

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/318767684>

Improved Similarity-Based Modeling for the Classification of Rotating-Machine Failures

Article in *Journal of the Franklin Institute* · July 2017

DOI: 10.1016/j.jfranklin.2017.07.038

CITATIONS

17

READS

3,795

4 authors, including:



Matheus Marins
Federal University of Rio de Janeiro

10 PUBLICATIONS 26 CITATIONS

[SEE PROFILE](#)



Sergio Lima Netto
Federal University of Rio de Janeiro

156 PUBLICATIONS 1,223 CITATIONS

[SEE PROFILE](#)



Eduardo A. B. da Silva
Federal University of Rio de Janeiro

290 PUBLICATIONS 2,553 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Energy Conservation in Wireless Sensor Networks for IoT Applications [View project](#)



DORIS project [View project](#)

Improved Similarity-Based Modeling for the Classification of Rotating-Machine Failures

Matheus A. Marins, Felipe M. L. Ribeiro*, Sergio L. Netto, Eduardo A. B. da Silva

*Program of Electrical Engineering, Universidade Federal do Rio Janeiro.
Rio de Janeiro, RJ, 21941-972, Brazil.*

Abstract

Similarity-based modeling (SBM) is a technique whereby the normal operation of a system is modeled in order to detect faults by analyzing their similarity to the normal system states. First proposed around two decades ago, SBM has been successfully used for fault detection in varied systems. In spite of this success, there is not much study performed in the literature regarding its design, that encompasses both similarity metrics and model training. This work aims at contributing with an in-depth study of SBM for fault detection considering these two design aspects. This is done in the context of proposing a novel system to identify rotating-machinery faults based on SBM, that is employed either as a standalone classifier or to generate features for a random forest classifier. New approaches for training the model and new similarity metrics are investigated. Experimental results are shown for the recently developed Machinery Fault Database (MaFaulDa) that has

*Corresponding author.

Email addresses: matheus.marins@smt.ufrj.br (Matheus A. Marins),
felipe.ribeiro@smt.ufrj.br (Felipe M. L. Ribeiro), sergioln@smt.ufrj.br (Sergio L. Netto), eduardo@smt.ufrj.br (Eduardo A. B. da Silva)

an extensive set of sequences and fault types, and for the Case Western Reserve University (CWRU) bearing database. Results for both databases indicate that the proposed techniques increase the generalization power of the similarity model and of the associated classifier, achieving accuracies of 98.5% on MaFaulDa and 98.9% on CWRU database.

Keywords: Rotating machinery fault diagnosis, Condition monitoring, Feature extraction, Similarity-based modeling.

¹ 1. Introduction

² Maintenance of critical equipment to ensure high levels of reliability,
³ availability, and performance is one of the major concerns on today's in-
⁴ dustrial sector [1]. Unexpected failures can lead to substantial losses, ei-
⁵ ther from the maintenance procedure itself or from the resulting production
⁶ halts [2]. To achieve an effective and cost-efficient procedure, new mainte-
⁷ nance strategies are being devised based on real-time and continuous moni-
⁸ toring, allowing one to detect and classify operational anomalies at an early
⁹ stage, thus limiting additional system degradation [2]. Applications of such
¹⁰ techniques include, for instance, flight paths [3], natural gas and nuclear
¹¹ power plants [4, 5, 6, 7], wind turbines [8], and bearing or rotating-machine
¹² faults [9, 10, 11, 12, 13, 14]. Among these equipments, rotating machines
¹³ are some of the most important, being a key element used in a variety of
¹⁴ applications, including airplanes, automobiles, power turbines, oil and gas
¹⁵ refineries, and so on [12, 15].

¹⁶ There are many approaches for detecting faults in rotating machines.
¹⁷ Most of them consist of extracting features from the vibration signal to as-

18 sess the machine current condition, in a supervised or automatic manner.
19 Different features are needed to extract useful information relevant to de-
20 tect faults from the original sources over multiple conditions. These features
21 can be classified considering their domain (time, spatial, time-frequency, fre-
22 quency) or its computation method (e.g. transform coefficients or aggregated
23 statistics) [16, 17, 18, 19].

24 An illustrative example is the approach in Yang et al. [20]. There, a
25 system is presented which uses an adaptive resonance theory Kohonen neural
26 network (ART-KNN) for fault diagnosis, having as inputs features derived
27 from the discrete Wavelet transform coefficients. Unfortunately, the fault
28 database used is not publicly available, making its comparison with other
29 approaches impractical.

30 A methodology for detecting broken and half broken bars using spectral
31 information over a FPGA is presented in [21]. This methodology is latter
32 extended in [22] and in [23], adding the discrete wavelet transform (DWT)
33 coefficients as features, and combining it with discrete frequency transform
34 coefficients. These works also treat the detection of other faults and failures.
35 The broken bar detection problem is also approached by the authors of [24]
36 using motor current signature analysis and mathematical morphology.

37 The authors of [25] focus on the feature extraction procedure proposing
38 a novel feature extraction scheme which utilizes the generalized S transform
39 and 2D non-negative matrix factorization to detect possible faults. Three
40 classifiers were used to assess the system: k -nearest neighbors (k NN), naive
41 Bayes, and support vector machines (SVM), all achieving good results. A
42 similar approach is presented in [26] using multiscale permutation entropy

43 for feature extraction and an SVM classifier for fault diagnosis. The work
44 of Rauber et. al. [18] also studies the effect of the features in the system
45 performance. It tests multiple features of different types, such as complex
46 envelope spectrum, statistical time- and frequency-domain parameters, as
47 well as wavelet packet analysis, together with a feature selection algorithm. A
48 fault classification database was used as testbed, and three different classifiers
49 (k NN, feedforward artificial neural networks (ANN), and SVMs) were used
50 during the assessment, achieving good performance.

51 This work proposes an automatic fault detector and classifier that uses
52 similarity-based modeling (SBM) to identify rotating-machine failures such
53 as imbalanced load, (horizontal or vertical) shaft misalignment, and bearing
54 defects (in rolling elements or inner/outer tracks). The similarity model can
55 be used either as an auxiliary model to generate features for the classifier
56 (a random forest classifier in this case) or as a standalone classifier. In this
57 context, new approaches for training the similarity model and new similar-
58 ity metrics are investigated. Two databases were employed to evaluate the
59 performance of the proposed techniques. The first one is the machinery fault
60 database (MaFaulDa) [27], a relatively new, large database of problematic
61 scenarios of rotating-machine operations [13, 14]. Performance evaluation
62 on this database included continuous monitoring of six vibration sensors,
63 one microphone, and one tachometer [14]. The second database is the Case
64 Western Reserve University (CWRU) bearing database [28]. This database
65 has become a standard reference in the bearing diagnostics field [29, 19]
66 and is used as testbed for comparison between the proposed methodology
67 against other algorithms [25, 26, 30, 18]. Results indicate that the proposed

68 methodology is capable of correctly diagnosing the machine operating states,
69 achieving an accuracy of 98.5% on the MaFaulDa dataset and 98.9% on the
70 CWRU database.

71 This paper is organized as follows: Section 2 presents the original SBM
72 technique [3, 5, 7, 9, 10, 11], devised for detecting unusual patterns in some
73 system or machine operation. Section 3 describes the proposed modifications
74 to the standard SBM technique that allow the detection and classification of
75 different types of anomalous machine operations in an efficient and robust
76 manner. Section 4 details the MaFaulDa database, used to design and evalua-
77 ate the system’s performance and the CWRU database, used for comparison.
78 The methodology of performance assessment is described in Section 5. This
79 section also describes the designed system, including the preprocessing and
80 validation procedures. Section 6 discusses the experimental results obtained
81 during the processes of training and selection of the best model, as well the
82 assessment results. Comparisons to other works are also included in this
83 section. Finally, conclusions and discussions emphasizing the main contribu-
84 tions of this paper are provided in Section 7.

85 **2. Similarity-Based Modeling (SBM)**

86 The SBM is a simple and yet powerful nonparametric modeling technique
87 that puts together an ensemble of previous state vectors in a single matrix
88 \mathbf{D} to represent the normal behaviour of a given system, process, or machine.
89 The SBM then evaluates the similarity of the current state vector with all
90 vectors within \mathbf{D} to assess the normality or not of the current system opera-
91 tion. This technique was first proposed in [4], and since then has been used

92 in a variety of industrial applications, such as fault diagnosis in a machinery
 93 fault simulator [9, 11], modeling airplanes flight paths [3], and anomaly de-
 94 tection in power plants [7]. In this work the SBM technique is adapted to
 95 monitor and classify the current operation of a given rotating machine.

96 In the original SBM framework, a system state at time n is represented
 97 by a vector $\mathbf{x}_n = [x_n(1), x_n(2), \dots, x_n(M)]^T$ comprising M measurements or
 98 features from multiple sources, such as sensors or signals. Given a set of
 99 L historical data \mathbf{x}_n , with $n = n_1, n_2, \dots, n_L$, corresponding to the normal
 100 behavior of a given system, one can then represent this operational mode by
 101 the $L \times M$ “memory” matrix \mathbf{D} which stacks the \mathbf{x}_n in a line-by-line manner,
 102 as given by [4]

$$\mathbf{D} = \begin{bmatrix} \mathbf{x}_{n_1} & \mathbf{x}_{n_2} & \cdots & \mathbf{x}_{n_L} \end{bmatrix}^T. \quad (1)$$

103 Using this matrix, one may evaluate if a new input state \mathbf{x} corresponds to
 104 the system normal operation by attempting to describe this state as a linear
 105 combination of the previously selected representative states contained in \mathbf{D} ,
 106 that is

$$\hat{\mathbf{x}}_n = \mathbf{D}^T \mathbf{w}_n. \quad (2)$$

107 By defining the estimation error and estimation error function respectively
 108 as

$$\mathbf{e}_n = \mathbf{x}_n - \hat{\mathbf{x}}_n, \quad (3)$$

$$e_n = \|\mathbf{x}_n - \hat{\mathbf{x}}_n\|_2^2, \quad (4)$$

109 the optimal linear estimate of \mathbf{x}_n becomes

$$\hat{\mathbf{x}}_n = \mathbf{D}^T \mathbf{w}_n^o = \mathbf{D}^T (\mathbf{D} \mathbf{D}^T)^{-1} \mathbf{D} \mathbf{x}_n, \quad (5)$$

110 where \mathbf{w}_n^o is the optimal estimator.

111 However, in [4], the authors argue that this result has numerous limita-
112 tions, such as an inability to accommodate random uncertainties and non-
113 random defects, the need of a very large number L of archetypical states that
114 define \mathbf{D} , and the requirement that \mathbf{DD}^T must be nonsingular.

115 The SBM was proposed as an alternative approach that copes with all
116 these issues by substituting the dot product implicit in Eq. (2) by a *similarity*
117 *operation* $s(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i \circ \mathbf{x}_j$. This operation provides a similarity score
118 $0 \leq s \leq 1$ between two vectors \mathbf{x}_i and \mathbf{x}_j , such that non-similar vectors yield
119 $s \approx 0$ and very similar vectors correspond to $s \approx 1$. An example of similarity
120 function considered in the original SBM formulation is given by [9]

$$s(\mathbf{x}_i, \mathbf{x}_j) = \frac{1}{1 + \|\mathbf{x}_i - \mathbf{x}_j\|_2}. \quad (6)$$

121 The incorporation of distinct similarity operations into the SBM formulation
122 is one of the contributions of the present work, as detailed in Subsection 3.2.

123 Following the linear approach and using the similarity operation, the
124 current state vector \mathbf{x}_n in the SBM algorithm can be estimated as

$$\hat{\mathbf{x}}_n = \mathbf{D}^T \frac{\mathbf{w}_n}{\|\mathbf{w}_n\|_1}, \quad (7)$$

125 with

$$\mathbf{w}_n = (\mathbf{D} \circ \mathbf{D}^T)^{-1} (\mathbf{D} \circ \mathbf{x}_n) = \mathbf{G}^{-1} \mathbf{a}_n. \quad (8)$$

126 that is, $\mathbf{G} = \mathbf{D} \circ \mathbf{D}^T$ and $\mathbf{a}_n = \mathbf{D} \circ \mathbf{x}_n$. The vector \mathbf{a}_n evaluates the similarity
127 between the current state and the representative states in matrix \mathbf{D} and
128 matrix \mathbf{G} transforms the similarity vector \mathbf{a}_n in a set of weights for each state
129 in \mathbf{D} . When $\mathbf{G} = \mathbf{I}$, the model is called auto-associative kernel regression

₁₃₀ (AAKR) [6], a particular case of SBM, equivalent to assuming no similarity
₁₃₁ between the state samples within \mathbf{D} . We mean by no similarity $s(\mathbf{x}_i, \mathbf{x}_j) = 1$
₁₃₂ if $i = j$ and zero otherwise. In this case, if \mathbf{x}_n belongs to \mathbf{D} , one then gets
₁₃₃ $\hat{\mathbf{x}}_n = \mathbf{x}_{n_l}$ such that the vector \mathbf{w}_n becomes all null with the exception of an
₁₃₄ ‘1’ entry at its n_l -th position.

₁₃₅ A key aspect within the SBM formulation is the strategy for composing
₁₃₆ matrix \mathbf{D} . Using all L historical samples for the normal behavior would
₁₃₇ incur in high computational expenses and very redundant data. Choosing
₁₃₈ an inadequate vector set when opting for a smaller L leads to performance
₁₃₉ impairments. The best possible set, therefore, would have the minimal num-
₁₄₀ ber of vectors still yielding the same performance level as the complete set.
₁₄₁ In [5], a strategy is proposed for selecting a proper reduced set of historical
₁₄₂ samples. It comprises two selection steps:

- ₁₄₃ 1. One chooses as representatives the samples with index in the set $J =$
₁₄₄ $\{j_1, j_2, \dots, j_K\}$, $K \leq 2M$, built such that

$$j \in J \quad \text{if} \quad \exists m : x_j(m) = \min_n \{x_n(m)\} \vee x_j(m) = \max_n \{x_n(m)\}; \quad (9)$$

- ₁₄₅ 2. The other samples are sorted by their ℓ_2 norm in decreasing order and
₁₄₆ decimated by a factor of t . The ones that remain after the decimation
₁₄₇ complement the set of representative samples.

₁₄₈ The first step inserts in \mathbf{D} all vector states which present the minimum
₁₄₉ and maximum value of each vector component. The second step decimates
₁₅₀ the remaining vectors using the ℓ_2 norm as ordering criterion. However, this
₁₅₁ second step may lead to not so good choices, because vectors with similar
₁₅₂ (even identical) norm values can be completely different [8]. Also, given

153 a small decimation factor t and the number of samples L , the number of
 154 chosen samples is $\bar{L} = K + \lfloor(L - K)/t\rfloor$, which may be not much lower
 155 than L . Additional strategies for composing the matrix \mathbf{D} are proposed in
 156 Subsection 3.3 in an attempt to overcome these issues.

157 **3. Proposed SBM Enhancements**

158 This section presents the proposed enhancements to the SBM formulation,
 159 which include: a generalization of the SBM framework that allows it to
 160 operate in a multiclass (more than two classes) scenario; introduction of
 161 alternative similarity operations; and the development of a new strategy to
 162 compose the matrix \mathbf{D} .

163 *3.1. Multiclass SBM*

164 SBM was originally devised to detect abnormal operating conditions,
 165 which are associated with a low similarity level between a current state vector
 166 \mathbf{x}_n and its SBM estimate $\hat{\mathbf{x}}_n$ given in Eq. (7).

167 Such framework can be extended, however, to detect and classify several
 168 types of system operational modes by defining a distinct model-matrix \mathbf{D}_c
 169 for each operational class c . In the proposed multiclass SBM formulation,
 170 given a new input state \mathbf{x}_n , a different estimate can be determined for each
 171 class

$$\hat{\mathbf{x}}_{n,c} = \mathbf{D}_c^T \frac{\mathbf{w}_{n,c}}{\|\mathbf{w}_{n,c}\|_1}, \quad (10)$$

172 where

$$\mathbf{w}_{n,c} = (\mathbf{D}_c \circ \mathbf{D}_c^T)^{-1} (\mathbf{D}_c \circ \mathbf{x}_{n,c}) = \mathbf{G}_c^{-1} \mathbf{a}_{n,c}. \quad (11)$$

₁₇₃ The current state \mathbf{x}_n is then associated to the class c_n^* which maximizes the
₁₇₄ similarity,

$$c_n^* = \arg \max_c \{s_{n,c}\} = \arg \max_c \{s(\mathbf{x}_n, \hat{\mathbf{x}}_{n,c})\}. \quad (12)$$

₁₇₅ *3.2. Alternative Similarity Functions*

₁₇₆ A *distance metric* is a function which maps a pair of elements in a set to a
₁₇₇ non-negative real number and satisfies a set of conditions. A classic example
₁₇₈ of distance metric is the family of the p -distance metrics [31], defined as

$$d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|_p = \left(\sum_{m=1}^M |x_{im} - x_{jm}|^p \right)^{\frac{1}{p}}. \quad (13)$$

₁₇₉ The above equation is also known as ℓ_p norm. It is typically used with either
₁₈₀ $p = 2$, the ℓ_2 norm or Euclidean distance, or $p = 1$, the ℓ_1 norm or Manhattan
₁₈₁ distance.

₁₈₂ In this work, five distinct similarity functions are employed, which can
₁₈₃ be separated in two main families: the multiquadric set and the exponential
₁₈₄ set. The multiquadric kernels include three of the selected functions which
₁₈₅ are based on the *inverse multiquadric function*, defined as [32]

$$f(d_{ij}) = \frac{1}{(r^2 + d_{ij}^2)^\alpha}. \quad (14)$$

₁₈₆ The first similarity function is a direct application of the above equation.
₁₈₇ By making $\alpha = 1/2$, $r^2 = 1/\gamma^2$, and $s_{IMK}(d_{ij}) = f(d_{ij})/\gamma$, one gets

$$s_{IMK}(d_{ij}) = \frac{1}{\sqrt{1 + \gamma^2 d_{ij}^2}}, \quad (15)$$

₁₈₈ which is the *inverse multiquadric kernel* (IMK) similarity function.

189 The second similarity function, the *Cauchy kernel* [33], is a direct variation
 190 of the IMK function, and can be defined as

$$s_{\text{CCK}}(d_{ij}) = s_{\text{IMK}}^2(d_{ij}) = \frac{1}{1 + \gamma^2 d_{ij}^2}. \quad (16)$$

191 The third and last multiquadric function is a modified version of the original
 192 SBM function, so-called *Wegerich similarity function*, shown in Eq. (6),
 193 and defined as

$$s_{\text{WSF}}(d_{ij}) = \frac{1}{1 + \gamma d_{ij}}. \quad (17)$$

194 The last two functions are representatives of the exponential set. The
 195 *exponential* or *Laplacian kernel* is defined as

$$s_{\text{EXP}}(d_{ij}) = e^{-\gamma d_{ij}}, \quad (18)$$

196 and the *radial basis function kernel* as

$$s_{\text{RBF}}(d_{ij}) = e^{-\gamma d_{ij}^2}. \quad (19)$$

197 Each one of these functions were employed and evaluated during the
 198 validation procedure, to select the combination of parameter γ and similarity
 199 function that would provide the system with best performance.

200 3.3. Improved Training Method

201 In addition to the original strategy proposed in [5] to determine the model-
 202 matrix, described in the end of Section 2, an alternative approach is proposed
 203 here in an attempt to reach the best compromise between the associated com-
 204 putational complexity and the resulting system fault-classification ability.

205 In the proposed approach, one selects the state vectors $\mathbf{x}_{n,c}$ to form a
 206 model-matrix \mathbf{D}_c iteratively. In each iteration a new vector is added to

207 \mathbf{D}_c taking into account the similarity between the currently selected state
208 vectors and the remaining vectors available for class c . More specifically,
209 given a vector set \mathcal{X}_c , the iterative procedure starts by selecting its geometric
210 median

$$\mathbf{v}_c = \arg \min_{\mathbf{z} \in \mathcal{X}_c} \sum_{\mathbf{x}_i \in \mathcal{X}_c} \|\mathbf{x}_i - \mathbf{z}\|_2 \quad (20)$$

211 as the first representative state. Since there is no closed form to compute this
212 median, it increases in complexity as the number of samples in \mathcal{X}_c increase.
213 In order to reduce the overall complexity, in this work we approximated the
214 vector median using the algorithm described in [34]. The subsequent states
215 that will compose \mathbf{D}_c are selected according to the following strategy: each
216 new sample $\mathbf{x}_{n,c}$ is compared against the current selected elements in \mathbf{D}_c . If
217 the similarity between the $\mathbf{x}_{n,c}$ and any element of \mathbf{D}_c is below a threshold τ ,
218 this sample is selected as an element of \mathbf{D}_c , otherwise the sample is discarded.
219 More formally, $\mathbf{x}_{n,c}$ is included in \mathbf{D}_c if

$$\mathbf{x}_{n,c} \circ \mathbf{x}_i < \tau, \quad \forall \mathbf{x}_i \in \mathbf{D}_c. \quad (21)$$

220 The effects in the SBM fault detection performance caused by the im-
221 provements proposed in this section are analyzed in the remaining of this
222 paper.

223 4. Databases Used

224 Two databases were used to evaluate the contributions of this paper.
225 The first one, named machinery fault database (MaFaulDa) [14, 27] is a
226 comprehensive database including multiple types of faults covering different
227 severities and rotation frequencies. This database was extensively used to

228 validate the proposed approach and to search for the models with the best set
229 of parameters based on their performance. The second database is the Case
230 Western Reserve University bearing database [28], the standard reference in
231 bearing faults [29, 19]. It is used to assess the proposed approach against
232 other ones found in the literature. A brief description of each database is
233 presented below.

234 *4.1. MaFaulDa Database*

235 This database is composed of multivariate time-series acquired by sensors
236 on a SpectraQuest's machinery fault simulator (MFS) alignment-balance-
237 vibration trainer (ABVT) [35]. This equipment emulates the dynamic of
238 motors with two shaft-supporting bearings and allows the study of multiple
239 faults, such as imbalanced mass, axis misalignment, and bearing problems.
240 The experimental setup used in this work is illustrated on Figure 1.

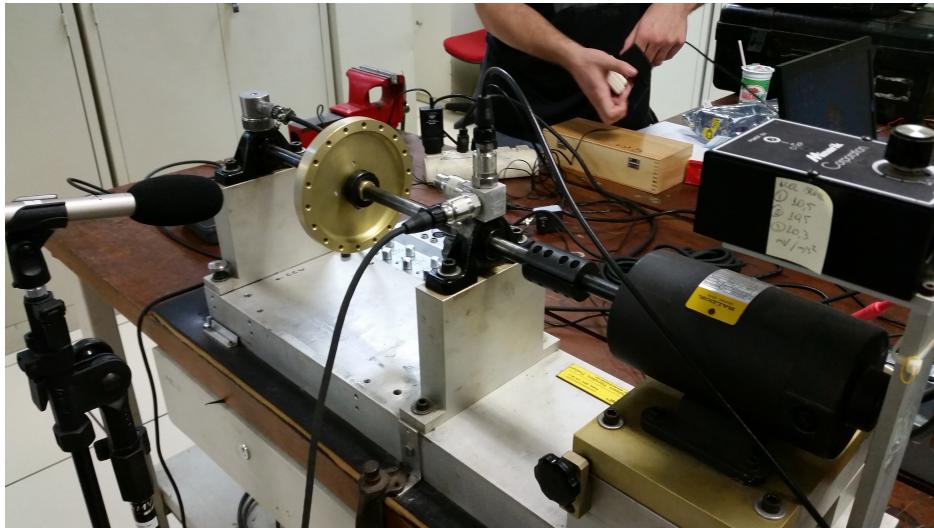


Figure 1: Experimental setup used to produce the MaFaulDa database.

241 The system was monitored by two distinct sets (one for each bearing)
242 of three accelerometers (on the axial, radial, and tangential directions), a
243 tachometer (for measuring the system rotation frequency), and a microphone
244 (for capturing the sound during the system operation). During the signal
245 acquisition procedure, a variety of faults were imposed on the MFS. These
246 faults are described below:

- 247 • **Normal operation:** this class represents the system operating under
248 normal condition without any fault. It includes a set of 49 distinct sce-
249 narios, each with a fixed rotating speed within the range from 737 rpm
250 to 3686 rpm with steps of approximately 60 rpm.
- 251 • **Imbalance:** To simulate different degrees of imbalanced operation,
252 distinct load values of 6 g, 10 g, 15 g, 20 g, 25 g, 30 g, and 35 g
253 were coupled to the rotor. For each load value below 30 g, the rota-
254 tion frequency assumed in the same 49 values employed in the normal-
255 operation case. For loads equal to or above 30 g, however, the result-
256 ing vibration makes impracticable for the system to achieve rotation
257 frequencies above 3300 rpm, limiting the number of distinct rotation
258 frequencies to only 44 in these cases. As such, the database includes a
259 total of 333 different imbalance-operation scenarios.
- 260 • **Horizontal Parallel Misalignment:** This type of fault was induced
261 into the MFS by shifting the motor shaft horizontally of 0.5 mm,
262 1.0 mm, 1.5 mm, and 2.0 mm. Using the same range for the rota-
263 tion frequency as in the normal operation for each horizontal shift, a
264 total of 197 different scenarios were considered for this class.

265 • **Vertical Parallel Misalignment:** This fault was induced into the
266 MFS by shifting the motor shaft vertically of 0.51 mm, 0.63 mm,
267 1.27 mm, 1.4 mm, 1.78 mm, and 1.9 mm. Using the same range for the
268 rotation frequency as in the normal operation for each vertical shift, a
269 total of 301 different scenarios were considered for this fault class.

270 • **Bearing faults:** As one of the most complex elements of the ma-
271 chine, the rolling bearings are the most susceptible elements to fault
272 occurrence. The ABVT manufacturer provided three defective bear-
273 ings, each one with a distinct defective element (outer track, rolling
274 elements, and inner track), that were placed one at a time in two dif-
275 ferent positions in the MFS experimental stand: between the rotor and
276 the motor (underhang position), or in the external position, having the
277 rotor between the bearing and the motor (overhang position). Bearing
278 faults are practically imperceptible when there is no imbalance. So, the
279 three masses of 6 g, 10 g, and 20 g were added to induce a detectable
280 effect, with different rotation frequencies as before, leading to a total
281 of 558 underhang scenarios and 513 overhang scenarios.

282 Considering all operating conditions described above, the MaFaulDa database
283 comprises a total of 1951 different scenarios, each one described by 8 signals
284 acquired at 50 kHz over a time interval of 5 s. The whole database is available
285 for download at [27].

286 4.2. *CWRU Bearing Database*

287 The data from this database was acquired from the bearing center of the
288 Case Western Reserve University (CWRU) [28]. It consists of 161 scenarios

grouped in four categories, as described in [29]. Each scenario can be composed of acceleration signals in three directions: on the drive-end bearing, which occurs in all scenarios; on the fan-end bearing housing, which occurs in most of the scenarios; and on the motor supporting base plate, which occurs in some scenarios. The sample rates used were 12 kHz for some scenarios and 48 kHz for others. The vibration signals were obtained from different states of the bearings: normal condition, inner race fault, ball fault, and outer race fault. A more complete description of this database is found in [28].

The CWRU bearing database was selected for two main reasons. The first one is its public availability. The second one is its wide use in the literature for reporting results of automatic bearing fault detection methods, which allows comparison of the performance of proposed method against the one of other works. In this paper only scenarios containing both the fan-bearing and the drive-end signals were used, reducing the total number of valid scenarios to 153.

5. Experimental Methodology

This section describes the experimental methodology employed to evaluate the modified SBM performance in detecting and classifying the ABVT's faulty scenarios within the databases described in Section 4.

The proposed system follows a modular architecture similar to the ones described in [2, 36] for a condition-based maintenance system. It comprises three blocks (see Figure 2): the *preprocessing* module converts the original data to a feature space which is more descriptive for the given application; the SBM model acts as a *state-monitoring* module, returning the similarity

313 between the current input data and the previously modeled conditions; and
 314 the *classifier* or *diagnostic* module uses the information from previous blocks
 315 to identify the current input among the pre-specified set of classes. In such a
 316 framework the SBM can act in a stand alone manner or can be combined to
 317 a specific classifier (*random forest* [37], for instance, as employed here). In
 318 this paper, both strategies are considered.

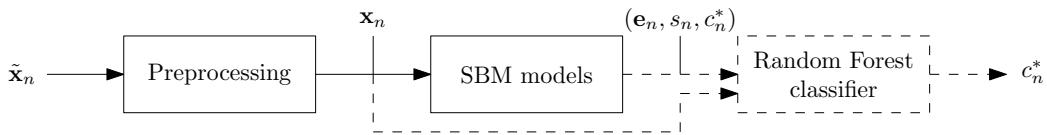


Figure 2: Block diagram of the proposed system, composed by a preprocessing module, followed by the SBM and, possibly, by a classifier.

319 The preprocessing block has the objective of reducing the original data to
 320 a set of more informative, relevant, and less redundant set of values. This is
 321 often important for reducing the system burden of learning and generalizing
 322 on the original data [38].

323 Given the distinct nature of each of the databases employed in this work,
 324 the preprocessing block should be different for each database, although its
 325 purpose is the same for both. The two preprocessing blocks are described as
 326 follows:

327 (i) *MaFaulDa*: three types of features were extracted from the original
 328 multivariate time-series: the *rotation frequency*, 21 additional *spectral*
 329 *features*, and 24 other *statistical features*.

330 The rotation frequency f_r was determined directly from the discrete
 331 Fourier transform (DFT) of the tachometer signal, as detailed in [13,

332 15]. This procedure is very similar to the one presented in [21] for
333 half-broken bars on induction motors.

334 The other spectral features correspond to the magnitudes of the spec-
335 trum of the signals other than the tachometer at frequencies f_r , $2f_r$,
336 and $3f_r$.

337 The additional statistical features include, for each of the eight mea-
338 sured signals in each operational scenario, the statistical mean, the
339 entropy, and the kurtosis. The variance feature is not employed as the
340 signals are normalized to unit variance to reduce the effect of energy
341 variations caused by changes in the acquisition setup.

342 (ii) *CWRU*: The statistical features presented in [18], together with the
343 mean, variance and entropy were extracted from each signal, totaling
344 36 features.

345 The extracted features are then input to the subsequent stages in order
346 to perform fault detection and classification. The two databases are treated
347 independently for performance assessment of the proposed methods. The
348 whole MaFaulDa database was randomly separated in two disjoint training
349 and test sets, comprising respectively 90% and 10% of the given scenarios.

350 The random choice of each set was constrained so that both presented the
351 same fault proportion as the whole database. The best set of parameters
352 was chosen using a k -fold cross-validation procedure on the training samples,
353 with $k = 10$. Then, the performance of the best models are evaluated on the
354 test set, producing the final results shown in Section 6.

355 As for the CWRU database, a process similar to cross dataset validation
356 is applied. The best setups found for the MaFaulDa are directly used on the

357 CWRU database. As such, this database is used to assess the generalization
358 power of the classifiers obtained using the proposed methodology. Results
359 have been obtained using k -folds with $k = 10$.

360 **6. Experimental Results and Discussion**

361 *6.1. Experiment Description*

362 This subsection describes the experiments made during the validation
363 procedure to select the best model for the proposed task considering all the
364 following system variations:

- 365 • Feature types: only spectral features, only statistical features, or both
366 families of features, as discussed in Section 5;
- 367 • Use of full SBM formulation (as given in Eq. (8)) or the AAKR scheme,
368 which considers $\mathbf{G} = \mathbf{I}$ in this same equation;
- 369 • Choice of similarity function, as presented in Subsection 3.2, with dis-
370 tinct values of $\gamma \in \{0.01, 0.1, 0.5, 1, 10\}$ and different ℓ_p norms ($p \in$
371 $\{1, 2\}$);
- 372 • Classification procedure either solely based on the stand-alone SBM or
373 combining it to a specific classifier algorithm (e.g. random forest);
- 374 • SBM strategy for building model-matrix \mathbf{D} : full matrix with all train-
375 ing feature vectors, original SBM training method [5] (see Section 2),
376 named “decimation method”, for decimation factors $s \in \{2, 3, 5, 7, 11\}$,
377 or proposed threshold-based method (see Subsection 3.3) for threshold
378 values $\tau \in \{0.05, 0.1, \dots, 0.95\}$.

379 Clearly the full combination of the options described above leads to a
380 prohibitively large number of possible system configurations. Therefore, in
381 this work we significantly reduce this number by presenting the validation
382 results following a sequential order of decisions, where each new decision
383 seeks an improvement on the resulting performance. These decisions can be
384 grouped in four main experiments:

- 385 • *Experiment 1* evaluates the influence of used feature types;
- 386 • *Experiment 2* compares the standard SBM model against the AAKR
387 particularization;
- 388 • *Experiment 3* evaluates the different classification strategies, investi-
389 gating whether one should use the stand-alone SBM to accomplish this
390 function or the SBM output should be fed to a classification algorithm.
391 In latter case, we also investigate the type of feature (similarity value
392 or estimation error vector (Eq. 4)) that the SBM module should deliver
393 to the subsequent classifier;
- 394 • *Experiment 4* selects the set of the remaining parameters (including
395 procedure for building the model-matrix \mathbf{D} and choice of similarity
396 function) that produces the best SBM model overall.

397 *6.2. Validation results*

398 This subsection presents the validation results using cross-validation for
399 each of the experiments. The experiments are presented in the aforemen-
400 tioned order, where the best configurations found in one experiment are car-
401 ried out to the next one.

402 *6.2.1. Experiment 1*

403 This experiment assesses the influence of the feature types in the resulting
404 validation performance of the classification system. To reduce the influence
405 of other parameters during this evaluation, a very simple system was used,
406 which employed the SBM as a classifier using the AAKR particularization,
407 and the RBF similarity function with ℓ_2 norm. The kernel width γ assumed
408 all values within the set $\{0.1, 0.5, 1.0\}$. All methods for building the SBM
409 model-matrix \mathbf{D} were assessed, with the decimation parameter fixed at $t = 5$,
410 and the threshold parameter fixed at $\tau = 0.6$.

411 Table 1 presents the obtained cross-validation results for each parameter
412 combination in Experiment 1. From this table, one can readily notice the
413 superior performance achieved by the use of all combined (spectral and sta-
414 tistical) 46 features, which is carried on to all configurations considered in
415 the subsequent experiments.

416 *6.2.2. Experiment 2*

417 This experiment compares the AAKR particularization with the standard
418 SBM model by using the same set of parameters as of Experiment 1. Results
419 from this experiment are summarized in Table 2.

420 From this table, one notices that the standard SBM outperformed its
421 AAKR particularization in all configurations considered here. Given these
422 results, the standard SBM approach was selected as the best performing
423 option to be considered in the experiments that follow.

Table 1: Experiment 1 cross-validation accuracy(%), using the AAKR particularization as a classifier with an ℓ_2 -norm RBF similarity function. f_r is the rotation frequency.

SBM model-matrix building	γ	$f_r + \text{spectral features}$	$f_r + \text{statistical features}$	All features
Full D	0.1	40.36 ± 3.28	68.27 ± 3.96	71.61 ± 3.72
	0.5	50.76 ± 3.18	76.55 ± 3.36	81.00 ± 3.07
	1.0	57.75 ± 3.18	77.0 ± 3.40	81.91 ± 2.77
Decimation [5] ($t = 5$)	0.1	37.20 ± 2.10	63.64 ± 3.55	67.53 ± 5.04
	0.5	46.60 ± 2.48	66.19 ± 4.56	72.45 ± 4.69
	1.0	49.01 ± 3.21	63.97 ± 3.88	69.50 ± 4.15
Threshold ($\tau = 0.6$)	0.1	41.02 ± 2.68	61.38 ± 2.78	67.08 ± 3.02
	0.5	50.36 ± 2.55	76.01 ± 3.28	80.81 ± 2.47
	1.0	56.70 ± 2.24	77.08 ± 3.69	81.55 ± 3.08

424 6.2.3. *Experiment 3*

425 This experiment evaluates the SBM method either as a stand-alone clas-
 426 sifier or as an auxiliary input to an off-the-shelf random forest (RF) classifier
 427 (see Figure 2). To this end, we have evaluated four different system config-
 428 urations: (i) stand-alone SBM classifier; (ii) stand-alone RF classifier; (iii)
 429 combined SBM-RF classifier using the SBM similarities to each class as a
 430 complementary feature; (iv) combined SBM-RF classifier using the SBM es-
 431 timation error vector Eq. (3) as a complementary feature. In this experiment,
 432 the SBM model used the best configurations found in the previous experi-
 433 ments, which include all 46 features previously considered and its standard
 434 SBM formulation.

Table 2: Experiment 2 cross-validation accuracy(%), comparing the SBM and AAKR.

SBM model-matrix building	γ	SBM	AAKR
Full D	0.1	84.80 ± 2.81	71.61 ± 3.72
	0.5	83.66 ± 2.45	81.00 ± 3.07
	1.0	82.50 ± 2.39	81.91 ± 2.77
Decimation [5] ($t = 5$)	0.1	78.75 ± 3.39	67.53 ± 5.04
	0.5	72.80 ± 3.26	72.45 ± 4.69
	1.0	70.29 ± 3.46	69.50 ± 4.15
Threshold ($\tau = 0.6$)	0.1	74.00 ± 2.28	67.08 ± 3.02
	0.5	82.77 ± 2.20	80.81 ± 2.47
	1.0	83.01 ± 2.40	81.55 ± 3.08

435 Table 3 presents the cross-validation results for all tested SBM-based
 436 configurations. As a basis for comparison, note that in our simulations the
 437 stand-alone RF configuration achieved an accuracy score of 92.70%.

438 These results show that the combined SBM-RF schemes are more dis-
 439 criminative than the stand-alone SBM or RF models. We can see that the
 440 case where the original features are extended by the similarities to each class
 441 estimated by the SBM produces consistently good results. However, the
 442 best results were obtained by extending the original features with the SBM
 443 estimation error vector (Eq. (3)) instead of similarities.

444 Table 3 also indicates that, when one uses the combined SBM-RF config-
 445 uration with the additional SBM estimation error features, the full model-
 446 matrix **D** is greatly outperformed by the ones built using the other two

Table 3: Experiment 3 cross-validation accuracy(%), comparing different SBM-based classifier configurations. The accuracy score of a stand-alone RF configuration is 92.70%.

SBM model-matrix building	γ	Classifier		
		SBM	RF + SBM estimation error	RF + SBM similarities
Full D	0.1	84.80 ± 2.81	44.19 ± 4.38	96.32 ± 1.66
	0.5	83.66 ± 2.45	39.66 ± 3.31	93.88 ± 1.88
	1.0	82.50 ± 2.39	41.97 ± 3.50	93.24 ± 1.39
Decimation [5] ($t = 5$)	0.1	78.75 ± 3.39	98.20 ± 1.11	96.43 ± 0.89
	0.5	72.80 ± 3.26	97.66 ± 0.91	94.98 ± 1.25
	1.0	70.29 ± 3.46	97.44 ± 0.96	94.77 ± 1.21
Threshold ($\tau = 0.6$)	0.1	74.00 ± 2.28	94.66 ± 1.19	96.59 ± 1.70
	0.5	82.77 ± 2.20	93.87 ± 0.88	94.64 ± 1.87
	1.0	83.01 ± 2.40	85.92 ± 2.56	94.02 ± 1.28

447 methods. Therefore, in Experiment 4, we consider only the decimation [5]
 448 and threshold methods for building **D**.

449 6.2.4. Experiment 4

450 This last experiment performs a fine tuning of the SBM method by se-
 451 lecting the best possible procedure for building the model-matrix **D** together
 452 with the best similarity function, including all their parameters.

453 Table 4 presents the 10 best configurations for this experiment, where the
 454 ‘Parameters’ column shows the chosen parameters for each of these scenarios.
 455 The similarity functions used are WSF (Eq. (17)), RBF (Eq. (19)), IMK
 456 (Eq. (15)), CCK (Eq. (16)), and EXP (Eq. (18)).

Table 4: Experiment 4 cross-validation accuracy(%) for the best 10 SBM configurations. The parameter of the last column is t for the decimation method for building the model-matrix and τ for the proposed threshold method.

SBM model-matrix building	Similarity function	Accuracy(%)	Parameters		
			Norm	γ	t or τ
Decimation [5]	WSF	98.91 ± 0.58	1	0.01	7
	RBF	98.02 ± 0.82	2	0.1	11
	IMK	98.62 ± 0.90	2	0.1	5
	CCK	98.56 ± 0.79	1	1	7
	EXP	98.57 ± 0.93	1	0.1	11
Threshold	WSF	98.68 ± 0.89	1	0.01	0.9
	RBF	96.66 ± 1.17	2	0.01	0.55
	IMK	98.56 ± 1.37	2	0.1	0.8
	CCK	98.72 ± 0.78	2	0.1	0.5
	EXP	98.33 ± 1.14	1	0.1	0.5

457 All the 10 models discriminated in Table 4 present superior performance
 458 than the one found in previous works using the same database [13, 14], attest-
 459 ing the SBM capability to successfully solve the fault-classification problem
 460 in rotating machines.

461 The analysis of the validation results in Table 4 leads us to choose three
 462 models as the best ones. They are:

- 463 1. Model A: Similarity function WSF, $\gamma = 0.01$, ℓ_1 norm. Using the
 464 decimation method for building \mathbf{D} [5], with $t = 7$.
- 465 2. Model B: Similarity function WSF, $\gamma = 0.01$, ℓ_1 norm. Using the

466 proposed threshold method for building \mathbf{D} , with $\tau = 0.9$.

467 3. Model C: Similarity function CCK, $\gamma = 0.1$, ℓ_2 norm. Using the pro-
468 posed threshold method for building \mathbf{D} , with $\tau = 0.5$.

469 The complexity of a given SBM model is given by its number of repre-
470 sentative states. Therefore, in order to analyze the model complexities when
471 using the three above models, we present in Table 5 the average number
472 of representative states, over the 10 validation folds, for each combination
473 of selected model and class (normal \mathbf{N} , imbalanced \mathbf{I} , horizontal misalign-
474 ment \mathbf{H}_M , vertical misalignment \mathbf{V}_M , underhang faulty bearing \mathbf{U}_B , over-
475 hang faulty bearing \mathbf{O}_B).

Table 5: Average number of representative states for each combination of model-matrix building model and class (normal \mathbf{N} , imbalanced \mathbf{I} , horizontal misalignment \mathbf{H}_M , vertical misalignment \mathbf{V}_M , underhang faulty bearing \mathbf{U}_B , overhang faulty bearing \mathbf{O}_B).

Configuration	\mathbf{N}	\mathbf{I}	\mathbf{H}_M	\mathbf{V}_M	\mathbf{U}_B	\mathbf{O}_B
Model A	33.1	87.6	65.7	80.1	131.2	118.2
Model B	5.8	73.8	8	8	94.8	29.8
Model C	5	73.5	4.9	5	74.5	6

476 From this table, one can readily draw two conclusions regarding the av-
477 erage number of representatives in each case: first, considering each model-
478 matrix building scheme, the table shows, as expected, that the proposed
479 threshold method is more selective than the decimation method. This is an
480 important result, as the proposed method requires less storage space and
481 processing time, making it well suited for deployment in a real-time. As an
482 example, Model B requires about $162 \mu s$ to process a sample, while Model A

483 requires $350 \mu s$ ¹. However, in the context of this work, this difference is irrele-
484 vant, as the preprocessing phase consumes $156 ms$ to process a single sample.
485 One should notice that the system is capable to work on real-time, since each
486 sample is $5s$ long. Second, analyzing the size of the model-matrices for each
487 failure, we observe that the imbalance failure and the underhang bearing
488 fault require larger number of states, and are thus difficult to discriminate.

489 *6.3. Results on the Testing Sets*

490 In this subsection the performance of the proposed system on the testing
491 set is analyzed. For this study, the three models chosen in Subsection 6.2.4
492 will be considered: Model A, Model B and Model C. It is important to notice
493 that the original scheme leads to a simpler model-matrix building stage but
494 to a larger matrix which results in a more complex classification procedure,
495 as observed in Table 5.

496 As mentioned on Section 5, the test dataset is composed by 10% of the
497 available MaFaulDa scenarios. Using the test dataset, Model A and Model
498 B achieved an accuracy of 98.49% and Model C achieved an accuracy of
499 97.47%, indicating that all three models are capable of generalizing well for
500 other samples. The confusion matrices for the first two models are shown in
501 Tables 6a and 6b.

502 Results in these tables are consistent with some already discussed as-
503 pects of the MaFaulDa database. As stated in Section 4, there are much
504 less scenarios when the machinery operates on normal conditions than there
505 are faulty cases. This discrepancy makes the model-matrix building more

¹All times were measured on an Intel(R) Core(TM) i7-4790K CPU @ 4.00GHz machine.

Table 6: Confusion Matrices in Test Dataset.

(a) Model A.

Class	N	I	H _M	V _M	U _B	O _B
N	4	0	1	0	0	0
I	0	34	0	0	0	0
H _M	0	0	20	0	0	0
V _M	0	0	0	31	0	0
U _B	0	1	0	0	55	0
O _B	0	1	0	0	0	51

(b) Model B.

Class	N	I	H _M	V _M	U _B	O _B
N	5	0	0	0	0	0
I	0	34	0	0	0	0
H _M	0	0	19	1	0	0
V _M	0	0	0	31	0	0
U _B	0	1	0	0	54	1
O _B	0	0	0	0	0	52

506 difficult for the normal class (N). Still analyzing the confusion matrices, it
 507 is possible to observe that sometimes bearing faults are misclassified as im-
 508 balance faults. We argue that this is somewhat expected, since the system
 509 needs to be unbalanced in order for bearing faults to be observed.

510 Also in Section 4, when the bearing faults are described, it was mentioned
 511 that each one of the bearings (underhang and overhang) where substituted
 512 by one out of three defective bearings provided by the manufacturer. Taking
 513 this into consideration, the three best configurations, Model A, Model B,
 514 and Model C, were also used to classify the signals in 10 classes. These
 515 classes were derived by further subdividing each bearing fault in 3 classes
 516 according to the defective element (outer race, inner race, or rolling ball)
 517 employed. The results are presented in Table 7, where the good accuracy
 518 figures indicate that the proposed system is also robust when applied to
 519 more complex fault classification problems.

Table 7: Accuracy results for the 10-class identification problem on the MaFaulDa database

Model	Model-matrix building method	Similarity function	γ	t or τ	p	Acc. (%)
Model A	Decimation	WSF	0.01	7	2	98.48
Model B	Threshold	WSF	0.01	0.9	1	98.48
Model C	Threshold	CCK	0.1	0.5	2	97.48

520 *6.4. CWRU Results and Discussion*

521 This subsection presents the results on the CWRU bearing dataset. As
 522 described in Section 5, this dataset is used for assessing the performance of
 523 the three best models selected on the MaFaulDa dataset, namely the Model
 524 A, Model B, and Model C schemes (see Subsection 6.2.4).

525 Using the same methodology as [18], each CWRU signal was divided
 526 into 15 segments, and the extended dataset was subdivided into the training
 527 and testing sets following a 9/1 ratio. The results presented in Table 8 are
 528 accuracy averages over 10 folds chosen randomly. For each fold configuration,
 529 the model-matrix \mathbf{D} is computed using the data in 9 folds and the accuracy
 530 result is measured in the remaining fold. From this table, one can observe
 531 that the SBM-based classifier has good generalization capability for all three
 532 configurations considered here.

533 *6.5. Comparison with Previous Works*

534 As mentioned in Section 1, several other works in the literature addressed
 535 the same problem that we have addressed in this paper, that is, the auto-
 536 matic detection and classification of faults in rotating machines. Some of

Table 8: Accuracy (%) results of SBM-based classifiers on the CWRU database.

Model	Mean
Model A	98.95 ± 0.72
Model B	98.91 ± 0.75
Model C	98.91 ± 0.95

these works have used the MaFaulDa database. In [14] the faults in the MaFaulDa database have been classified using perceptron neural networks with multiple layers, considering several subsets of the features investigated here. Six classes have been considered: normal, overhang and underhang faults, imbalance, horizontal and vertical misalignment. The accuracy obtained was 95.8%, inferior to the ones obtained with the proposed use of SBM and described in Table 7, that reach, for one configuration, the average figure of 98.48%.

For the CWRU database, even though there are many works using such dataset [29], its very difficult to make a direct comparison, as most works do not present their results in a quantitative manner, but only in a qualitative manner. As such, the comparison is restricted to a small set of works. In [25] the k NN, naive Bayes, and SVM classifiers achieved accuracies of 98.83%, 98% and 98.97%, respectively. The SVM classifier found in [26] obtained accuracies above 98% for different rotation frequencies. The SVM and ELM classifiers using the procedure described in [30] achieved accuracies of 82.4% and 97.5%, respectively. Lastly, the k NN, SVM, and ANN classifiers using the feature selection method proposed in [18] obtained accuracies between 93% and 100%. From the above results and Table 8, one can conclude that

556 the proposed SBM-based fault classifier achieves, for the CWRU database,
557 competitive results to those. It is important to point out that, as demon-
558 strated by the results over the MaFaulDa database, the proposed system
559 is able to detect and classify, with high accuracy, a wide range of machine
560 faults, including misalignment and unbalanced faults.

561 **7. Conclusion**

562 This paper addressed the automatic fault diagnosis in rotating machines.
563 The use of similarity based modeling (SBM) was investigated, either as a
564 stand-alone classification method or in combination with an off-the-shelf clas-
565 sifier, in this case a random forest classifier. The system is evaluated in two
566 databases. One of them is a comprehensive database with multiple faults re-
567 ferred to as MaFaulDa [27]. The other is the CWRU bearing database [28],
568 that is the current standard database for bearing fault diagnosis.

569 The extension of the SBM to a multiclass model is an important contri-
570 bution of this paper. Other contributions include the use of new similarity
571 metrics and the development of a novel method for building the SBM model-
572 matrices. This work also presented an extensive study of the evaluation of all
573 the proposed modifications on the MaFaulDa database. These contributions
574 achieved the goal of increasing the SBM performance in a fault classification
575 scenario while reducing its computational complexity. The usage of SBM
576 either as a stand-alone classifier or as a feature generator for off-the-shelf
577 classifiers has also been investigated. Our results have shown that the use of
578 the proposed enhancements to the SBM consistently increased the accuracy
579 of a random forest classifier.

580 The proposed system showed to be robust, reaching an accuracy of around
581 98.5% in the MaFaulDa database, higher than previous works along the same
582 base. For the CWRU dataset the proposed system yielded an accuracy level
583 of 98.9%, which is as good as previous results reported in the literature. It is
584 worth emphasizing that the proposed class of methods is able to detect and
585 classify, with high accuracy, a wide range of faults, which indicates that the
586 proposed approach based on SBM is worth further investigation.

587 **Acknowledgements**

588 This research was supported by CNPq and Petrobras. The authors would
589 like to thank Dr. Kenneth A. Loparo and the Case Western Reserve Uni-
590 versity Bearing Data Center for providing the CWRU dataset for this study,
591 and Dr. Wade A. Smith for his help.

592 **References**

- 593 [1] J. P. Herzog, D. Gandhi, B. Nieman, Making decisions about the best
594 technology to implement, Technical report, Smartsignal/GE Intelligent
595 Platforms (2011).
- 596 [2] J. B. Coble, Merging data sources to predict remaining useful life –
597 an automated method to identify prognostic parameters, Ph.D. Thesis,
598 University of Tennessee (2010).
- 599 [3] J. Mott, M. Pipke, Similarity-based modeling of aircraft flight paths, in:
600 IEEE Proc. Aerospace Conference, Vol. 3, 2004.

- 601 [4] R. M. Singer, K. C. Gross, J. P. Herzog, R. W. King, S. Wegerich,
602 Model-based nuclear power plant monitoring and fault detection: Theo-
603 retical foundations, in: International Conference on Intelligent Systems
604 Applications to Power Systems, 1997.
- 605 [5] J. P. Herzog, S. W. Wegerich, K. C. Gross, F. K. Bockhorst, MSET
606 modeling of Crystal River-3 venturi flow meters, in: International Con-
607 ference on Nuclear Engineering, 1998.
- 608 [6] J. Garvey, D. Garvey, R. Seibert, J. W. Hines, Validation of on-line
609 monitoring techniques to nuclear plant data, Nuclear Engineering and
610 Technology 39 (2) (2006) 133–142.
- 611 [7] F. A. Tobar, L. Yacher, R. Paredes, M. E. Orchard, Anomaly detection
612 in power generation plants using similarity-based modeling and multi-
613 variate analysis, in: Proc. American Control Conference, Vol. 3, 2011.
- 614 [8] P. Guo, N. Bai, Wind turbine gearbox condition monitoring with AAKR
615 and moving window statistic methods, Energies 4 (11) (2011) 2077–2093.
- 616 [9] S. W. Wegerich, A. D. Wilks, R. M. Pipke, Nonparametric modeling
617 of vibration signal features for equipment health monitoring, in: IEEE
618 Proc. Aerospace Conference, Vol. 7, 2003.
- 619 [10] S. W. Wegerich, Similarity-based modeling of time synchronous aver-
620 aged vibration signals for machinery health monitoring, in: IEEE Proc.
621 Aerospace Conference, Vol. 6, 2004.
- 622 [11] S. W. Wegerich, Similarity based modeling of vibration features for fault
623 detection and identification, Sensor Review 25 (2) (2005) 114–122.

- 624 [12] J. Liu, W. Wang, F. Golnaraghi, An enhanced diagnostic scheme for
625 bearing condition monitoring, *IEEE Transactions on Instrumentation*
626 and *Measurement* 59 (2) (2010) 309–321.
- 627 [13] A. A. de Lima, T. d. M. Prego, S. L. Netto, E. A. B. da Silva, R. H. R.
628 Gutierrez, U. A. Monteiro, A. C. R. Troyman, F. J. d. C. Silveira, L. Vaz,
629 On fault classification in rotating machines using Fourier domain fea-
630 tures and neural networks, in: *Proc. Latin American Symposium on*
631 *Circuits and Systems*, 2013.
- 632 [14] D. Pestana-Viana, R. Zambrano-López, A. A. de Lima, T. d. M. Prego,
633 S. L. Netto, E. A. B. da Silva, The influence of feature vector on the
634 classification of mechanical faults using neural networks, in: *Proc. Latin*
635 *American Symposium on Circuits and Systems*, 2016.
- 636 [15] P. Li, F. Kong, Q. He, Y. Liu, Multiscale slope feature extraction for
637 rotating machinery fault diagnosis using wavelet analysis, *Measurement*
638 46 (19) (2013) 497–505.
- 639 [16] R. B. Randall, J. Antoni, Rolling element bearing diagnostics – a tuto-
640 rial, *Mechanical Systems and Signal Processing* 25 (2) (2011) 485–520.
- 641 [17] J. H. Zhou, L. Wee, Z. W. Zhong, A knowledge base system for rotary
642 equipment fault detection and diagnosis, in: *Proc. Control Automation*
643 *Robotics Vision*, 2010.
- 644 [18] T. W. Rauber, F. de Assis Boldt, F. M. Varejão, Heterogeneous feature
645 models and feature selection applied to bearing fault diagnosis, *Mechan-
646 ical Systems and Signal Processing* 62 (1) (2015) 637–646.

- 647 [19] A. Boudiaf, A. Moussaoui, A. Dahane, I. Atoui, A comparative study
648 of various methods of bearing faults diagnosis using the Case Western
649 Reserve University Data, *Journal of Failure Analysis and Prevention*
650 16 (2) (2016) 271–284.
- 651 [20] B. Yang, T. Han, J. An, ART-KOHONEN neural network for fault diag-
652 nosis of rotating machinery, *Mechanical Systems and Signal Processing*
653 18 (3) (2004) 645–657.
- 654 [21] J. de Jesus Rangel-Magdaleno, , R. de Jesus Romero-Troncoso, R. A.
655 Osornio-Rios, E. Cabal-Yepez, L. M. Contreras-Medina, Novel method-
656 ology for online half-broken-bar detection on induction motors, *IEEE*
657 *Transactions on Instrumentation and Measurement* 58 (5) (2009) 1690–
658 1698.
- 659 [22] L. M. Contreras-Medina, R. de Jesus Romero-Troncoso, E. Cabal-
660 Yepez, J. de Jesus Rangel-Magdaleno, J. R. Millan-Almaraz, Fpga-based
661 multiple-channel vibration analyzer for industrial applications in induc-
662 tion motor failure detection, *IEEE Transactions on Instrumentation and*
663 *Measurement* 59 (1) (2010) 63–72.
- 664 [23] J. de Jesus Rangel-Magdaleno, R. de Jesus Romero-Troncoso, R. A.
665 Osornio-Rios, E. Cabal-Yepez, A. Dominguez-Gonzalez, Fpga-based vi-
666 bration analyzer for continuous cnc machinery monitoring with fused
667 fft-dwt signal processing, *IEEE Transactions on Instrumentation and*
668 *Measurement* 59 (12) (2010) 3184–3194.
- 669 [24] J. de Jesus Rangel-Magdaleno, H. Peregrina-Barreto, J. M. Ramirez-

- 670 Cortes, P. Gomez-Gil, R. Morales-Caporal, Fpga-based broken bars de-
671 tector on induction motors under different load using motor current
672 signature analysis and mathematical morphology, IEEE Transactions
673 on Instrumentation and Measurement 63 (5) (2014) 1032–1040.
- 674 [25] B. Li, P. Zhang, D. Liu, S. Mi, G. Ren, H. Tian, Feature extraction
675 for rolling element bearing fault diagnosis utilizing generalized S trans-
676 form and two-dimensional non-negative matrix factorization, Journal of
677 Sound and Vibration 330 (10) (2011) 2388–2399.
- 678 [26] S.-D. Wu, P.-H. Wu, C.-W. Wu, J.-J. Ding, C.-C. Wang, Bearing fault
679 diagnosis based on multiscale permutation entropy and support vector
680 machine, Entropy 14 (8) (2012) 1343–1356.
- 681 [27] MaFaulDa - Machinery Fault Database, <http://www02.smt.ufrj.br/~offshore/mfs/>, accessed November 22, 2016 (2016).
- 683 [28] K. A. Loparo, Bearings vibration data set, Case Western Reserve Uni-
684 versity, <http://csegroups.case.edu/bearingdatacenter/home>, ac-
685 cessed November 25, 2016 (2003).
- 686 [29] W. A. Smith, R. B. Randall, Rolling element bearing diagnostics us-
687 ing the Case Western Reserve University data: A benchmark study,
688 Mechanical Systems and Signal Processing 64-65 (2015) 100–131.
- 689 [30] Y. Li, X. Wang, J. Wu, Fault diagnosis of rolling bearing based on
690 permutation entropy and extreme learning machine, in: Proc. Chinese
691 Control and Decision Conference, 2016.

- 692 [31] G. H. Golub, C. F. V. Loan, Matrix Computations, 3rd Edition, Johns
693 Hopkins University Press, Baltimore, MD, USA, 1996.
- 694 [32] C. A. Micchelli, Interpolation of scattered data: Distance matrices and
695 conditionally positive definite functions, Constructive Approximation
696 2 (1) (1986) 11–22.
- 697 [33] J. Basak, A least square kernel machine with box constraints, in: Interna-
698 tional Conference on Pattern Recognition, 2008.
- 699 [34] H. Cardot, P. Cénac, P.-A. Zitt, Efficient and fast estimation of the
700 geometric median in hilbert spaces with an averaged stochastic gradient
701 algorithm, Bernoulli 19 (1) (2013) 18–43.
- 702 [35] SpectraQuest, Inc., <http://www.spectraquest.com/>, accessed
703 November 27, 2016 (2016).
- 704 [36] G. Palem, Condition-based maintenance using sensor arrays and telem-
705 atics, International Journal of Mobile Network Communications &
706 Telematics 3 (3) (2013) 19–28.
- 707 [37] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion,
708 O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vander-
709 plas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay,
710 Scikit-learn: Machine learning in Python, Journal of Machine Learning
711 Research 12 (2011) 2825–2830.
- 712 [38] Y. S. Abu-Mostafa, M. Magdon-Ismail, H.-T. Lin, Learning from data,
713 AMLBook, US, 2012.