An Evidential Evolving Prognostic Approach and its Application to PRONOSTIA's Data Streams

Lisa Serir, Emmanuel Ramasso, and Noureddine Zerhouni
FEMTO-ST Institute, UMR CNRS 6174 - UFC / ENSMM / UTBM,
Automatic Control and Micro-Mechatronic Systems Dep., 25000, Besançon, France
(lisa.serir, emmanuel.ramasso, noureddine.zerhouni)@femto-st.fr

ABSTRACT

The research activity in the PHM community is in full bloom and many efforts are being made to develop more realistic and reliable methodologies. However, there still exist very few real-world applications due to the complexity of the systems of interest. Nonlinear dynamical systems identification and behavior prediction are difficult problems encountered in prognosis. The difficulty in switching from theory to practice can partially be explained by the existence of different kinds of uncertainty at each step of the implementation that must be taken into account with the appropriate tools. In this paper, we propose an evolving multi-modeling approach for the detection, the adaptation and the combination of local models in order to analyze complex systems behavior. It relies on belief functions in order to take into consideration the uncertainty related to the available data describing the system as well as the uncertainty generated by the nonlinearity of the system. The information of doubt explicitly represented in the belief functions framework is exploited to properly segment the data and take into account the uncertainty related to the transitions between the operating regions. The proposed algorithm is validated on a data provided by PRONOSTIA platform.

1. INTRODUCTION

Although prognosis is acknowledged as a key element in industrial maintenance strategies, there still exist very few real-world applications due to the complexity of the systems of interest. The implementation of a data-driven PHM tool in real conditions requires a learning phase for the estimation of the different parameters, that represents one of the main barriers to its applicability, as it is very costly to collect data, in particular for the failure modes. Modeling the behavior of nonlinear systems from observed data is a difficult task to perform widely encountered in prediction and fault diagnosis (Angelov, Filev, & Kasabov, 2010). One way to overcome the complexity related to nonlinearity is to adopt multi-model approaches (Madani, Rybnik, & Chebira, 2003; Boukhris, Mourot, & Ragot, 2000; Murray-Smith, 1997; Nelles, 1995) by considering that the system's behavior gradually evolves along the operating range. Thus, the system could locally be described by simple functions corresponding to some operating regions. The description of the global system's behavior is then made by the combination of the local models, each of which being weighted by an *activation degree*. The identification consists in two main steps: the structural identification to determine the number of models, and a parametric identification to evaluate the parameters of the local models.

Understanding the response of nonlinear systems is a very challenging task leading to significant uncertainty. Uncertainty sources are numerous and may take the form of system variability, environmental and operational conditions, data acquisition errors, among other sources that vary depending on the application at hand. The quantity as well as the quality of the data are also very important aspects to take into account. Moreover, the segmentation of the data into meaningful modes is often required to help the methods of parameters' estimation to converge towards reasonable solutions. These constraints on the data are not clearly considered in the literature and we believe that the uncertainty related to the lack of data for the detection and prediction of failure modes must be quantified to understand the capabilities and limitations of the modeling process. While probability theory is well suited to deal with aleatory uncertainties (intrinsic variability), other formalisms exist that are more appropriate to manage epistemic (imprecision and incompleteness) uncertainty (G. J.Klir, 2006), among which, fuzzy sets or possibility theory and evidence theory, also known as belief functions theory (Shafer, 1976) which are the most prominent ones.

Fuzzy set theory has been used to deal with imprecision within data in multi-modeling approaches (Chandrashekhar & Ganguli, 2009; Haag, Herrmann, & Hanss, 2010), among which the fuzzy rule-based models of Takagi-Sugeno (TSK) type (Takagi & Sugeno, 1985), widely used in modeling applications of complex systems, due to their flexibility and

computational efficiency. Their main advantage dwells in the fact that since the local regions are fuzzily defined, the resulting global model can be nonlinear, while the local models can be very simple (linear) (Takagi & Sugeno, 1985; Angelov, Lughofer, & Zhou, 2008). This kind of approach has been applied to build a neuro-fuzzy predictor in the context of prognosis application in (El-Koujok, Gouriveau, & Zerhouni, 2011). It was based on the evolving extended Takagi-Sugeno system (exTS) proposed by Angelov (Angelov & Filev, 2004).

Ramdani *et al.* (Ramdani, Mourot, & Ragot, 2005) exploited the theoretical framework of belief functions to deal with uncertainties in multi-modeling founded on a TSK fuzzy model. The main advantage of this approach remains in the use of belief functions theory to determine the activation degrees of the local models. However, Their proposed methodology is an *offline* approach and requires the entire dataset to be available in advance for the modeling process. Belief functions are particularly interesting because they obviate the need to introduce unjustified a priori thanks to the relaxation of the additivity constraint, which makes possible the explicit representation of doubt. The work of Smets (Smets & Kennes, 1994) has shown that the information brought by the empty set could enable to take into account "open world" situations. Although many publications using belief functions theory have recently increased, very few are related to PHM applications.

One of the constraints to address in PHM applications is related to the fact that the data concerning the failure modes are generally rare compared to the data of normal functioning modes. This can be seen as a problem of unbalanced data (He & Garci, 2009). A solution is to construct an evolving algorithm that can adapt to the lack of data. This kind of approach enables the consideration of applications for which historical data cannot be stored. Then, for more coherence and to facilitate the implementation of the prognostic approach, the detection and the prediction must be performed within the same algorithm.

In this paper, we propose to adapt the offline approach developed in (Ramdani et al., 2005) to the online case, where the data arrive gradually. In the sequel, We qualify the proposed approach as "evolving", as it is able to adapt its parameters online. The proposed algorithm is called E2GKpro and relies on the Evidential Evolving Gustafson-Kessel algorithm (E2GK) initially developed in (Serir, Ramasso, & Zerhouni, 2012) to sequentially perform the clustering phase using the formalism of belief functions. The clustering is then followed by the online identification of local linear models. The activation degrees of each local model are directly provided by the E2GK algorithm in the form of belief masses, and the global model is a combination of all the local models. This work is presented as an improved version of a previous contribution (Serir, Ramasso, & Zerhouni, 2011).

The paper is organized as follows. First, the basics of belief functions theory are given and illustrated on a simple hypothetical example (Section 2). Then, the proposed E2GKpro algorithm is detailed (Section 3). And finally, the algorithm is applied on the data of PRONOSTIA platform (Section 4).

2. BACKGROUND MATERIALS ON BELIEF FUNCTIONS THEORY

The theory of belief functions, also called theory of evidence or Dempster-Shafer theory, is a formal framework for reasoning with partial (imprecise and uncertain) information. It was introduced by Dempster (1968) and Shafer (1976) and later developed by Smets and others. The theory of belief functions extends both the set-membership (intersection, union, inclusion, etc.) and probabilistic (conditioning, marginalization) approaches to uncertain reasoning, and a belief function may be viewed as both a generalized set and a non additive measure. Smets has developed the *Transferable Belief Model* (TBM) (Smets & Kennes, 1994) as a general framework for uncertainty representation and combination of various pieces of information without additional prior. In particular, TBM offers the possibility to explicitly emphasize doubt, representing ignorance, and conflict, emphasizing the contradiction in a fusion process. We give in this section some of the basic notions of the theory and refer the reader to (Smets & Kennes, 1994) for a more complete description.

Let ω be a variable taking values in a finite set Ω, called *frame of discernment*, and let 2^Ω be its power set. A belief function on Ω is mathematically defined by introducing a set function m, called the basic belief assignment (BBA) defined by:

$$\begin{array}{rccc} m: & 2^\Omega & \rightarrow & [0,1] \\ & A & \mapsto & m(A) \end{array}, \tag{1}$$

and satisfies: $\sum_{A\subseteq\Omega} m(A) = 1$. Each subset $A \subseteq \Omega$ such as $m(A) > 0$ is called a *focal elements* of m.

We collect a piece of evidence (information) about ω, which can have different interpretations weighted by subjective probabilities. One of these interpretations can mean that we only know that $\omega \in A$, for some $A \subseteq \Omega$, and nothing more. The probability that the evidence means exactly that $\omega \in A$ is $m(A)$. A BBA is assigned not only to singletons ($|A| = 1$), but also to *subsets* of Ω ($|A| > 1$), *without assumption concerning additivity* of the measure m (G. J.Klir, 2006). This property permits the explicit modeling of doubt and conflict which constitutes a fundamental difference with probability theory:

- Perfect knowledge of the value of $\omega \in \Omega$ is represented by the allocation of the unit of belief to ω: $m(\omega) = 1$. In this case, m is said to be *certain*.
- Complete ignorance corresponds to $m(\Omega) = 1$.
- In the case of all focal elements being singletons, m boils down to a probability function and is said to be *Bayesian*.

- The value of $m(\emptyset)$ is called *conflict* and can be positive, meaning that one accepts the *open-world assumption* stating that the set Ω might not be complete, and thus ω might take its value outside Ω.
- In the case of all focal elements being singletons, m boils down to a probability function and is said to be *Bayesian*.

A BBA m is said to be *normal* if $m(\emptyset) = 0$. A normalized BBA m^* can be computed from a BBA m by applying the Dempster *normalization*:

$$m^*(A) = \begin{cases} \dfrac{m(A)}{1 - m(\emptyset)} \text{ if } A \neq \emptyset \\ 0 \text{ otherwise} \end{cases} \quad (2)$$

Given a BBA m, a belief function bel and a plausibility function pl can be defined, respectively, as

$$bel(A) = \sum_{\emptyset \neq B \subseteq \Omega} m(B), \ \forall A \subseteq \Omega \ , \quad (3)$$

$$pl(A) = \sum_{B \cap A \neq \emptyset} m(B), \ \forall A \subseteq \Omega \ , \quad (4)$$

$bel\,(A)$ can be interpreted as a measure of our total belief committed to A after receiving the item of evidence, and $pl(A)$ represents the maximal degree of belief supporting the subset A (Cobb & Shenoy, 2006). If m is Bayesian, $pl = bel$ boils down to a probability measure. Note that a possibility measure is known to be formally equivalent to a *consonant* belief function, i.e., a belief function with nested focal elements (Dubois, Prade, & Smets, 2001). The functions bel, pl and m are in one-to-one correspondence and represent three facets of the same piece of information.

The TBM distinguishes the credal level where beliefs are formalized, revised and combined, and the *pignistic* level used for decision making, which consists in the choice of the best hypothesis using the *pignistic probability* distribution (Smets & Kennes, 1994) defined as:

$$\mathbf{BetP}(\omega) = \sum_{A \ni \omega} \frac{m(A)}{|A|} \frac{1}{1 - m\,(\emptyset)}, \quad \forall \omega \in \Omega \ . \quad (5)$$

where each mass of belief $m(A)$ is equally distributed among the elements of A and $\mathbf{BetP}(A) = \sum_{\omega \in A} \mathbf{BetP}(\omega), \forall A \subseteq \Omega$.

Example 1 *Let us consider the example described in Fig. 1. Suppose that the figure results from a monitoring of a bearing system, and that it represents the evolution along time of a feature provided by a sensor (the available piece of evidence). There are* 4 *possible operating modes: Normal (N), Degrading 1 (D1), Degrading 2 (D2) and Faulty (F) (*$\Omega = \{N, D1, D2, F\}$*). One can notice that the feature evolution curve can be segmented into* 4 *clear operating regions (arrows in magenta color). However, in the transitions (magenta ellipses) from one operating region to another (*$N \rightarrow D1, D1 \rightarrow D2, D2 \rightarrow F$*), one can express doubt regarding whether the bearing is still in the same operating mode or if it had evolved to a new one. One's belief regarding the operating mode of the bearing along time can "consecutively" be expressed by:* $m(N) = 1$, $m(\{N, D1\})$, $m(D1) = 1$, $m(\{D1, D2\})$, $m(D2) = 1$, $m(\{D2, F\})$ *and* $m(F) = 1$.

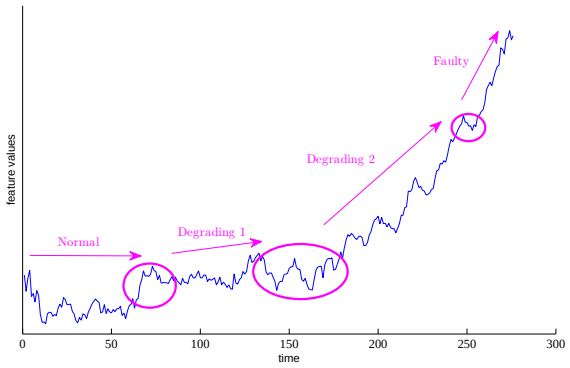

Figure 1. An example of basic belief assignments.

3. An Evidential Evolving Prediction Methodology

The proposed methodology, called E2GKpro, follows a three-phase scheme, commonly used in the evolving systems modeling approaches (Angelov et al., 2010). The novelty of E2GKpro dwells in the use of belief functions for the determination of the local models and the estimation of their activation degrees.

The system under study is supposed to gradually evolve through different operating modes (Fig. 1), each of which corresponding to unknown local linear models to be identified online. At the current time-instant k, a n-dimensional input feature vector $x_k = [x_1 \dots x_n]^{\mathrm{T}} \in \mathbb{R}^n$ and an output $y_k \in \mathbb{R}$ are observed. After an initialization phase, given (x_k, y_k), E2GKpro starts with a clustering phase to detect the current operating regions of the system. Then, a regression phase is performed to update the local linear models corresponding to each cluster. Finally, the prediction phase estimates the value of the output $\hat{y}_k$ given the input x_k.

3.1. Initialization

To explain the belief functions generation process, let consider the existence of an initial set of available data so that two operating regions can be identified by applying a standard clustering algorithm on the input-output space. Given an initial observed datum $z_k = (x_k, y_k)$, the Gustafson-Kessel algorithm is used, with $c = 2$, to find the coordinates $v_{i,k}, i = 1, 2$ of the two first clusters and to initialize their covariance matrices $\Sigma_{i,k}, i = 1, 2$.

3.1.1. Computing the evidential partition

Let Ω be the set of the so far existing clusters. Given the centers and the covariance matrices, the degree of belief regarding the membership of the k-th data point to each possible subset of clusters $A_i \in 2^{\Omega} \backslash \emptyset$, can be computed as in the evidential evolving clustering algorithm (E2GK) (Serir et al., 2012):

$$m_{ik} = \frac{|A_i|^{-1} \cdot d_{ik}^{-2}}{\sum_{A_l \neq \emptyset} |A_l|^{-1} \cdot d_{lk}^{-2} + \delta^{-2}} \ , \tag{6}$$

and the mass assigned to the emptyset is equal to:

$$m_{\emptyset k} = 1 - \sum_{A_i \neq \emptyset} m_{ik} \ , \tag{7}$$

where d_{ik} denotes the Mahalanobis-like distance between data point z_k and subset A_i, and $\delta \in \mathbb{R}^+$ controls the amount of data considered as outliers. the distance d_{ik} is computed by first defining the center of A_i as the barycenter $\overline{v}_{i,k}$ of clusters' centers composing A_i 2. The corresponding covariance matrix $\Sigma_{i,k}$ can then be computed as:

$$\Sigma_{i,k} = \frac{1}{N-1} \cdot \sum_{k=1}^{N} \sum_{A_j \ni \omega_i} (z_k - \overline{v}_{i,k}) \cdot (z_k - \overline{v}_{i,k})^{\mathrm{T}} \ , \tag{8}$$

and the distance $d_{i,k}$ is given by:

$$d_{ik}^2 = \|z_k - \overline{v}_{i,k}\|_{S_i}^2 = (z_k - \overline{v}_{i,k}) \cdot S_i \cdot (z_k - \overline{v}_{i,k})^{\mathrm{T}} \ , \tag{9a}$$

$$S_{i,k} = [\det(\Sigma_{i,k})]^{1/n} \cdot \Sigma_{i,k}^{-1} \ , \tag{9b}$$

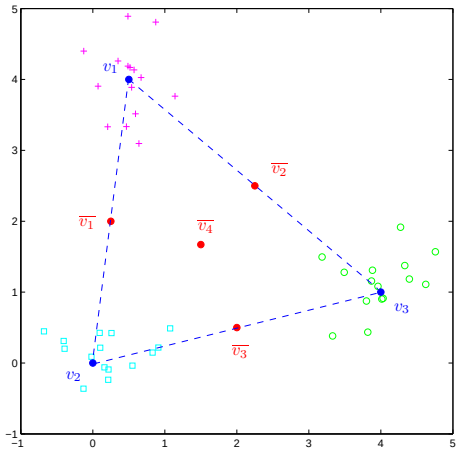

Figure 2. Centers of subsets of clusters in Ω.

Remark 1 *Only singleton focal elements (clusters $\omega_k \in \Omega$) are associated with centroids but the particularity of the BBA computation holds in the consideration of virtual centroids located at the barycenters of subsets of clusters (Masson & Denoeux, 2008). This makes the proposed approach different from usual possibilistic and probabilistic clustering approaches.*

3.1.2. Initializing the local models

Let X_i be the set of input points belonging to the i-th cluster, Y_i the corresponding outputs, and θ_i the parameters of the local linear model for the i-th cluster that can be optimized by a standard least squared approach:

$$\theta_i = \left(X_i^{\mathrm{T}} \cdot X_i\right)^{-1} \cdot X_i^{\mathrm{T}} \cdot Y_i \tag{10}$$

After completing the initialization, E2GKpro can be run online: if the observed data is an input-output couple z_k then the clustering and the local models can be updated, whereas if only the input data x_k is observed, a prediction $\hat{y}_k$ of the output can be estimated.

3.2. Online clustering phase

This phase aims at detecting the current operating regions of the system. The number of clusters c can evolve, in particular when a new operating region is detected. At the current instant k, each cluster $i = 1 \dots c$ is identified by two parameters: a center $v_{i,k} \in \mathbb{R}^n$ and a covariance matrix $\Sigma_{i,k} \in \mathbb{R}^n \times \mathbb{R}^n$, both adapted according to (x_k, y_k). This phase relies on the evidential evolving clustering algorithm called E2GK (Serir et al., 2012).

When a new input-output datum z_k is observed, the data structure can be updated. The boundary of each cluster i is first estimated by computing its radius: The radius r_i of the i-th cluster is computed by:

$$r_i = \underset{\forall z_k \in \ i\text{-th cluster and } \ \lambda_{ik} > c^{-1}}{\text{median}} \|z_k - v_{i,k}\|_{S_i} \ , \tag{11}$$

where λ_{ik} is the confidence degree that point x_k belongs to a singleton cluster ($\omega_i \in \Omega$) estimated by the pignistic transformation (Eq. 5) (Serir et al., 2012):

$$\lambda_{i,k} = \sum_{A_j, \omega_i \in A_j} \frac{m_{j,k}}{|A_j|} \tag{12}$$

The closest cluster to z_k is then found by:

$$p = \underset{i=1}{\overset{c}{\operatorname{argmin}}} \, d_{ik}^2 \tag{13}$$

Once cluster p is found, two cases are possible:

- Case 1: $d_{pk}^2 \leq r_p$, i.e. z_k belongs to an existing cluster, inducing an update of the p-th cluster:

$$v_{p,k+1} = v_{p,k} + \theta \cdot \Delta \ , \tag{14}$$

where $\Delta = z_k - v_{p,k}$ and $\theta \in [0, 1]$ is the updating rate. The inverse of the covariance matrix and its determinant can be recursively adapted by (Georgieva & Filev, 2009):

$$\Sigma_{p,k+1}^{-1} = \left(I - \frac{\theta \cdot \Sigma_{p,k}^{-1} \cdot \Delta^{\mathrm{T}} \cdot \Delta}{\xi} \right) \cdot \Sigma_{p,k}^{-1} \cdot \frac{1}{1-\theta}, \tag{15}$$

where I is the identity matrix and

$$\xi = 1 - \theta + \theta \cdot \Delta \cdot \Sigma_{p,k}^{-1} \cdot \Delta^{\mathrm{T}} \tag{16}$$

and

$$\det\left(\Sigma_{p,k+1}\right) = (1-\theta)^{n-1} \cdot \det\left(\Sigma_{p,k}\right) \cdot \xi. \tag{17}$$

The partition matrix can then be computed (Eq. 9b, 6 and 7).

- Case 2: $d_{pk}^2 > r_p$, i.e. z_k is too far from any existing cluster, involving the creation and the validation of a new cluster: the number of clusters is incremented and the incoming data z_k is chosen as the center of the new cluster, whose parameters are initialized as those of the closest cluster p:

$$c \leftarrow c + 1 \ , \tag{18a}$$

$$v_{c,k+1} = z_k \ , \tag{18b}$$

$$\Sigma_{c,k+1} = \Sigma_{p,k} \ . \tag{18c}$$

In (Serir et al., 2012), a constraint of validity of the partition was added by imposing that each cluster in the new partition has a minimum of points denoted P_{tol} and is simply removed otherwise. Doing so, the number of clusters can increase or decrease explains a better modeling of the data structure (Serir et al., 2012). This flexibility guarantees the validity of both the covariance matrices and the local models.

3.3. Online Regression phase

In this phase, the local linear models corresponding to each cluster are updated. A local model relates x_k to y_k such that, given only the input x_k, the predicted output $\hat{y}_k$ is as close as possible to the true value y_k. Therefore, in this phase, the goal is to optimize the vector of parameters of the i-th local model.

The estimates of the parameters of the local linear models at a given instant k, $\pi_{i,k} = [a_{i0,k} \quad a_{i1,k} \quad a_{i2,k} \ldots a_{in,k}]^{\mathrm{T}}$ are computed by a recursive least squared approach, given their previous estimates $\pi_{i,k-1}$ and the new input-output vector z_k:

$$\mathcal{C}_{i,k} = \mathcal{C}_{i,k-1} - \frac{\lambda_i(x_{k-1}) \cdot \mathcal{C}_{i,k-1} \cdot x_{e,k-1}^{\mathrm{T}} \cdot x_{e,k-1} \cdot \mathcal{C}_{i,k-1}}{1 + \lambda_i(x_{k-1}) \cdot x_{e,k-1}^{\mathrm{T}} \cdot \mathcal{C}_{i,k-1} \cdot x_{e,k-1}} \tag{19a}$$

$$\hat{\pi}_{i,k} = \hat{\pi}_{i,k-1} + \mathcal{C}_{i,k} \cdot x_{e,k-1} \cdot \lambda_i(x_{k-1}) \cdot \left(y_k - x_{e,k-1}^{\mathrm{T}} \cdot \hat{\pi}_{i,k-1}\right) \tag{19b}$$

with $k > 2$ (time-step), $i = [1, \ c]$ (c is the number of clusters = the number of local models), $x_{e,k} = [1; x_k]$ is the extended input data vector, and $\mathcal{C}_{i,k} \in \mathbb{R}^{(n+1)\times(n+1)}$ is the covariance matrix of the i-th local model at time k, which is updated by the Ricatti equation as in the Kalman filter. The initial conditions ($k = 1$) are set to $\hat{\pi}_1 = 0$ and $c_{i,1} = \alpha \cdot I$, where I is the identity matrix and α a large value (for example 100).

3.4. Prediction phase

Given an input datum x_k, the goal of the prediction phase is to which estimate the value $\hat{y}_k$ taken by the output. The prediction is computed by a simple weighted sum of the local linear models, where the weights are provided directly by the clustering phase:

$$\hat{y}_k = \Psi^{\mathrm{T}} \cdot \theta \tag{20}$$

where $\Psi = [\lambda_{1,k} \cdot x_{ek}; \ldots; \lambda_{i,k} \cdot x_{ek}; \ldots; \lambda_{i,c} \cdot x_{ek}]$ is the vector of the inputs weighted by the normalized activation degrees of the local models, and $\theta = [\pi_{1,k}; \ldots; \pi_{i,k}; \pi_{c,k}]$ is the vector of the parameters of the linear local models.

Remark 2 *When the observed datum at instant k is an input-output couple z_k, then the clustering and the local models can be updated, whereas the prediction of the output $\hat{y}_k$ can be estimated when only the input datum x_k is observed.*

4. APPLICATION ON A REAL-WORLD DATASET

To demonstrate the performance of the proposed method, a real-world multi-dimensional case is considered. E2GKpro is implemented on a dataset provided by the PRONOSTIA platform described in the sequel.

4.1. Description of PRONOSTIA

PRONOSTIA is an experimentation platform (Figure 3) dedicated to the test and validation of the machinery prognosis approaches, focusing on bearing prognostics. It was developed in the Department of Automatic Control and Micro-Mechatronic Systems (AS2M) of FEMTO-ST[1] institute. The main objective of PRONOSTIA is to provide real experimental data (Serir, Ramasso, Nectoux, Bauer, & Zerhouni, 2011) that characterize the degradation of a ball bearing along its whole operational life (until fault/failure). The collected data are vibration and temperature measurements of the rolling bearing during its functioning mode.

The internal bearing ring is put in rotation, while the external bearing ring is maintained fixed. A radial load is applied on the external bearing ring in order to simulate its functioning. To speed up the degradation, the load exceeds the maximal load recommended by the supplier. The originality of this experimental platform lies not only in the conjunction of the characterization of both the bearing functioning (speed, torque and radial force) and its degradation (vibrations and temperature), but also in the possibilities, offered by the platform, to make the operating conditions of the bearing vary during its useful life. Figure 3(c) depicts a bearing before and after the experiment.

The bearing operating conditions are determined by instantaneous measures of the radial force applied on the bearing,

[1] Franche-Comte, Electronics, Mechanics, Thermal Processing, Optics - Science and Technology

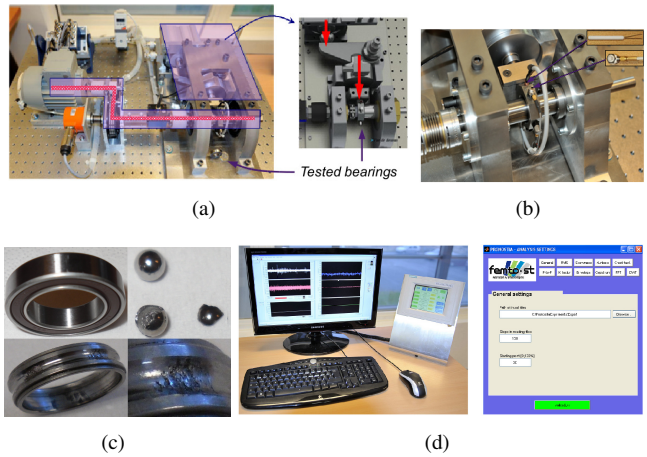

Figure 3. 3(a) The PRONOSTIA platform, 3(b) close-up view on sensors for degradation measurement, 3(c) example of degraded bearings and 3(d) the software for degradation analysis.

the rotation speed of the shaft handling the bearing, and of the torque inflicted on the bearing. During a test, the rolling bearing starts from its nominal mode until the fault state. The bearing behavior is measured using different types of sensors (Figure 3(b)) such as miniaturized acceleration sensors and temperature probe.

The raw signals provided by the sensors are processed in order to extract relevant information concerning bearings states. Several techniques have been implemented and gathered in a signal processing toolbox with Matlab (Fig. 3(d)): time-domain methods (RMS, skewness and kurtosis, crest factor, K-factor, Peak-to-Peak), frequency-domain methods (spectral and cepstrum analysis, envelope detection), time-frequency domain (short-time Fourier transform) and wavelets (discrete transform).

4.2. Prognostic on PRONOSTIA

The experiments are conducted on data used for the IEEE PHM 2012 Prognostic Data Challenge provided by PRONOSTIA and available on its dedicated website[2]. More precisely, the data related to the first functioning condition are used. From the horizontal accelerometer data, two features are generated and used for testing (Fig. 4):

- The RMS which is computed at a given time k in a window of size 50. Let S be the signal of the accelerometer and $\mu(S_W)$ its average value, then:

$$RMS(k) = \sqrt{\frac{1}{W}\sum_{i=k-W+1}^{k}\left(S(i)-\mu(S_W)\right)^2} \quad (21)$$

[2] http://www.femto-st.fr/ieee-PHM2012-data-challenge.

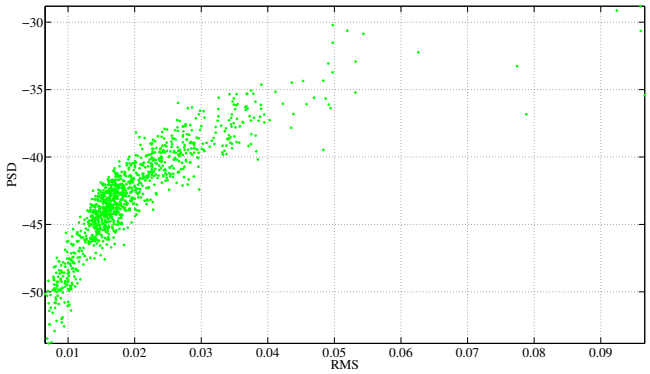

Figure 4. A set of data obtained from the PRONOSTIA platform.

- The mean value of the power spectral density in the same window as the RMS, denoted $PSD(k)$.

The data depict a typical behavior of wear with a small amount of data with high variance close to the end of the experiment (top, right-hand side), while the normal behavior is represented by a larger amount of data points gathered in a smaller space (bottom, left-hand side). The variance is partly due to the high level of noise generated by the vibrations during the experiment.

Let $x(k) = [RMS(k) \quad PSD(k)]^{\mathrm{T}}$ the feature vector at instant k. The goal is to predict $x_{RMS}(k+100)$ and $x_{PSD}(k+100)$ given $x_{in} = [x(k-18) \;\; x(k-12) \;\; x(k-6) \;\; x(k)]^{\mathrm{T}}$. For each data point, x_{in} is a vector with 4×2 elements and the output vector is $x_{out} = [x_{RMS}(k+100) \;\; x_{PSD}(k+100)]^{\mathrm{T}}$. The clustering is performed in the input-output space composed of 10 elements. Given an input data vector (with 8 elements), the prediction is estimated by projecting the centers and the covariance matrix onto the input space, then computing the degree of membership to each cluster, followed by the regression.

The engine is launched and the bearing is gradually degraded. The data are processed along time by E2GKpro, so that clusters representing operating regions are created and the local models estimated. Only one experiment is used to tune the parameters of E2GKpro. The data related to the training dataset is given in Figure 4. The local models are updated as displayed in Figure 5 for the four first rules. The local model being initialized, the predictions 100 steps ahead can be estimated. The result of the prediction on the training data is depicted on Figure 6 (NDEI $= 0.4986$).

The second experiment is then used as a testing dataset. An interesting characteristic of the PRONOSTIA's data holds in the fact that all experiments depict different degradation trends, although bearings with similar mechanical properties were used. An illustration of this variability is represented in Figure 7. In this kind of application, the use of online

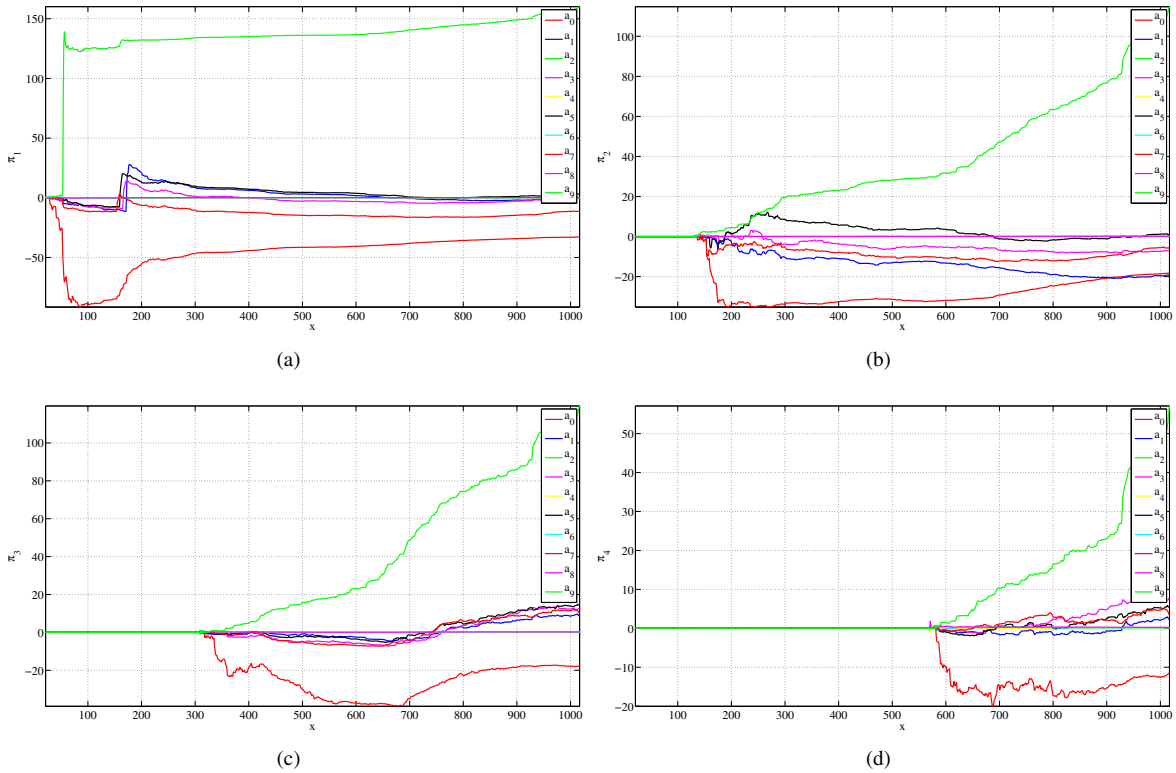

Figure 5. Evolution of the parameters for rules 1 to 4.

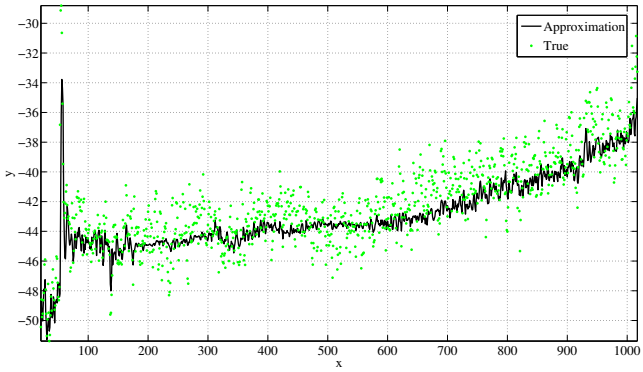

Figure 6. Real data and prediction on the training dataset.

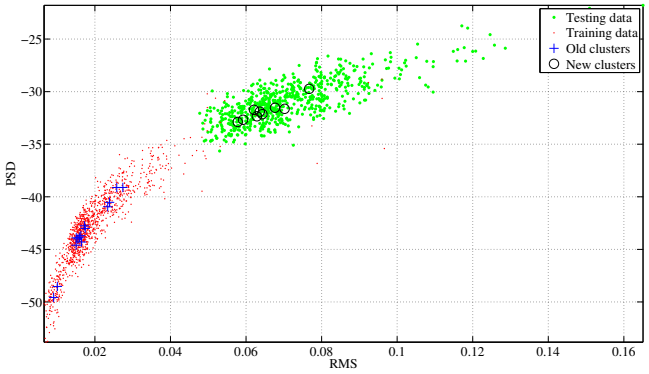

Figure 7. Illustration of high variability: training and testing data.

methods is thus well justified. This figure also shows the position of the clusters found during both training and testing. In the latter case, E2GKpro starts with the local models estimated during the former but E2GKpro still adapts the models accounting for new clusters in the new operating regions encountered in the testing dataset.

The result of the prediction on the second experiment (with features depicted in Figure 4 and on the left-hand side of Figure 7) is shown in Figure 6. In this example, E2GKpro generates smooth predictions, which appear useful in this real-

world application because it gives the global trend of the functioning behavior. Another interesting aspect is the updated set of operating regions found during testing (Figure 7), which can then be used for another experiment.

5. CONCLUSION

An evidential approach to analyze complex systems behavior was proposed in the context of sequential data. The proposed approach is in line with evolving systems modeling approaches and consists in three main phases performed online: 1) the on-line clustering of the data describing the system to determine the different operating regions, 2) the creation, adaptation or removing of models locally estimated for each cluster, and 3) the prediction of the future evolution. E2GKpro was implemented on a real-world dataset provided by PRONOSTIA platform. The results demonstrated the ability of the proposed method for an online segmentation of multi-dimensional time-series and reliable predictions. Further comparative studies have shown the great interest of using belief functions. In particular, the number of rules is decreased compared to usual approaches, while ensuring limited error by using the concept of virtual centroids to represent transitions between operating regions.

REFERENCES

Angelov, P., & Filev, D. (2004). An approach to online identification of Takagi-Sugeno fuzzy models. *IEEE Trans. on Systems, Man, and Cybernetics*, *34*, 484 – 98.

Angelov, P., Filev, D., & Kasabov, N. (2010). *Evolving intelligent systems: Methodology and applications*. New york: IEEE Press Series on Computational Intelligence, John Wiley.

Angelov, P., Lughofer, E., & Zhou, X. (2008). Evolving fuzzy classifiers using different model architectures. *Fuzzy Sets and Systems*, 3160 – 3182.

Boukhris, A., Mourot, G., & Ragot, J. (2000). Nonlinear dynamic system identification: a multiple-model approach. *International Journal of Control*, *72*, 591 – 604.

Chandrashekhar, M., & Ganguli, R. (2009). Uncertainty handling in structural damage detection using fuzzy logic and probabilistic simulation. *Mechanical Systems and Signal Processing*, *23*(2), 384 – 404.

Cobb, B. R., & Shenoy, P. P. (2006). On the plausibility transformation method for translating belief function models to probability models. *International Journal of Approximate Reasoning*, *41*, 314 – 330.

Dubois, D., Prade, H., & Smets, P. (2001). New Semantics for Quantitative Possibility Theory. In *2nd International Symposium on Imprecise Probabilities and Their Applications.* Ithaca, New York.

El-Koujok, M., Gouriveau, R., & Zerhouni, N. (2011). Reducing arbitrary choices in model building for prognostics: An approach by applying parsimony principle on an evolving neuro-fuzzy system. *Microelectronics Reliability*, *51*, 310 – 330.

G. J.Klir. (2006). *Uncertainty and information. Foundations of generalized information theory*. New York: Wiley.

Georgieva, O., & Filev, D. (2009). Gustafson-Kessel algorithm for evolving data stream clustering. In *International Conference on Computer Systems and Technologies - CompSysTech 09.*

Haag, T., Herrmann, J., & Hanss, M. (2010). Identification procedure for epistemic uncertainties using inverse fuzzy arithmetic. *Mechanical Systems and Signal Processing*, *24*(7), 2021 – 2034.

He, H., & Garci, E. A. (2009). Learning from Imbalanced Data. *IEEE Tr. On Knowledge and Data Engineering*, *21*, 1263 – 1284.

Madani, K., Rybnik, M., & Chebira, A. (2003). Non Linear Process Identification Using a Neural Network Based Multiple Models Generator. *LNCS series*, 647 – 654.

Masson, M.-H., & Denoeux, T. (2008). ECM: An evidential version of the fuzzy c-means algorithm. *Pattern Recognition*, *41(4)*, 1384 – 1397.

Murray-Smith, R. (1997). *Multiple model approaches to modeling and control*. Taylor & Francis Publishers.

Nelles, O. (1995). On the identification with neural networs as series-parallel and parallel models. In *International Conference on ANN.* Paris, France.

Ramdani, M., Mourot, G., & Ragot, J. (2005). A Multi-Modeling Strategy based on Belief Function Theory. In *CDC-ECC 05.*

Serir, L., Ramasso, E., Nectoux, P., Bauer, O., & Zerhouni, N. (2011). Evidential Evolving Gustafson-Kessel Algorithm (E2GK) and Its Application to PRONOSTIA's Data Streams Partitioning. In *IEEE Int. Conf. on Decision and Control* (p. 8273-8278).

Serir, L., Ramasso, E., & Zerhouni, N. (2011). E2GKpro: An evidential evolving multimodeling approach for systems behavior prediction. In *PHM Society.* Montreal, Canada.

Serir, L., Ramasso, E., & Zerhouni, N. (2012, July). Evi-

dential evolving Gustafson-Kessel algorithm for online data streams partitioning using belief function theory. *International Journal of Approximate Reasoning*, *53*, 747 – 768.

Shafer, G. (1976). *A mathematical theory of evidence*. Princeton University Press.

Smets, P., & Kennes, R. (1994). The Transferable Belief Model. *Artificial Intelligence*, *66*, 191 – 234.

Takagi, T., & Sugeno, M. (1985). Fuzzy identification of systems and its application to modeling and control. *IEEE Trans. on Systems, Man, and Cybernetics*, *15*, 116 – 132.