paper has two objectives: firstly, to estimate the statistical parameters of indicators in relation

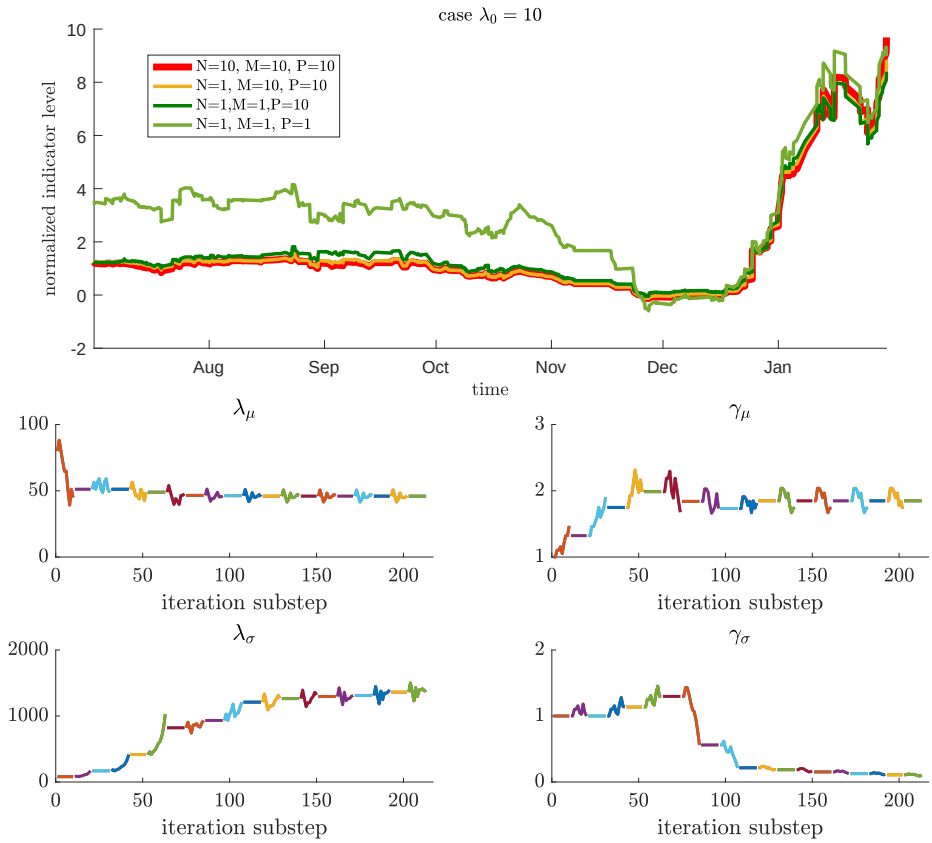


Figure 4: Influence of the initial condition $\lambda_0$ on the number of iteration needed to get a suitable regression. ($\lambda_0 = 1000$)
Top: averaged normalized indicator time series for different values of $M$, $N$ and $P$.
Bottom: successive values of the hyper parameters if $N = 10$ (1 color = 1 sub loop).

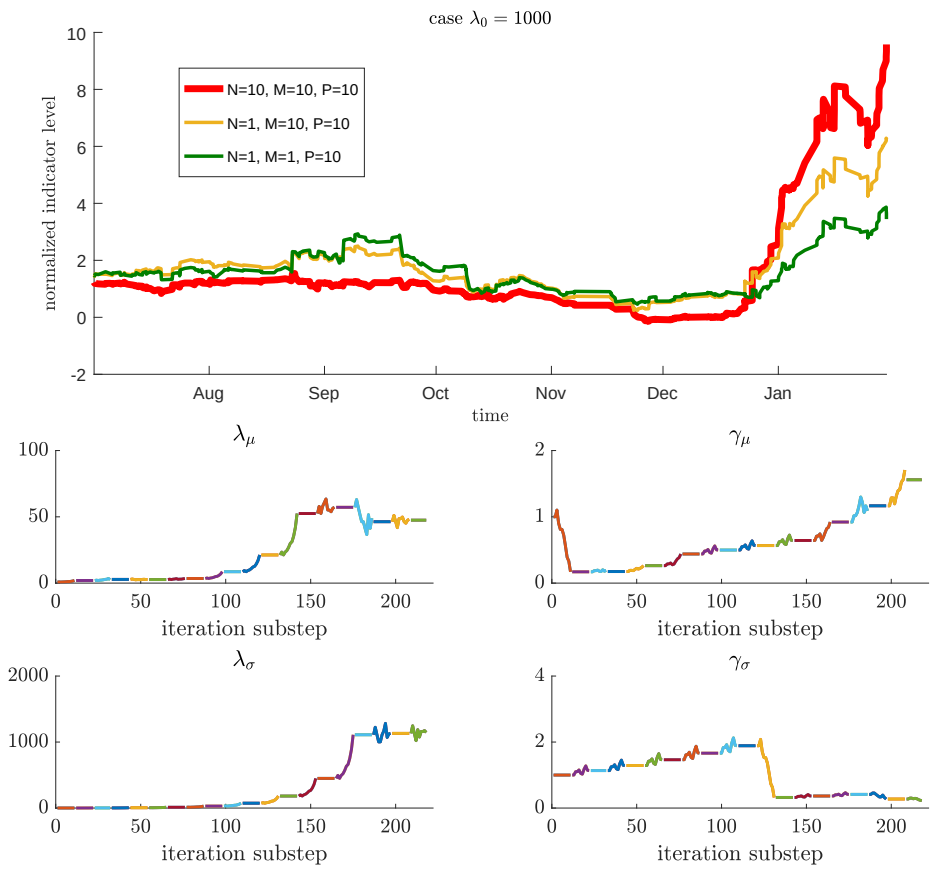


Figure 5: Influence of the initial condition $\lambda_0$ on the number of iteration needed to get a suitable regression. ($\lambda_0 = 10$)
Top: averaged normalized indicator time series for different values of $M$, $N$ and $P$.
Bottom: successive values of the hyper parameters if $N = 10$ (1 color = 1 sub loop).

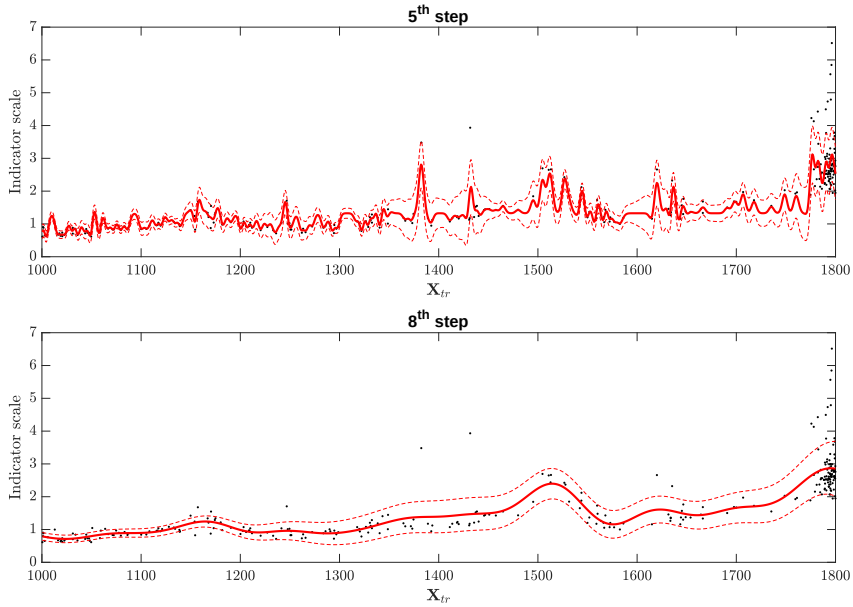


Figure 6: Intermediary steps of the regression using $\lambda_0 = 1000$, $M = 10$, $N = 10$ and $P = 10$. continuous and dash curves are respectively the $\mu(\mathbf{x})$ and $\mu(\mathbf{x}) \pm \sigma(\mathbf{x})$ functions estimated after the $5^{th}$ step (top) and the $8^{th}$ step (bottom).

to operating conditions, and secondly to use these parameters to normalise the indicators. The first stage could be described as regression: overall statistical behaviour is estimated from data constituting a baseline. The second stage could be described as standardisation: the indicators are reduced in such a way as to follow a normal distribution centred on 0 with unit standard deviation. This paper aims at providing a summary of the entire procedure and to describe in detail the influence model configuration parameters have on the efficiency of the monitoring system. The approach set out in this paper is a combination of various tools which have already been used successfully in other fields. Whether or not this approach is of interest to those working in rotating machine monitoring remains to be seen. This paper sets out one possible application where automatic detection of a bearing defect would not be possible without using this method. However, this approach relies on the hypothesis that the healthy state indicators follow a gaussian law. Therefore, this approach could be improved through an adaptation of its process to other statistical behaviour.

## 5. Acknowledgement

The authors would like to thank ENGIE Green for their financial support.

## References


[1] H. André, F. Girardin, A. Bourdon, J. Antoni, and D. Rémond. Precision of the ias monitoring system based on the elapsed time method in the spectral domain. *Mechanical Systems and Signal Processing*, 44(1):14 – 30, 2014. Special Issue on Instantaneous Angular Speed (IAS) Processing and Angular Applications.

[2] H. André, A. Bourdon, and D. Rémond. On the use of the instantaneous angular speed measurement in non stationary mechanism monitoring. In *IDETC,*, 10.1115/DETC2011-47470, pages 15–24, 2011.

[3] H. André, I. Khelf, and Q. Leclere. Wind Turbine Bearing fault detected with IAS combined with Harmonic Product Spectrum. In *COMADEM 2017*, Preston, United Kingdom, July 2017.

[4] Q. Leclere H. Andre, I. Khelf. Harmonic Product Spectrum revisited and adapted for rotating machine monitoring based on IAS. In *SURVEILLANCE 9*, April 2017.

[5] J.L. Gomez, A. Bourdon, H. André, and D. Rémond. Modelling deep groove ball bearing localized defects inducing instantaneous angular speed variations. *Tribology International*, 98:270 – 281, 2016.

[6] X.G. Hua, Y.Q. Ni, J.M. Ko, and K.Y. Wong. Modeling of temperature and frequency correlation using combined principal component analysis and support vector regression technique. *Journal of Computing in Civil Engineering*, 21(2):122–135, 2007.

[7] H. Sohn, M. Dzwonczyk, E. G. Straser, A. S. Kiremidjian, K. H. Law, and T. Meng. An experimental study of temperature effect on modal parameters of the alamosa canyon bridge. *Earthquake Engineering & Structural Dynamics*, 28(8):879–897, 1999.

[8] C.J. Stander, P.S. Heyns, and W. Schoombie. Using vibration monitoring for local fault detection on gears operating under fluctuating load conditions. *Mechanical Systems and Signal Processing*, 16(6):1005 – 1024, 2002.

[9] K. Worden, H. Sohn, and C.R. Farrar. Novelty detection in a changing environment: Regression and interpolation approaches. *Journal of Sound and Vibration*, 258(4):741 – 761, 2002.

[10] M. Timusk J. McBain. Fault detection in variable speed machinery: Statistical parameterization. *Journal of Sound and Vibration*, 327(3):623 – 646, 2009.

[11] C. Saunders, A. Gammerman, and V. Vovk. Ridge regression learning algorithm in dual variables. In *Proceedings of the Fifteenth International Conference on Machine Learning*, ICML '98, pages 515–521, San Francisco, CA, USA, 1998. Morgan Kaufmann Publishers Inc.

[12] H. Drucker, C.J.C. Burges, L. Kaufman, A. Smola, and V. Vapnik. Support vector regression machines. In *Proceedings of the 9th International Conference on Neural Information Processing Systems*, NIPS'96, pages 155–161, Cambridge, MA, USA, 1996. MIT Press.

[13] M.A. Aizerman, E. A. Braverman, and L. Rozonoer. Theoretical foundations of the potential function method in pattern recognition learning. In *Automation and Remote Control,*, number 25 in Automation and Remote Control,, pages 821–837, 1964.

[14] H. Drucker, C.J.C. Burges, L. Kaufman, A.J. Smola, and Vladimir Vapnik. Support vector regression machines. In M. C. Mozer, M. I. Jordan, and T. Petsche, editors, *Advances in Neural Information Processing Systems 9*, pages 155–161. MIT Press, 1997.

[15] T. Hofmann, B. Schlkopf, and A. J. Smola. Kernel methods in machine learning. *Ann. Statist.*, 36(3):1171–1220, 06 2008.

[16] R. Lefort, G. Real, and A. Drmeau. Direct regressions for underwater acoustic source localization in fluctuating oceans. *Applied Acoustics*, 116:303 – 310, 2017.

[17] A. Bartkowiak and R. Zimroz. Dimensionality reduction via variables selection linear and nonlinear approaches with application to vibration-based condition monitoring of planetary gearbox. *Applied Acoustics*, 77:169 – 177, 2014.

[18] M. Li, X. Liu, and X. Liu. Infrasound signal classification based on spectral entropy and support vector machine. *Applied Acoustics*, 113:116 – 120, 2016.

[19] G.C. Cawley, N.L.C. Talbot, R.J. Foxall, S.R. Dorling, and D.P. Mandic. Heteroscedastic kernel ridge regression. *Neurocomputing*, 57:105 – 124, 2004. New Aspects in Neurocomputing: 10th European Symposium on Artificial Neural Networks 2002.

[20] I.T. Nabney. Efficient training of rbf networks for classification. In *1999 Ninth International Conference on Artificial Neural Networks ICANN 99. (Conf. Publ. No. 470)*, volume 1, pages 210–215 vol.1, Sep. 1999.

[21] E. P. Grimit, T. Gneiting, V. J. Berrocal, and N. A. Johnson. The continuous ranked probability score for circular variables and its application to mesoscale forecast ensemble verification. *Quarterly Journal of the Royal Meteorological Society*, 132(621C):2925–2942, 2006.

[22] H. Hersbach. Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting*, 15(5):559–570, 2000.

[23] G.C. Cawley and N.L.C. Talbot. Fast exact leave-one-out cross-validation of sparse least-squares support vector machines. *Neural Networks*, 17(10):1467 – 1475, 2004.

[24] J. Lagarias, J. Reeds, M. Wright, and P. Wright. Convergence properties of the nelder–mead simplex method in low dimensions. *SIAM Journal on Optimization*, 9(1):112–147, 1998.

[25] J.A. Nelder and RA Mead. A simplex method for function minimization comput. *The Computer Journal*, 7, 01 1965.