Open access • Journal Article • DOI:10.1109/TR.2012.2220700

Connexionist-Systems-Based Long Term Prediction Approaches for Prognostics

Rafael Gouriveau, Noureddine Zerhouni

Institutions: ASM International

Published on: 22 Oct 2012 - IEEE Transactions on Reliability (IEEE)

Topics: Prognostics

Connexionist-Systems-Based Long Term Prediction Approaches for Prognostics.

Rafael Gouriveau, Noureddine Zerhouni

Connexionist-Systems-Based Long Term Prediction Approaches for Prognostics

Rafael Gouriveau, *Member, IEEE,* Noureddine Zerhouni, *Member, IEEE*

***Abstract*—Prognostics and Health Management aims at estimating the remaining useful life of a system (RUL), i.e. the remaining time before a failure occurs. It benefits thereby from an increasing interest: prognostic estimates (and related decision-making processes) enable increasing availability and safety of industrial equipment while reducing costs. However, prognostics is generally based on a prediction step which, in the context of data-driven approaches as considered in this paper, can be hard to achieve because future outcomes are in essence difficult to estimate. Also, a prognostic system must perform sufficient long term estimates, whereas many works focus on short term predictions. Following that, the aim of this paper is to formalize and discuss the connexionist-systems-based approaches to ensure multi-step ahead predictions for prognostics. Five approaches are pointed out: the Iterative, Direct, DirRec, Parallel, and MISMO approaches. Conclusions of the paper are based, on one side, on a literature review; and on the other side, on simulations among 111 time series prediction problems, and among a real engine fault prognostics application. These experiments are performed using the exTS (evolving extended Takagi-Sugeno system). As for comparison purpose, three types of performances measures are used: prediction accuracy, complexity (computational time), and implementation requirements. Results show that all three criteria are never optimized at the same time (same experiment), and best practices for prognostics application are finally pointed out.**

***Index Terms*—Prognostics and health management, multi-step ahead predictions, connexionist system, evolving extended Takagi-Sugeno system.**

ACRONYMS

CBM	Condition Based Maintenance
exTS	evolving extended Takagi-Sugeno system
pdf	probability density function
RLS	Recursive Least Squares
RMSE	Root Mean Square Error
RUL	Remaining Useful Life
TS	Takagi-Sugeno Model

NOTATIONS

$\mathbf{X}$, $\mathbf{Y}$, $\hat{\mathbf{Y}}$	input, and output data sets, estimation of $\mathbf{Y}$
$\epsilon = \mathbf{Y} - \hat{\mathbf{Y}}$	residual of estimates
$\Gamma(.)$	real function which governs the input-output law
$\hat{\Gamma}(.)$	approximation of $\Gamma(.)$
$f(.)$, $[\theta]$	structure and set of parameters of the estimated law
p	number of regressors of the prediction model
t	time index
$\mathbf{S}_t$	univariate time series: $\mathbf{S}_t = \{x_1, x_2, \ldots, x_t\}$
H	final prediction horizon
$\hat{x}_{t+1}$	one-step ahead prediction
$\hat{x}_{t+H}$	final-step ahead prediction
$\hat{\mathbf{X}}_{t+1 \rightarrow t+H}$	set of predictions: $[\hat{x}_{t+1}\ ,\ \hat{x}_{t+2}\ ,\ \ldots\ ,\ \hat{x}_{t+H}]$
μ_e, σ_e	mean, and standard deviation of the errors of prediction

R. Gouriveau and N. Zerhouni are with FEMTO-ST Institute, Automatic Control and Micro-Mechatronic Systems Department (AS2M), UMR CNRS 6174 - UFC / ENSMM / UTBM, 24 rue Alain Savary, Besançon, 25000 France. e-mails: firstname.lastname@femto-st.fr

I. INTRODUCTION

TO avoid high costs while increasing the safety and availability of equipment, researchers and engineers show interest in Condition-Based Maintenance (CBM). More precisely, prognostics becomes a major area of focus. The core purpose of prognostics is to estimate the remaining useful life (RUL) of a system before a failure occurs [1], [2]. It is thereby a promising activity that benefits planning, safety, availability, and maintenance cost reduction [3]. However, real prognostics systems are scarce in industry. Nobody is able to *a priori* ensure that an accurate prognostic model can be built. In other words, the applicability of a prognostics approach is still an open area, mainly because of the prediction required. Also, maintenance managers need the RUL to be greater than the decision, scheduling, and maintenance tasks cumulative times (Fig. 1). Otherwise, prognostics would be useless because maintenance workers would not be able to achieve maintenance before failure occurs. Following that, developing a suitable prognostics system requires performing multi-step ahead predictions to get mid-term or long term estimates of the system's health [4]. This problem of long term prognostics is a central point of this paper.

Three main prognostic approaches are generally distinguished [2], [5]–[8]: model-based, data-driven, and experience-based prognostic approaches. Experience-based prognostic methods are used in statistical reliability applications to predict the probability of a failure at any time. Model-based approaches suppose that the degradation process can be formalized in a mathematical and analytical form.

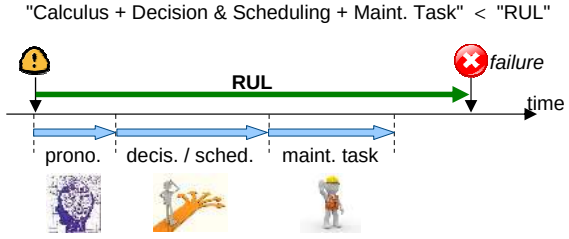


Fig. 1. Useful prognostics must enable maintenance.

Data-driven prognostic methods rely on the assumption that the statistical characteristics of data are relatively unchanged unless a malfunction occurs. These methods aim thereby at transforming raw monitoring data into relevant information and behavior models (including the degradation) of the system. They take as inputs the current monitoring data and return as outputs predictions or trends about the health state of the system. Data-driven approaches offer an alternative, especially in cases where obtaining *in-situ* data is easier than constructing physical or analytical behavior models. Indeed, in many applications, measured input and output data are the major source of information for a deeper understanding of the system degradation. Following that, data-driven approaches are increasingly applied to machine prognostics (mainly techniques from artificial intelligence). More precisely, neural networks and neuro-fuzzy systems (that are widely known as connexionist networks) benefit from a growing interest. Indeed, their approximation capability makes them powerful candidates to achieve the prediction step of prognostics. Actual developments confirm the interest of using this class of approaches in forecasting applications [9]–[17].

Nevertheless, many works focus on short term predictions [13], [15], [18], [19], which does not meet our requirement of a sufficient forecasting horizon. Furthermore, there is no widely accepted way of building long term connexionist-based prediction systems. Various architectures and learning processes can be used whose accuracy performances depend on several factors like the type of connexionist system, the nature of the data to be predicted, and the horizon of prediction. Applicability characteristics like processing time and complexity also vary widely. Following that, the aim of this paper is to review and discuss the connexionist-systems-based approaches to ensure long term predictions for prognostics. Developments emphasize univariate time series forecasting. This paper extends work published in [4], to which we brought improvements on the problem statement, notations, literature review, and tests.

The paper is organized in four main parts. First, a classical data-driven prognostics procedure is replaced within the Condition-Based Maintenance (CBM) concept to point out the problem of long term predictions. Following that, the multi-step ahead prediction problem with connexionist systems is formalized, and the underlying learning phase is shortly explained. At this stage, differences between prediction approaches, tools, learning algorithms, and structures are proposed. In the second part, five types of connexionist-based multi-step prediction approaches are presented and discussed. For that purpose, performances criteria are proposed. The next section aims at testing the five approaches on a benchmark set of time series from NN3 competition (111 data series). This part enables us to complete the discussion, and to identify the most relevant approaches. Developments are finally applied on a real engine fault prognostics problem to validate conclusions on a real world case, and to point out some best practices for prognostics applications. Note that all experiments are performed using the evolving extended Takagi-Sugeno system (exTS).

II. Connexionist system based predictions - problem statement

A. Data-driven prognostics and prediction

Prognostics cannot be seen as a single task; the whole aspects of failure analysis and prediction must be viewed as a set of activities that are necessary to be performed. This aspect is highlighted within the CBM concept. According to CBM practitioners, various activities, ranging from data collection through the recommendation of specific maintenance actions, must be carried out to perform predictive maintenance (and thereby improve maintenance performance). Generally, a CBM system is seen as the integration of seven modules, one of them being that of prognostics [7], and the entire set being sensor, signal processing, condition monitoring, health assessment, prognostic, decision making, and presentation. When focusing on the prognostic process (of data-driven approaches), one can underline a flow that goes from multidimensional data through the remaining useful life of a system. This procedure consists of three main phases (Fig. 2). Data are first acquired from sensor sources, and are then pre-processed before feeding a prognostic model. The pre-processing step is composed of a features extraction module based on signal processing techniques, and of a features selection module that relies on data mining approaches. The prognostic phase is also composed of two complementary modules. A prediction engine forecasts observations in time. These predictions are then analyzed by a classifier which provides the most probable state of the system. The RUL is finally deduced thanks to the estimated time to reach the failure mode. Obviously, prediction phase is critical, and must be dealt with in an appropriate manner to provide accurate predictions, and thereby better RUL estimation. Also, as stated in the introduction of this paper, predictions must be sufficiently long to ensure usefulness of the full prognostic process. In the following sections, we discuss how to perform long term predictions with connexionist systems, which are a kind of approximation tools.

B. Approximation approaches - a formalization

Connexionist systems like neural networks or neuro-fuzzy systems approximate an input-output function. This kind of system must be tuned to fit the studied problem through

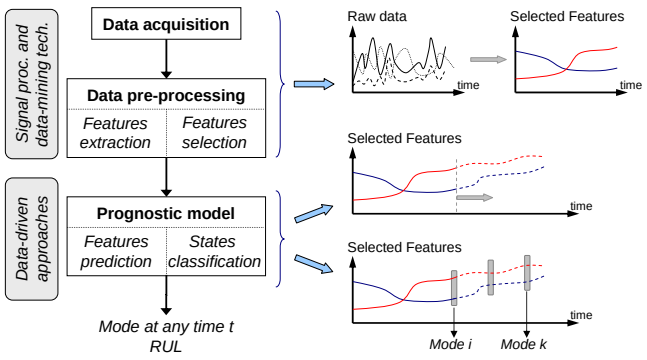


Fig. 2. Data-driven prognostics from data to RUL.

a learning phase. This identification problem is defined as follows.

Let $\mathbf{X}$ be an input data set, $\mathbf{Y}$ be an output data set, and $\Gamma(.)$ be the real function which governs the input-output law of

$$\mathbf{Y} = \Gamma(\mathbf{X}) \tag{1}$$

We use an approximation model to estimate the output set $\mathbf{Y}$ (note the estimate as $\hat{\mathbf{Y}}$). For that purpose, the real function $\Gamma(.)$ is approximated ($\hat{\Gamma}(.)$) so that the global model can be expressed as

$$\hat{\mathbf{Y}} = \hat{\Gamma}(\mathbf{X}) \tag{2}$$

The estimated input-output law $\hat{\Gamma}(.)$ is obtained thanks to a learning phase. For that purpose, $\hat{\Gamma}(.)$ is expressed as the combination of a structure $f(.)$ and a set of parameters $[\theta]$ that both are estimated by using a learning algorithm $La(.)$ that aims at making the residual $\epsilon = \mathbf{Y} - \hat{\mathbf{Y}}$ as close as possible to the null vector.

$$\begin{array}{l} \{f, [\theta]\} \leftarrow La\left(\mathbf{X}, \mathbf{Y}\right) \\ \hat{\Gamma}(.) = f\left([\theta]\right) \end{array} \tag{3}$$

The input-output law being estimated thanks to the learning phase, the approximation function can finally be formalized as

$$\hat{\mathbf{Y}} = f(\mathbf{X}, [\theta]) \tag{4}$$

Building an approximation model requires a structure, and a set of parameters estimated using a learning algorithm. Various structures and algorithms can be used to approximate an input-output function. In the following sections, we thereby distinguish the concepts of approximation approach, and approximation tool. The first one is the way of reaching $\hat{\Gamma}(.)$, while the second one is the basic connexionist system used for that purpose.

C. Multi-step ahead prediction with connexionist systems

Let us now use the concepts introduced in Section II-B to formalize the problem of connexionist-based multi-step ahead prediction of a univariate time series.

A univariate time series $\mathbf{S}_t$ is a chronological sequence of values describing a physical observation made at equidistant intervals [20] $\mathbf{S}_t = \{x_1, x_2, \ldots, x_t\}$. The multi-step ahead prediction problem consists of estimating a set of future values of the time series $\hat{\mathbf{X}}_{t+1 \rightarrow t+H}$. According to (2), this approximation can be expressed as

$$\hat{\mathbf{X}}_{t+1 \rightarrow t+H} = \widehat{msp}(\mathbf{X}_t) \tag{5}$$

where, msp is the multi-step ahead prediction, and $\mathbf{X}_t \in \mathbf{S}_t$ is known as the set of regressors used (for example $\mathbf{X}_t = [x_t \ , \ x_{t-1} \ , \ x_{t-2}]$).

Like in the previous section, a multi-step ahead prediction approach $\widehat{msp}$ can be obtained using different methods and by using different connexionist tools (structure + learning algorithm). Furthermore, various tools can be needed for a single approach. As an example, consider Fig. 3. In this illustration, n tools are needed to perform the global approximation. Each tool has a specific set of inputs $\mathbf{X}^i$, and provides an output set $\hat{\mathbf{Y}}^i$ (where $i = 1 \ldots n$). Depending on the approach, the input set $\mathbf{X}^i$ can be composed of regressor values of the time series, or estimated values of other tools, or both. The global output approximation is a combination of the elements of local function outputs

$$\hat{\mathbf{X}}_{t+1 \rightarrow t+H} \in \hat{\mathbf{Y}}^1 \cup \hat{\mathbf{Y}}^2 \cup \ldots \cup \hat{\mathbf{Y}}^n \tag{6}$$

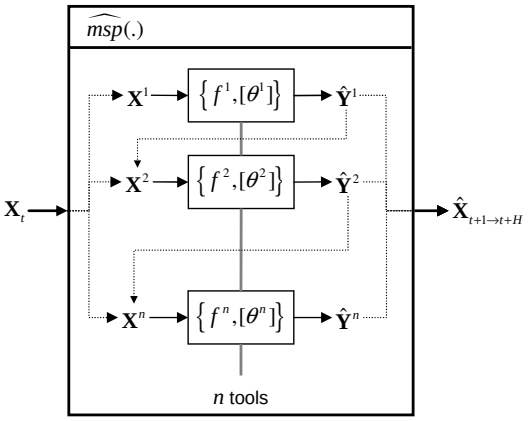


Fig. 3. Representation of a multi-step ahead prediction approach based on various approximation tools.

D. First discussion

This first part of the paper points out that there is no unique way of performing multi-step ahead predictions. Moreover, the performances of a specific approach depend on many aspects, such as the size of the set of regressors $\mathbf{X}_t$, the final horizon H, or the nature of the time series. The choice of the basic prediction tool also is influential. For an example, consider neuro-fuzzy systems. The same type of structure (a first order Takagi Sugeno fuzzy inference model) can be tuned with various algorithms like gradient descent or clustering techniques, which will imply very different approximation (prediction) capabilities. The aim of the following section of the paper is thereby to identify and discuss the main multi-step ahead prediction approaches based on connexionist systems. This approach requires some performances criteria to be defined.

III. Connexionist-based multi-step ahead prediction approaches - an overview

A. A taxonomy of multi-step ahead prediction approaches

Long-term prediction based on connexionist systems benefit from a large interest [21]–[36]. Multi-step prediction approaches can be divided into two main categories [24]: ones that are based on the combination of single output tools (Iterative, Direct, and DirRec approaches), and ones that requires multiple outputs models (Parallel, and MISMO approaches). An illustration of this taxonomy is given in Fig. 4. As for the name of the approaches, there is no absolute consensus in the literature. The Parallel approach defined in this article is called the Direct approach in [28], and the Joint approach in [30].

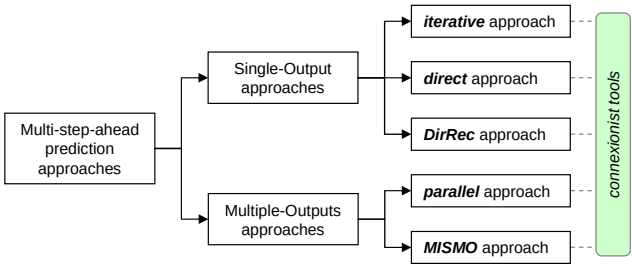


Fig. 4. Taxonomy of multi-step ahead prediction approaches.

B. Iterative approach

The Iterative approach is the most common. Multi-step predictions are provided using a single tool that is tuned to perform a one-step ahead prediction $\hat{x}_{t+1}$. This estimated value is used as one of the regressors of the model to estimate another, and the operation is repeated until the estimation of $\hat{x}_{t+H}$ (see Fig. 5a). Formally,

$$\hat{x}_{t+h} = \begin{cases} \text{if } h = 1, f^1\left(x_t, \ldots, x_{t+1-p}, [\theta^1]\right) \\ \text{elseif } h \in \{2, \ldots, p\}, \\ f^1\left(\hat{x}_{t+h-1}, \ldots, \hat{x}_{t+1}, x_t, \ldots, x_{t+h-p}, [\theta^1]\right) \\ \text{elseif } h \in \{p+1, \ldots, H\}, \\ f^1\left(\hat{x}_{t+h-1}, \ldots, \hat{x}_{t+h-p}, [\theta^1]\right) \end{cases} \tag{7}$$

where $\left\{f^1, [\theta^1]\right\}$ is the one-step ahead prediction model with its parameters set calculated during the learning phase, and p is the number of regressors used, i.e. the number of past discrete values used for prediction. When $h > p$, predictions are made only by using evaluated data, i.e. without observed data. The Iterative approach is the simplest to implement [22], [26]. However, this approach suffers from propagation error: the accuracy decreases as the length of the prediction horizon increases [12], [22], [24], [29]. Moreover, this approach does not take into account the temporal behavior [24].

C. Direct approach

The Direct approach is the combination of H models that aim at predicting $\hat{x}_{t+h}$ (with $h \in [1, H]$). All models use the same observed data (see (8), and Fig. 5b).

$$\left| \begin{array}{l} \hat{x}_{t+1} = f^1(x_t, x_{t-1}, \ldots, x_{t+1-p}, [\theta^1]) \\ \ldots \\ \hat{x}_{t+h} = f^h(x_t, x_{t-1}, \ldots, x_{t+1-p}, [\theta^h]) \\ \ldots \\ \hat{x}_{t+H} = f^H(x_t, x_{t-1}, \ldots, x_{t+1-p}, [\theta^H]) \end{array} \right. \tag{8}$$

where $\left\{f^h, [\theta^h]\right\}$ is the model tuned to provide predictions at $t+h$, and p is again the number of regressors used. The Direct model advantage is that each model is dedicated to the prediction of its own horizon. However, the Direct approach does not take into account the complex dependencies between variables, which influences the prediction accuracy [24]. Moreover, this approach is not easy to implement [26].

D. DirRec approach

The DirRec approach was presented by [34]. As shown in (9) and Fig. 5c, the DirRec approach is similar to the Iterative approach except that each prediction step forecasting model is distinct. The learning procedure is quite different because each model $(\left\{f^1, [\theta^1]\right\}, (\left\{f^2, [\theta^2]\right\}, \ldots)$ must be sequentially tuned. Indeed, predictions at $t+1$ of $\left\{f^1, [\theta^1]\right\}$ are used to tune $\left\{f^2, [\theta^2]\right\}$, and so on until all prediction tools are trained.

$$\hat{x}_{t+h} = \begin{cases} \text{if } h = 1, f^h\left(x_t, \ldots, x_{t+1-p}, [\theta^h]\right) \\ \text{elseif } h \in \{2, \ldots, p\}, \\ f^h\left(\hat{x}_{t+h-1}, \ldots, \hat{x}_{t+1}, x_t, \ldots, x_{t+h-p}, [\theta^h]\right) \\ \text{elseif } h \in \{p+1, \ldots, H\}, \\ f^h\left(\hat{x}_{t+h-1}, \ldots, \hat{x}_{t+h-p}, [\theta^h]\right) \end{cases} \tag{9}$$

where $\left\{f^h, [\theta^h]\right\}$ is the model tuned to provide predictions at $t+h$, and p is again the number of regressors used. According to [12], the DirRec approach has the same disadvantage as the Iterative approach with respect to the propagation of the error, although the new model is created after each step of the prediction process.

E. Parallel approach

The Parallel approach is a multiple output prediction model. This approach calculates all prediction steps with a single model, as illustrated by (10), and Fig. 5d.

$$\begin{array}{rcl} \hat{\mathbf{X}}_{t+1 \to t+H} & = & [\hat{x}_{t+1}, \ldots, \hat{x}_{t+H}] \\ & = & f(x_t, x_{t-1}, \ldots, x_{t+1-p}, [\theta]) \end{array} \tag{10}$$

where $\{f, [\theta]\}$ is the model tuned to provide predictions, and p is again the number of regressors used. This approach provides all step predictions with less computing time than the Direct approach because there is only one model to tune [29]. But it raises serious rounding errors: the number of output nodes is equal to the length of the prediction horizon [32].

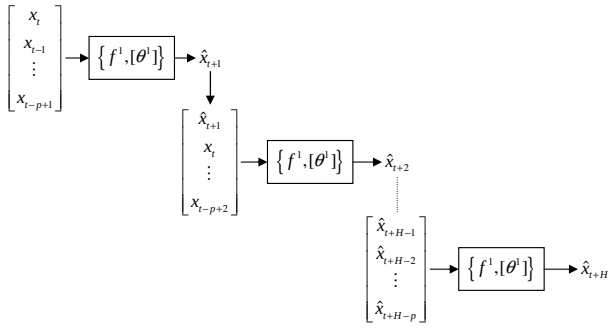
F. MISMO approach

The Multiple-Input Several Multiple-Outputs (MISMO) approach was introduced in [23]. This approach consists of several Multiple-Input Multiple-Output (MIMO) with a parameter s that determines the output number for all MIMO (Fig. 5e). If $s = 1$, this amounts to the Direct approach, whereas if $s = H$ this corresponds to the Parallel approach.

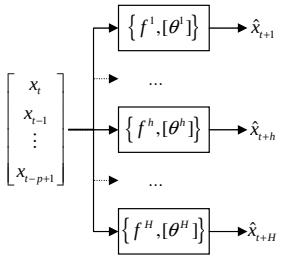
$$\begin{array}{rcl} \widehat{\mathbf{OUT}}^k & = & \left[\hat{x}_{t+ks}, \ldots, \hat{x}_{t+(k-1)s+1}\right] \\ & = & f^k\left(x_t, x_{t-1}, \ldots, x_{t+1-p}, [\theta^k]\right) \end{array} \tag{11}$$

where s is the number of outputs of each model, $m = H/s$ is the total number of models, and $k \in [1, m]$ is the model number. $\left\{f^k, [\theta^k]\right\}$ is the k^{th} model tuned, and p the number of regressors used. According to [24], predictions are expected to be $s-$dependent because of the stochastic properties of the series. At the same time, their degree of dependency is difficult to set *a priori*, and is typically unrelated to the horizon H fixed by the user. The greater adaptability of MISMO comes at the cost of an additional parameter s.

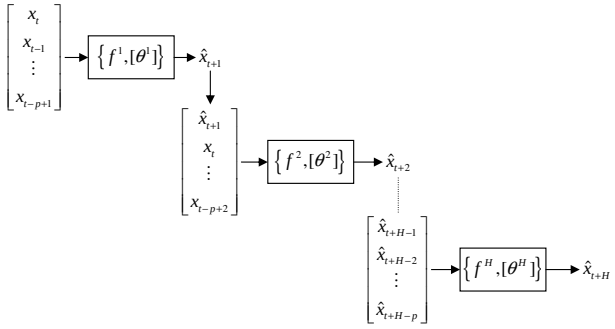
a) Iterative approach representation



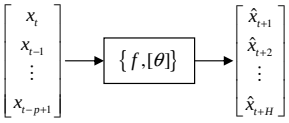
b) Direct approach representation



c) DiRec approach representation



d) Parallel approach representation



e) MISMO approach representation

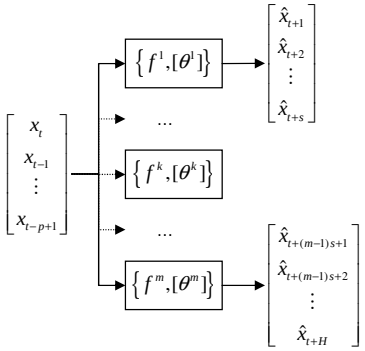


Fig. 5. Representations of multi-step ahead prediction approaches

G. Analysis criteria, and discussion

The scientific literature does not suggest the superiority of any approach. To evaluate and compare the approaches, relevant criteria must be chosen. At least three groups of criteria can be defined:

- prediction approach accuracy,
- computational complexity, and
- implementation difficulty.

Prediction approach accuracy. The choice of an error measure to quantify the accuracy of predictions has been much discussed (see for example [37], [38]). Prediction performances are used to be assessed using the root mean square error criterion (RMSE), which is the most popular prediction error measure, the mean absolute percent error (MAPE), the mean absolute scaled error (MASE), or the coefficient of determination (R2) which is a measure of how well future outcomes are likely to be predicted by the model. In any case, those error measures are only intended as summaries for the error distribution for a specific model. Thereby, the use of the mean (μ_e) and standard deviation (σ_e) of the errors of prediction is also of interest. Moreover, and according to section II-B, the approaches can be compared if the same basic connexionist tool is used.

Computational complexity. A complexity criterion assesses the amount of committed computing resources or time that are necessary during the learning and execution phases, or the number of evaluated parameters. From this point of view, the multiple-tools approaches (Direct, DirRec, and MISMO) should take more training time than do single-tool approaches (Iterative and Parallel) because models have to be tuned. If the learning time is a critical criterion, these three first approaches should be avoided.

Implementation difficulty. The implementation difficulty criterion is more subjective: its aim is to quantify the effort in implementing the prediction approach. Whatever the ability of the practitioner is, it seems adequate to sort the approaches from the most simple to the most complex: Parallel, Iterative, Direct, DirRec, and finally MISMO.

Synthesis. The main characteristics of the approaches are given in Table I. Accordingly, Parallel and Iterative approaches seem interesting from the complexity and implementation points of view. However, prediction accuracy is a critical performance criterion, and definitive conclusions cannot be done until tests are made. In the rest of the paper, the approaches are first compared on prediction benchmarks, and then on a real prognostic problem (Sections IV, and V).

IV. Comparing the approaches with the NN3 competition data sets

A. NN3 data sets

To compare different multi-step prediction approaches, tests are made on data sets from the NN3 competition, which was provided to test the accuracy of computational intelligence methods (notably neural networks) in time series forecasting [39]. The advantage of using NN3 resides in the quantity and diversity of time series: these data sets consist in 111 monthly time series derived from homogeneous population of empirical business time series. Fig. 6 depicts time series numbers 22, 47, 53, 71, 88, and 96 (taken randomly, but ordered). Note that NN3 has been used in [24] for the same type of analysis as in this paper.

TABLE I
CONNEXIONIST-BASED MULTI-STEP AHEAD PREDICTION APPROACHES - LITERATURE REVIEW SYNTHESIS

Approach	Addressed by	Principle	Main advantages and drawbacks	Accu.	Comp.	Impl.
Iterative	[22], [24]–[26], [28], [29], [32]	* Based on a single tool tuned to perform a one-step ahead prediction. * The estimated value is used as a regressor to estimate the following ones. * Operation repeated until final-step estimation.	++ The simplest to implement. −− Suffers from propagation of errors. −− Does not take into account the temporal behaviors.	+−	+	+
Direct	[21], [22], [24], [26]–[29], [31], [33], [35]	* Combination of H models. * Each model h aims at predicting step h. * All models use the same observed data.	++ Each model is dedicated to the prediction of its own horizon. −− Does not take into account complex dependencies between variables. −− Not easy to implement.	+	−	+−
DiRec	[22], [34]	* Similar to the Iterative approach except that all prediction models are distinct. * Each model must be sequentially tuned: predictions of $tool1$ are used to tune $tool2$, and so one until all tools are trained.	++ Quite easy to implement. −− Suffers also from propagation of errors.	+−	−	−
Parallel	[24], [28], [29], [32]	* Multiple outputs prediction model. * All prediction steps calculated with a single model.	++ Provides all steps predictions with less computing time (single model). −− Can raise rounding errors: number of output nodes is equal to H.	+	+	++
MISMO	[23], [24]	* Several Multiple-outputs. * Parameter s determines the output number for all MIMO. * Number of models: $m = H/s$ - If $s = 1$; $m = H$ -> Direct approach - If $s = H$; $m = 1$ -> Parallel approach	++ Greater adaptability thanks to additional parameter s. −− Difficult to set *a priori* - typically unrelated to H. −− Learning phase is time consuming.	+	−	−−

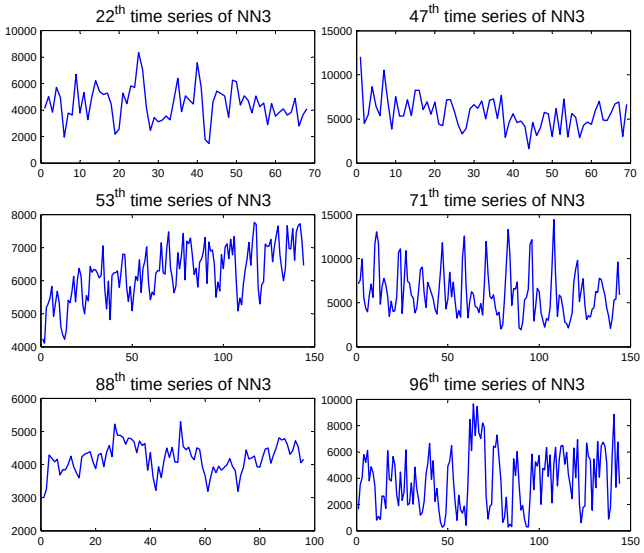


Fig. 6. Six NN3 time series taken randomly.

approximates an input-output function. It can be seen as a multi-linear-model structure: 1) the input space is fuzzily partitioned, 2) a fuzzy rule is assigned to each region of the input space and provides a local linear approximation of the output, and 3) the final output is a combination of the whole rules. Consider Fig. 7 as an example. This model has two inputs variables. Two membership functions (antecedent fuzzy sets) are assigned to each input. The TS model is finally composed of two fuzzy rules.

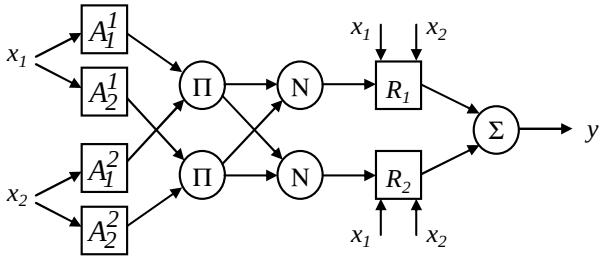


Fig. 7. First order Takagi-Sugeno model.

B. Basic connexionist tool

As stated in Section II-B, a basic connexionist tool has to be chosen to test the different approaches. In this paper, experiments have been made by using the exTS (evolving extended Takagi-Sugeno system) proposed by [40], [41]. The exTS approximation tool consists of a first order Takagi-Sugeno inference structure whose parameters are learned, thanks to the combination of a clustering algorithm with the RLS (Recursive Least Squares) algorithm. This tool has been used in several application areas such as fault detection and diagnosis, and has shown good prediction performances. In previous work, we proposed to use it for prognostics [11].

Takagi-Sugeno models: principles. A first order TS model

The rules perform a linear combination of inputs

$$R_i : \text{if } x_1 \; is \; A_i^1 \text{ and } \ldots \text{and } x_n \; is \; A_i^n \\ \text{then } y_i = a_{i0} + a_{i1}x_1 + \ldots + a_{in}x_n \tag{12}$$

where R_i is the i^{th} fuzzy rule, N is the number of rules, $\mathbf{X}_n = [x_1, ..., x_n]^T$ is the input vector, A_i^j denotes the antecedent fuzzy sets, $j = [1, n]$, y_i is the output of the i^{th} linear subsystem, and a_{iq} are its parameters, $q = [0, n]$.

Due to their generalization capabilities, Gaussian antecedent fuzzy sets are generally assumed to define the regions of fuzzy rules in which the local linear sub-models are valid:

$$\mu_i^j = exp^{-(4\left\|x-x^{i*}\right\|_j)/(\sigma_i^j)^2} \tag{13}$$

with σ_i^j as the spread of the membership function, and x^{i*} as the center of the i^{th} rule antecedent. The firing level τ_i, and the normalized firing level λ_i of each rule are obtained as

$$\tau_i = \mu_i^1(x_1) \times \ldots \times \mu_i^n(x_n) \ , \ \lambda_i = {\tau_i}\Big/{\sum_{v=1}^{N} \tau_v} \qquad (14)$$

The model output is the weighted average of individual rules contributions. With notations, $\pi_i = [a_{i0}, \ldots, a_{in}]$ the parameters vector of the i^{th} sub-model, and $\mathbf{X}_e = [1 \ \mathbf{X}_n^T]^T$ the expanded data vector, this output is expressed as

$$y = \sum\nolimits_{i=1}^{N} \lambda_i y_i = \sum\nolimits_{i=1}^{N} \lambda_i \mathbf{X}_e^T \pi_i \qquad (15)$$

Note that a TS model has two types of parameters. Non-linear parameters are those of the membership functions (centers and spread deviations in (13)). These parameters are referred to as antecedent parameters. Other parameters are the linear ones that form the consequent part of each rule (a_{iq} in (12).

Learning procedure of the eXTS. The eXTS is singular in that the number of rules is equal to the number of membership's functions per input. Parameters are learned (online) thanks to the combination of an unsupervised data clustering algorithm with the Recursive Least Squares (RLS) algorithm. Both phases cannot be fully described here.

- The clustering phase enables adjusting antecedent parameters. It processes on the global input-output data space $\mathbf{z} = [\mathbf{x}^T, \ \mathbf{y}^T]^T$. Each sub-model of the eXTS operates in a sub-area of z. This clustering scheme is based on the calculus of a potential for each new learning sample, which is the capability of data (a sample) to form a cluster (antecedent of a rule). The procedure starts from scratch, and as more data are available, the model evolves by replacement or upgrade of rules [41]. Thanks to this evolving capability, the eXTS system does neither require the user to define the structure of the model nor to initialize the parameters.
- The RLS phase updates the consequent parameters. At any learning step k, (15) can be expressed as

$$\hat{y}_{k+1} = \sum\nolimits_{i=1}^{N} \lambda_i y_i = \sum\nolimits_{i=1}^{N} \lambda_i \mathbf{X}_e^T \pi_i = \psi_k^T \hat{\theta}_k \qquad (16)$$

where $\psi_k^T = [\lambda_1 x_1^T, ..., \lambda_n x_n^T]_k^T$ is the vector of the inputs weighted by normalized firing (λ) of the rules (updated thanks to the clustering phase). $\hat{\theta}_k = [\hat{\pi}_1^T, ..., \hat{\pi}_N^T]_k^T$ is an estimation of the linear parameters of the sub-models obtained by applying the RLS procedure

$$\hat{\theta}_k = \hat{\theta}_{k-1} + C_k \psi_k (y_k - \psi_k^T \hat{\theta}_{k-1}) \ ; \ k = 2, 3, ... \qquad (17)$$

$$C_k = C_{k-1} - \left[C_{k-1} \psi_k \psi_k^T C_{k-1}\right] \Big/ \left[1 + \psi_k^T C_{k-1} \psi_k\right] \qquad (18)$$

with C_k the $R(n+1) \times R(n+1)$ co-variance matrix of parameters errors, and initial conditions $\theta_1 = 0$, $C_1 = \Omega I$, where Ω is a large positive number.

C. Simulation conditions

Horizon of prediction and set of regressors. Tests were carried out on all 111 NN3 time series, without data processing beyond data normalization, and with identical initial conditions for each approach. As for the horizon of prediction (and according to NN3 competition [39]), the last 18 values of each time series were used for test ($H = 18$). To extract more solid conclusions from tests on the comparison of the approaches, the numbers of inputs used has been set from 1 to 5 (regressors p).

Accuracy criteria. The accuracy criteria retained were the root mean square error RMSE, the mean μ_e, and the standard deviation σ_e of the errors of prediction of the 111 time series tests (19). Because the MISMO approach is based on the generation of various models that perform the same predictions (parameter s), the most suitable model for each series (that one with the lowest RMSE) has been retained for comparison before calculating the global accuracy criteria.

$$\begin{aligned} \text{RMSE} &= \sqrt{\frac{1}{H} \sum_{h=1}^{h=H} (x_{t+h} - \hat{x}_{t+h})^2} \\ \mu_e &= \frac{1}{H} \sum_{h=1}^{h=H} (x_{t+h} - \hat{x}_{t+h}) \\ \sigma_e^2 &= \frac{1}{H} \sum_{h=1}^{h=H} (x_{t+h} - \hat{x}_{t+h} - \mu_e)^2 \end{aligned} \qquad (19)$$

Complexity criterion. Complexity has been assessed by estimating the processing time, i.e., the required time to transform time series into interpretable data, to learn the models, and to perform the predictions. For each prediction approach, the complexity criterion is thereby the cumulate processing time to test the 111 time series.

D. Results, and discussion

Discussion on accuracy. Simulation results are given in Table II. From this table, we can see that, whatever the set of regressors is, the MISMO approach appears to be the most accurate (with the lowest RMSE), followed by the Direct and Parallel approaches. The Iterative approach has the worst results. As the RMSE is an aggregated measure, it does not facilitate study of the dispersion of the error (couple (μ_e, σ_e)). Fig. 8 depicts the probability density function (pdf) of the errors of prediction for the three more accurate approaches (MISMO, Direct, and Parallel). One can notice that these pdfs are very similar, the main difference of accuracy being a slight difference between the corresponding spread deviations. However, whatever the number of inputs is, the MISMO approach still presents the lowest error deviation, and appears again to be the most accurate. Table III enables us to have a closer look at the results. This table depicts, for various regressors, the percentage of time series that have been better predicted (lowest (μ_e, σ_e)) when varying the s parameter of the MISMO approach. According to this table, and noting that if $s = H$ then the MISMO is equivalent to the Parallel approach (Section III-F), one can notice that the best predictions are achieved with the Parallel approach: whatever the number of regressors is, the MISMO model with $s = 18$ (the Parallel approach) outperforms all other MISMO for more than 50% of the NN3 time series. This result strengthens the Parallel approach capacity for prediction accuracy.

Discussion on computational complexity. As for the execution time (see Table II), the Iterative and Parallel approaches

TABLE II
RESULTS ON NN3 FROM 1 TO 5 INPUTS

1 input

Approach	RMSE	μ_e	σ_e	Proc. time
Iterative	0.24017	-0.04062	0.23676	8.35
Direct	0.20496	-0.01252	0.20462	130.48
DirRec	0.22686	-0.00979	0.22670	183.91
Parallel	0.19845	-0.01620	0.19784	**7.59**
MISMO	**0.18421**	-0.01559	0.18359	282.00

2 inputs

Approach	RMSE	μ_e	σ_e	Proc. time
Iterative	2.03104	-0.03647	2.03122	9.09
Direct	0.21253	-0.01018	0.21234	141.13
DirRec	0.35858	-0.00581	0.35862	197.19
Parallel	0.22399	-0.01680	0.22342	**8.18**
MISMO	**0.18700**	-0.01615	0.18634	304.36

3 inputs

Approach	RMSE	μ_e	σ_e	Proc. time
Iterative	2.41402	-0.06499	2.41375	9.46
Direct	0.23584	-0.00765	0.23577	147.97
DirRec	1.13689	+0.01070	1.13712	204.36
Parallel	0.26894	-0.01397	0.26865	**8.69**
MISMO	**0.20720**	-0.01584	0.20664	323.44

4 inputs

Approach	RMSE	μ_e	σ_e	Proc. time
Iterative	2.97077	+0.04149	2.97122	10.16
Direct	0.27990	-0.01120	0.27975	156.13
DirRec	1.56018	-0.04948	1.55979	212.53
Parallel	0.32049	-0.01164	0.32036	**9.44**
MISMO	**0.23741**	-0.01401	0.23706	339.53

5 inputs

Approach	RMSE	μ_e	σ_e	Proc. time
Iterative	3.30838	-0.16983	3.30484	10.80
Direct	0.28882	-0.00780	0.28879	164.12
DirRec	0.88700	+0.01866	0.88702	219.38
Parallel	0.32231	-0.01245	0.32215	**10.11**
MISMO	**0.24670**	-0.00675	0.24667	359.27

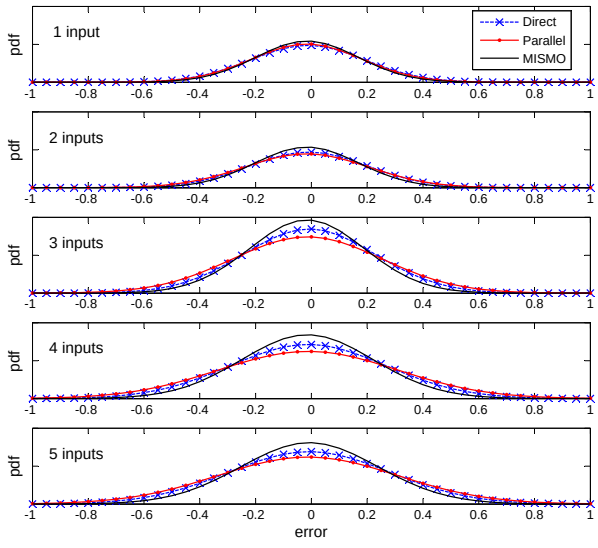


Fig. 8. NN3 error dispersion from 1 to 5 inputs.

are equally fastest, and thereby the most suitable ones with respect to implementation constraints. From this point of view, the MISMO approach is the worst. It balances the results on accuracy, and there is no way to optimize both criteria.

TABLE III
MISMO APPROACH - % OF TIME SERIES BEST PREDICTED ACCORDING TO s AND p PARAMETERS

s **param.**	**1 input**	**2 inputs**	**3 inputs**	**4 inputs**	**5 inputs**
1	14	13	14	18	9
2	4	3	2	3	7
3	3	3	0	8	8
6	6	7	7	9	10
9	12	18	21	14	16
18	**61**	**57**	**57**	**49**	**50**
	100%	100%	100%	100%	100%

Fig. 9 offers a clear representation of this problem. According to this figure (and according to previous conclusions), the Parallel approach appears to be the one which offers the greater compromise between accuracy and complexity, regardless of the number of inputs.

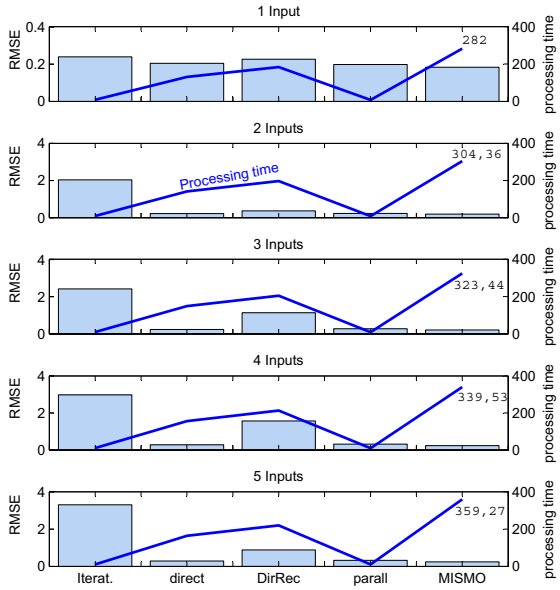


Fig. 9. RMSE versus computing time for various number of inputs.

Discussion on implementation difficulty. As expected, and even if it cannot be shown with tests, when performing experiments, the MISMO approach appeared to be by far the hardest multi-step ahead prediction approach to implement, because many cases must be taken into account (various models, research of the best one). As well, DirRec and Direct approaches require the practitioner to be quite comfortable with computing techniques. By contrast, the Parallel approach is the most natural to program as a particular case of classical approximation tasks. The Iterative approach also is quite easy to deploy because there is a single approximation tool to be computed.

Table IV depicts a synthesis on NN3 time series tests.

V. APPLICATION ON A REAL DEGRADATION DATASET

A. Dataset on a real engine health

Developments were applied to the challenge dataset of diagnostics and prognostics of machine faults from the first

TABLE IV
SYNTHESIS OF TESTS ON NN3 TIMES SERIES

Approach	Accur.	Compl.	Impl.	Balance
Iterative	−−	++	+	+ + + − −
Direct	+	+	+/−	+ + +/−
DirRec	−	−	−	− − −
Parallel	+	++	++	+ + + + +
MISMO	++	−−	−−	+ + − − −−

International Conference on Prognostics and Health Management (2008) [42]. This dataset consists of multiple multivariate time series signals (26 features) with sensor noise (see Fig. 10 for some examples). Each set of time series comes from a different engine of the same fleet. Each engine starts from a specific degree of initial wear. Manufacturing variations are unknown by the user. The engine operates normally at the beginning, but develops a fault. The fault grows until system failure.

As illustrated in Fig. 2, two steps are necessary to perform prognostics. The first one aims at forecasting features in time, i.e. at predicting observations. The second one, the classification step, aims at providing the health state of the engine at any time. RUL estimation is derived, and is the difference in between current time and the time at which failure mode will be reached. In this section, results on prediction by applying connexionist-based multi-step ahead prediction approaches are given (ï¿½ V-C). Also, the analysis is extended to the classification phase to assess the impact of prediction approach capabilities on RUL estimation (ï¿½ V-D).

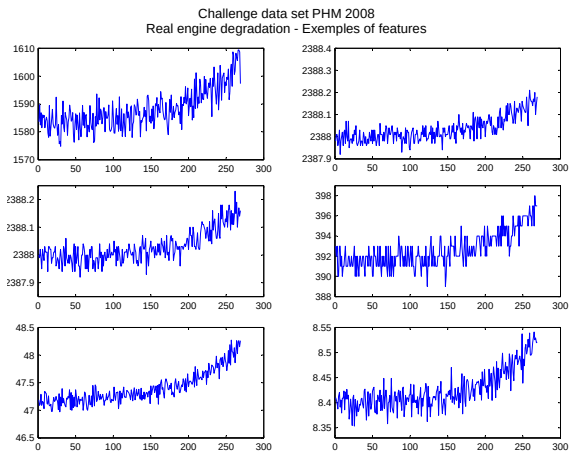


Fig. 10. Real degradation dataset - examples of features from the same engine.

B. Simulation conditions

From the dataset, among 26 available features, 8 were pre-selected in a previous work thanks to information theory, and Choquet Integral [43]. As for the number of prediction inputs, two regressors have been used in addition to a third input, which is the time index as suggested in [43]. The learning data set was composed of 40 runs from previous degraded systems. During the test, 15 distinct degradation data sets have been predicted. For that purpose, the first 50 values of each feature were learned as if they have been provided by a monitoring system, and predictions were made with a horizon of 80 steps-ahead ($H = 80$: from time 51 to 130).

Considering the results of Section IV synthesized in Table IV, the MISMO approach has been removed from tests. Indeed, its processing time is prohibitive with regard to a real world case study. Also, the DirRec approach, that neither appears to be accurate nor has a slight computation time, has also been excluded from tests. Finally, experiments have been made with the Iterative, the Direct, and the Parallel approaches.

Similarly to NN3 competition experiments, the approaches have been assessed by using the accuracy criteria RMSE, μ_e and σ_e, and by calculating the overall processing time.

C. Prediction phase - Results and discussion

As for readability purpose, only results on the fourth feature are presented and discussed. However, conclusions are the same with remaining features. Fig. 11 depicts an example of prediction results. Whereas the first 50 values were used for training for each multi-step prediction approach, the remaining 80 were predicted at time 50. Even if the prediction curves are quite close together, it appears that Parallel approach provides bests results. That can be more closely discussed by considering the whole tests on 15 degradation time series.

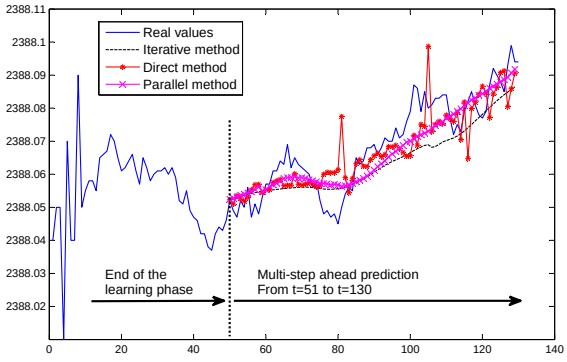


Fig. 11. Real degradation feature - example of predictions.

Table V resumes the performances criteria obtained by considering 15 degradations for test. Fig. 12 depicts the dispersion of the errors of prediction. One can note that the Parallel and Direct approach have quite the same accuracy, and are roughly equivalent whatsoever in dispersion or average. However, the Direct approach requires much more processing time than the two other approaches (it takes 120 times longer than the Parallel approach). Following that, the most suitable approach for real world problems seems to be the Parallel one because it is the one that offers the best compromise between accuracy and complexity. Those experiments confirm the NN3 tests.

D. Classification phase - Results and discussion

As stated in the introduction, performing accurate long term predictions is not an objective in itself, but is required to

TABLE V
RESULTS ON REAL DEGRADATION DATA

Approach	RMSE	μ_e	σ_e	Proc. time
Iterative	0.04601	-0.00985	0.04496	384.74
Direct	0.02658	+0.00404	0.02628	15923.32
Parallel	**0.02504**	+0.00409	0.02471	**133.26**

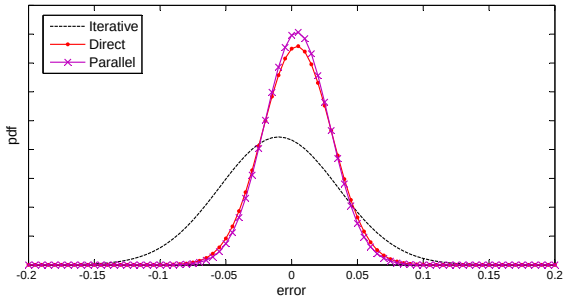


Fig. 12. Real degradation prediction - Error distribution.

enable practitioners to have enough time to plan and achieve maintenance tasks before failure occurs (otherwise prognostic would be useless; see Fig. 1). However, the time to failure is in essence unknown, and uncertain. Following that, an additional problem can be pointed out. How do we set in advance the final horizon of prediction H? Indeed, if $H < RUL$, the prognostic model does not allow deciders to anticipate failure with sufficient time. This aspect can be critical with respect to the study of multi-step ahead prediction approaches. Indeed, the Iterative approach is the only one that does not require the user to *a priori* set the final horizon step. In others words, the Iterative approach is the only approach that enables one to estimate the RUL whatever the actual value is. This advantage can be illustrated by considering the classification phase of prognostics.

For illustration purpose, classification has been performed by using a Fuzzy-Cmeans clustering algorithm [44]. Details of this step cannot be fully presented here. In a few words, a classifier has been built during the learning phase to identify the health state of the system. For that purpose, four functioning modes have been considered: steady state (the engine is working well), degrading state (the engine is being degraded), transition state (in between steady and degrading), and critical state (faulty state, when the engine has failed). When performing tests, predictions of features were used as input from this classifier to estimate the future probable functioning modes of the engine. An example is proposed in Fig. 13 that depicts the future states of a test engine according to predicted features. For clarity, a single predicted feature is presented.

One can note that, in this case, neither the Direct approach nor the Parallel can provide long term estimates of the RUL: the learning scheme has been set for $H = 80$, and greater predictions cannot be obtained. In opposite to that, the Iterative approach can perform predictions at any time, and thereby enables us to reach all health states. RUL can be estimated whatever the horizon of prediction.

According to all these results, and remembering that the Parallel approach is the best compromise between accuracy and complexity, one should compute two types of multi-step ahead prediction approaches:

- the Iterative approach in order to provide results for any future time, and
- the Parallel approach, to get increased confident RUL estimation results for short term predictions.

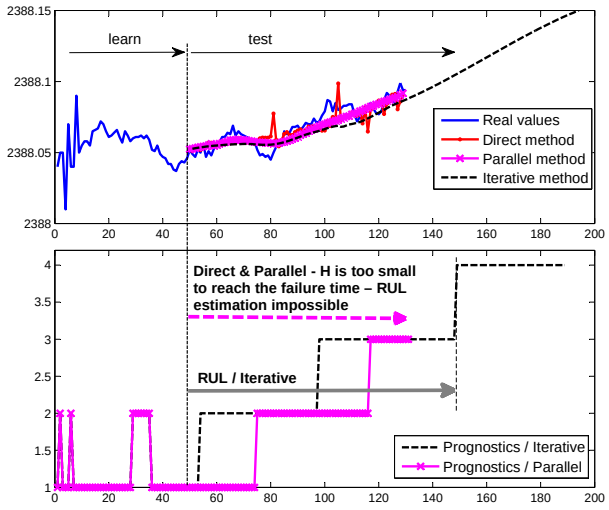


Fig. 13. Health state classification and RUL estimation.

VI. CONCLUSION, AND WORK IN PROGRESS

The aim of this paper is to point out an efficient connexionist-systems-based approach to ensure long term predictions for prognostics. The paper emphasizes univariate time series forecasting. Five connexionist-system-based approaches have been studied, namely the Iterative, Direct, DirRec, Parallel, and MISMO approaches. After defining and formalizing each approach, experiments have been made by using two types of data. First, the NN3 competition dataset has been used to discuss the accuracy and complexity performances of each approach. On this basis, three types of approaches have been applied on a real degradation problem. On both tests, the exTS neuro-fuzy system has been used as the basic tool of each one of the approaches.

Whatever the experiments, the conclusions are similar. The approach that was best is the Parallel approach because it provides a compromise between accuracy and complexity. However, an important point must be raised. The Iterative approach is the only one able to predict at any horizon of prediction. Indeed, in the other approaches, the practitioner must set in advance the final horizon of the prediction he would like. Following that, a good practice could be to implement both approaches: use the Parallel approach to get accurate predictions for a limited critical horizon, and use the Iterative approach to provide more long term tendencies.

The work must be reinforced by analyzing the influence of the learning size, and of the basic connexionist tool notably. An extension to multidimensional predictions is also planned.

REFERENCES

[1] ISO13381-1, *Condition monitoring and diagnostics of machines - prognostics - Part1: General guidelines*. Internat. Standard, ISO, 2004.

[2] A.-K.-S. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mech. Systems and Sig. Processing*, vol. 20, pp. 1483–1510, 2006.

[3] T. Brotherton, G. Jahns, J. Jacobs, and D. Wroblewski, "Prognosis of faults in gas turbine engines," in *IEEE Int. Conf. on Aerospace*, vol. 6, 2000, pp. 163–171.

[4] M.-D. Gauvain, R. Gouriveau, N. Zerhouni, and M. Hessabi, "Long term prediction approaches based on connexionist systems - a study for prognostics application," in *2011 IEEE Conf. on Prognostics and Health Management (PHM)*, june 2011, pp. 1–8.

[5] C. Byington, M. Roemer, G. Kacprzynski, and T. Galie, "Prognostic enhancements to diagnostic systems for improved condition-based maintenance," in *IEEE Aerospace Conf., Big Sky, USA*, 2002.

[6] A. Heng, S. Zhang, A.-C.-C. Tan, and J.Matwew, "Rotating machinery prognostic:state of the art,challenges and opportunities," *Mech. Systems. and Sig. Processing.*, vol. 23, no. 3, pp. 724–739, 2009.

[7] M. Lebold and M. Thurston, "Open standards for condition-based maintenance and prognostics systems," in *5th Annual Maintenance and Reliability Conference*, 2001.

[8] G. Vachtsevanos, F.-L. Lewis, M. Roemer, and A. Hess, *Intelligent Fault Diagnosis and Prognosis for Engineering Systems*. New Jersey, Hoboken: Wiley & Sons, 2006.

[9] Y.-L. Dong, Y.-J. Gu, K. Yang, and W.-K. Zhang, "A combining condition prediction model and its application in power plant," in *Int. Conf. on Mach.ine Learning and Cyber.*, vol. 6, 2004, pp. 474–3478.

[10] M. El-Koujok, R. Gouriveau, and N. Zerhouni, "Towards a neuro-fuzzy system for time series forecasting in maintenance applications," in *IFAC World Congress, Seaoul, Korea*, 2008.

[11] M. El-Koujok, R. Gouriveau, and N. Zerhouni, "Reducing arbitrary choices in model building for prognostics: An approach by applying parsimony principle on an evolving neuro-fuzzy system," *Microelectronics Reliability*, vol. 51, pp. 310–320, 2011.

[12] V.-T. Tran, B.-S. Yang, and A.-C.-C. Tan, "Multi-step ahead direct prediction for the machine condition prognosis using regression trees and neuro-fuzzy systems," *Expert Systems with Applications*, vol. 36, pp. 378–387, 2009.

[13] W.-Q. Wang, "An adaptive predictor for dynamic system forecasting," *Mechanical Systems and Signal Processing*, vol. 21, pp. 809–823, 2007.

[14] W.-Q. Wang and G. Vachtsevanos, "Fault prognostic using dynamic wavelet neural networks," *Artif. Intell. for Eng Design Analysis and Manufact*, vol. 15, pp. 349–365, 2001.

[15] W.-Q. Wang, M.-F. Goldnaraghi, and F. Ismail, "Prognosis of machine health condition using neuro-fuzzy systems," *Mechanical Systems and Signal Processing*, vol. 18, pp. 813–831, 2004.

[16] W.-Q. Wang, F. Ismail, and M.-F. Goldnaraghi, "A neuro-fuzzy approach to gear system monitoring," *IEEE Transaction Fuzzy Systems*, vol. 12, pp. 710–723, 2004.

[17] R.-C.-M. Yam, P.-W. Tse, L. Li, and P. Tu, "Intelligent predictive decision support system for condition-based maintenance," *Int. Jour. of Advanced Manufacturing Technology*, vol. 17, pp. 383–391, 2001.

[18] V.-T. Tran, B.-S. Yang, M.-S. Oh, and A.-C.-C. Tan, "Machine condition prognosis based on regression trees and one-step-ahead prediction," *Mech. Syst. and Signal Proc.*, vol. 22, no. 5, pp. 1179–1193, 2008.

[19] W.-Q. Wang, M.-F. Goldnaraghi, and F. Ismail, "A robust prognostic system for real time industrial applications," in *4th International Conference on Industrial Automation*, 2003.

[20] A.-K. Palit and D. Popovic, *Computational Intelligence in Time Series Forecasting: Theory and Engineering Applications*, 1st ed. Springer, Oct. 2005, vol. Chapitre 2.

[21] A. Andalib and F. Atry, "Multi-step ahead forecasts for electricity prices using NARX: a new approach, a critical analysis of one-step ahead forecasts," *Energy Conversion and Management*, vol. 50, no. 3, pp. 739–747, Mar. 2009.

[22] M. Assaad, R. Boné, and H. Cardot, "A new boosting algorithm for improved time-series forecasting with recurrent neural networks," *Information Fusion*, vol. 9, no. 1, pp. 41–55, 2008.

[23] S. Ben Taieb, G. Bontempi, A. Sorjamaa, and A. Lendasse, "Long-term prediction of time series by combining direct and mimo strategies," in *Neural Networks, 2009. IJCNN 2009. International Joint Conference on*. IEEE, 2009, pp. 3054–3061.

[24] S. Ben Taieb, A. Sorjamaa, and G. Bontempi, "Multiple-output modeling for multi-step-ahead time series forecasting," *Neurocomputing*, vol. 73, no. 10-12, pp. 1950–1957, Jun. 2010.

[25] S.-A. Billings and X. Hong, "Dual-orthogonal radial basis function networks for nonlinear time series prediction," *Neural Networks*, vol. 11, no. 3, pp. 479–493, Apr. 1998.

[26] C.-T. Cheng, J.-X. Xie, K.-W. Chau, and M. Layeghifard, "A new indirect multi-step-ahead prediction model for a long-term hydrologic prediction," *Journal of Hydrology*, vol. 361, no. 1-2, pp. 118–130, Oct. 2008.

[27] F.-A. Guerra and L. Dos S. Coelho, "Multi-step ahead nonlinear identification of lorenz's chaotic system using radial basis neural network with learning by clustering and particle swarm optimization," *Chaos, Solitons & Fractals*, vol. 35, no. 5, pp. 967–979, Mar. 2008.

[28] C. Hamzaçebi, D. Akay, and F. Kutay, "Comparison of direct and iterative artificial neural network forecast approaches in multi-periodic time series forecasting," *Expert Systems with Applications*, vol. 36, no. 2, pp. 3839–3844, 2009.

[29] N. Huck, "Pairs trading and outranking: The multi-step-ahead forecasting case," *European Journal of Operational Research*, vol. 207, no. 3, pp. 1702–1716, Dec. 2010.

[30] D.-M. Kline, "Methods for multi-step time series forecasting with neural networks," in *Neural networks in business forecasting*, G. P. Zhang, Ed. IGI Global, 2004, pp. 226–250.

[31] J. Liu, W. Wang, and F. Golnaraghi, "A multi-step predictor with a variable input pattern for system state forecasting," *Mechanical Systems and Signal Processing*, vol. 23, no. 5, pp. 1586–1599, Jul. 2009.

[32] H.-T. Pao, "Forecasting energy consumption in taiwan using hybrid nonlinear models," *Energy*, vol. 34, no. 10, pp. 1438–1446, Oct. 2009.

[33] A.-G. Parlos, O.-T. Rais, and A.-F. Atiya, "Multi-step-ahead prediction using dynamic recurrent neural networks," *Neural Networks*, vol. 13, no. 7, pp. 765–786, Sep. 2000.

[34] A. Sorjamaa and A. Lendasse, "Time series prediction using dirrec strategy," in *ESANN, European Symposium on Artificial Neural Networks*, 2006, pp. 143–148.

[35] Y. Sun, V. Babovic, and E.-S. Chan, "Multi-step-ahead model error prediction using time-delay neural networks combined with chaos theory," *Journal of Hydrology*, vol. 395, no. 1-2, pp. 109–116, 2010.

[36] J. Zhang, A.-J. Morris, and E.-B. Martin, "Long-term prediction models based on mixed order locally recurrent neural networks," *Computers & Chemical Engineering*, vol. 22, no. 7-8, pp. 1051–1063, Jul. 1998.

[37] J.G. De Gooijer and R.J. Hyndman, "25 years of time series forecasting," *International Journal of Forecasting*, vol. 22, pp. 443–473, 2006.

[38] R.J. Hyndman and A.B. Koehler, "Another look at measures of forecast accuracy," *Inter. Journal of Forecasting*, vol. 22, pp. 679–688, 2006.

[39] "NN3 forecasting competition." [Online]. Available: http://www.neural-forecasting-competition.com/NN3/index.htm

[40] P. Angelov and D. Filev, "An approach to online identification of takagi-sugeno fuzzy models," *IEEE Trans Syst Man Cybern B Cybern*, vol. 34, no. 1, pp. 484–98, 2004.

[41] P. Angelov and X. Zhou, "Evolving fuzzy systems from data streams in real-time," in *2006 IEEE Symposium on Evolving Fuzzy Systems*, 2006, pp. 26–32.

[42] A. Saxena, K. Goebel, D. Simon, and N. Eklund, "Damage propagation modeling for aircraft engine run-to-failure simulation," in *Proceedings of 2008 International Conference on Prognostics and Health Management*. Citeseer, 2008.

[43] E. Ramasso and R. Gouriveau, "Prognostics in switching systems: Evidential markovian classification of real-time neuro-fuzzy predictions," in *Prognostics and Health Management Conference, 2010. PHM'10.* IEEE, 2010.

[44] J.-C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithm*. Plenum New York, 1981.

Dr. Rafael Gouriveau received his engineering degree from National Engineering School of Tarbes (ENIT) in 1999. He then got his MS (2000), and his Ph.D. in Industrial Systems in 2003, both from the Toulouse National Polytechnic Institute (INPT). During his PhD, he worked in the field of risk management and dependability analysis. In September 2005, he joined the National Engineering Institute in Mechanics and Microtechnologies of Besanï¿½on (ENSMM) as Associate Professor. His main teaching activities are concerned with production, maintenance, manufacturing, and informatics domains. As for investigation, Rafael Gouriveau is with the AS2M (Automatic Control and Micro-Mechatronic Systems) department of FEMTO-ST Institute. He is currently at the head of the PHM team in this department. His research interests are in the development of industrial prognostics systems using connexionist approaches like neuro-fuzzy methods, and the investigation of reliability modeling using possibility theory. He is also the scientific

coordinator of PHM research axe from the FCLAB (Fuel Cell Lab) Research Federation (CNRS), devoted to Fuel Cell Systems.

Prof. Noureddine Zerhouni received his engineering degree from National Engineers and Technicians School of Algiers (ENITA) in 1985. After a short period in industry as an engineer, he received his Ph.D. Degree in Automatic Control from the Grenoble National Polytechnic Institute in 1991. In September 1991, he joined the National Engineering School of Belfort (ENIB) as Associate Professor. At this time, his main research activity is concerned with modelling, analysis, and control of manufacturing systems. Since September 1999, Noureddine Zerhouni has been a Professor at the national high school of mechanics and microtechniques of Besanï¿½on. As for investigation, he is with at the head of the AS2M (Automatic Control and Micro-Mechatronic Systems) department of the Femto-st Institute. His main research activities are concerned with intelligent maintenance systems, and e-maintenance. Professor Noureddine Zerhouni has been and is involved in various European and National projects on intelligent maintenance systems like the FP5 European Integrated Project of the ITEA program (Information Technology for European Advancement) PROTEUS, NEMOSYS (Naval E-Maintenance Oriented SYStem) with DCNS, and AMIMAC-FAME (Reliability Improvement of Embedded Machines) with ALSTOM.