Long term prediction approaches based on connexionist systems - A study for prognostics application.

Marie-Danièle Gauvain, Rafael Gouriveau, Mike Hessabi, Noureddine Zerhouni

Long term prediction approaches based on connexionist systems - A study for prognostics application

Marie-Danièle Gauvain, Rafael Gouriveau and Noureddine Zerhouni
FEMTO-ST Institute - 25000 Besançon, France
Email: rgourive@ens2m.fr

Mike Hessabi
Emasystec - 25000 Besançon, France
http://www.emasystec.com

***Abstract*—Data-driven approaches are increasingly applied to machine prognostics. More precisely, connexionist systems like neural networks and neuro-fuzzy systems benefit from a growing interest. Indeed, their approximation capability makes them as powerful candidates to achieve the prediction step of prognostics. Nevertheless, prognostic implies to be able to perform multi-step ahead predictions whereas many works focus on short term predictions. Following that, the aim of this paper is to review and discuss the connexionist-systems-based approaches to ensure long term predictions for prognostics. The paper emphasizes on univariate time series forecasting. Five connexionist-systems-based approaches are pointed and formalized, namely: the iterative, direct, DirRec, parallel and MISMO approaches. Their performances are analyzed according to three types of criteria: those one of prediction accuracy, of complexity (computational time) and of implementation requirements. In addition, simulations are made among 111 times series prediction problems in order to reinforce the discussion. These experiments are performed by using the exTS (evolving extended Takagi-Sugeno system). Finally developments are applied on a real engine fault prognostics problem in order to validate conclusions on a real world case and to point out some best practices for prognostics applications.**

I. Introduction

The main objective of prognostic is to estimate the remaining useful life (RUL) of a system. Moreover, maintenance managers need the RUL to be greater than the "decision, scheduling and maintenance tasks" cumulated times, otherwise prognostic would be useless since maintenance workers would not be able to achieve maintenance before failure occurs. Following that, developing a suitable prognostic system implies to be able to perform multi-step ahead predictions in order to provide mid-term or long term estimation of the system's health.

According to literature, three main prognostic approaches are generally retained [1]–[3]: model-based, data-driven and experience-based prognostic approaches. In a few words, 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. Data-driven prognostic methods rely on the assumption that the statistical characteristics of data are relatively unchanged unless a malfunctioning 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 to the others approaches, especially in cases where obtaining reliable data is easier than constructing physical or analytical behavior models. Indeed, in many applications, measured input/output data is the major source of information for a deeper understanding of the system degradation. Following that, data-driven approaches are increasingly applied to machine prognostic (mainly techniques from artificial intelligence). More precisely, connexionist systems like neural networks and neuro-fuzzy systems benefit from a growing interest. Indeed, their approximation capability makes them as powerful candidates to achieve the prediction step of prognostics. Actual developments confirm the interest of using this class of approaches in forecasting applications [4]–[7].

Nevertheless, many works focus on short term predictions which is not really matching the requirement of a sufficient forecasting horizon stated before. Furthermore, there is no consensual 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, the horizon of prediction. Applicability characteristics like processing time and complexity also vary from an extreme to an other. 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 on univariate time series forecasting.

The paper is organized in four mains parts. First, 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 algorithm and structure" 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 to complete the discussion in order to point out the most relevant approaches. Developments are finally applied on a real engine fault prognostics problem in order to validate conclusions on a real world case and to point out some best practices for prognostics applications. Note that all experiments are performed by using the exTS (evolving extended Takagi-Sugeno system) as the basic tool of each one of the approaches.

II. Connexionist system based predictions - problem statement

A. Approximation approaches - a formalization

Connexionist systems like neural networks or neuro-fuzzy systems aim at approximating an input-output function. This kind of systems must be tuned to fit the studied problem thanks to a learning phase of parameters (and/or structure). This "identification problem" is defined as follows.

Let $[X]$ be an input data set, $[Y]$ an output data set, and $r(.)$ the real function which governs the input/output law:

$$[Y] = r([X]) \tag{1}$$

The use of an approximation model aims at estimating the output set $[Y]$ (let note it $[\hat{Y}]$). For that purpose, the real function $r(.)$ is approximated ($\hat{r}(.)$) so that the global model can be expressed as follows.

$$[\hat{Y}] = \hat{r}([X]) \tag{2}$$

The estimated input/output law $\hat{r}(.)$ is obtained thanks to a learning phase. For that purpose, $\hat{r}(.)$ 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 $\epsilon = [Y] - [\hat{Y}]$ as close as possible to the null vector.

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

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

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

As a short synthesis, building an approximation model requires a structure and a set of parameters estimated thanks to a learning algorithm. Various structures and algorithms can be used to approximate a same input-output law. In the following sections, we thereby distinguish the concepts of "approximation approaches" and "approximation tools". The first one is the way of reaching $\hat{r}(.)$, the second one is the basic connexionist system used for that purpose.

B. Multi-step ahead prediction with connexionist systems

Let now use the concepts introduced in section II-A to formalize the problem of connexionist-based multi-step ahead prediction of an univariate time series.

A univariate time series S_t is a chronological sequence of values describing a physical observation made at equidistant intervals [8]: $S_t = \{x_1, x_2, \ldots, x_t\}$ (where t states for the temporal index variable).

The multi-step ahead prediction problem consists in estimating a set of future values of the time series: $[\hat{X}_{t+1 \rightarrow t+H}] = [\hat{x}_{t+1}\ ,\ \hat{x}_{t+2}\ ,\ \hat{x}_{t+3}\ ,\ \ldots\ ,\ \hat{x}_{t+H}]$ where H states for the final prediction horizon. According to eq. 2, this approximation can be expressed as:

$$[\hat{X}_{t+1 \rightarrow t+H}] = \widehat{msp}([X_t]) \tag{5}$$

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

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

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

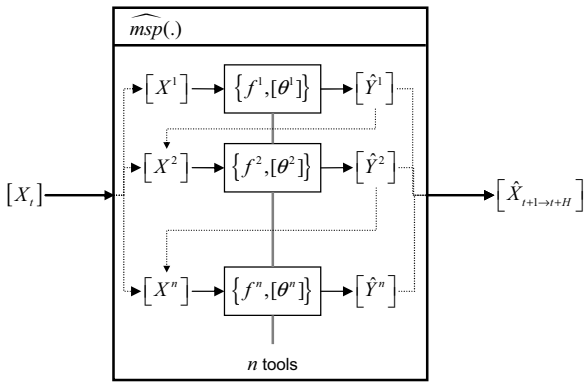

Fig. 1. Representation of a multi-step ahead prediction approach based on various approximation tools

C. First discussion

This first part of the paper enables to point 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 $[X_t]$, the final horizon H, the nature of the time series. The choice

of the basic prediction tool also is influent. As 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 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

Multi-step prediction approaches can be divided into two main categories [9]: that ones that are based on the combination of single output tools ("iterative", "direct" and "DirRec" approaches), and that ones that requires multiple outputs models ("parallel" and "MISMO" approaches). An illustration of this taxonomy is given in Figure 2.

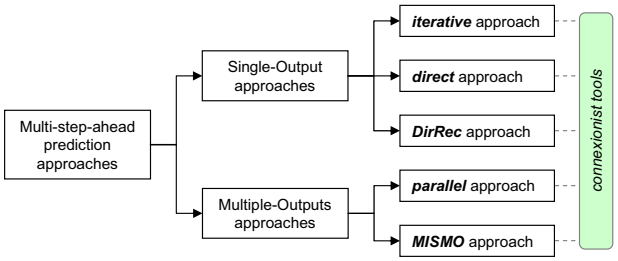

Fig. 2. Taxonomy of multi-step ahead prediction approaches

Note that, as for the name of the approaches, there is no absolute consensus in literature. The "parallel" approach defined in this article is called "joint" approach in [10] and "direct" approach in [11].

B. Iterative approach

The iterative approach is the most common one. Multi-step predictions are provided by 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 the following one and the operation is repeated until the estimation of $\hat{x}_{t+H}$. The procedure is illustrated in Fig 3. Formally:

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

where $\left\{f^1, [\theta^1]\right\}$ states for the one-step ahead prediction model with its parameters set calculated during the learning phase, p the number of regressors used, i.e. the number of past discrete values used for prediction. Note that from the time $h > p$, predictions are made only by using evaluated data, i.e. without observed data.

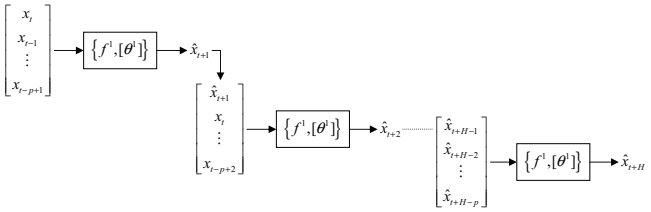

Fig. 3. Iterative approach representation

The iterative approach is the simplest to implement [12], [13]. However this approach suffers from propagation error: the accuracy decreases as the length of the prediction horizon increases [9], [12], [14], [15]. Moreover, this approach does not take into account the temporal behavior [9].

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 eq. 8 and Figure 4).

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

where $\left\{f^h, [\theta^h]\right\}$ states for the model tuned to provide predictions at $t+h$, and p the number of regressors used.

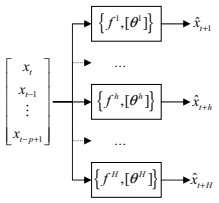

Fig. 4. Direct approach representation

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 [9]. Moreover, this approach is not easy to implement [13].

D. DirRec approach

DirRec approach has been presented by [16]. As shown in eq. 9 and Figure 5, the DirRec approach is similar to the iterative approach except that each prediction step forecasting model is distinct. Learning procedure is quite different since 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 one until all prediction tools are trained.

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

where $\left\{f^h, [\theta^h]\right\}$ states for the model tuned to provide predictions at $t+h$, and p the number of regressors used.

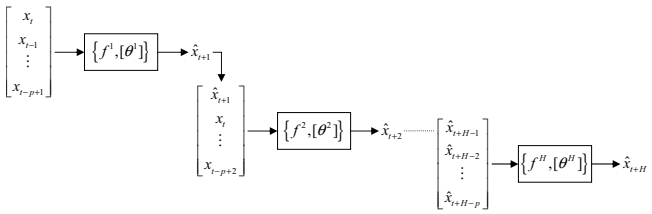

Fig. 5. DirRec approach representation

According to [15], 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 outputs prediction model. This approach calculates all prediction steps with a single model, as illustrated by eq. 10 and Figure 6.

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

where $\{f, [\theta]\}$ states for the model tuned to provide predictions, and p the number of regressors used.

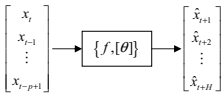

Fig. 6. Parallel approach representation

This approach enables to provide all steps predictions with less computing time than the direct approach since there is only one model to tune [14]. But it raises serious rounding errors: the number of output nodes is equal to the length of the prediction horizon [17].

F. MISMO approach

MISMO (Multiple-input Several Multiple-outputs) approach has been introduced in [18]. This approach consists of several MIMO (Multiple-input Multiple-output) with a parameter (s) that determines the output number for all MIMO. If $s = 1$, this amounts to the direct approach, whereas if $s = H$ this corresponds to the parallel approach.

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

where s is the number of outputs of each model, $m = H/s$ the total number of models, $k \in [1, m]$ the model number. $\left\{f^k, [\theta^k]\right\}$ states for the k^{th} model tuned, and p the number of regressors used.

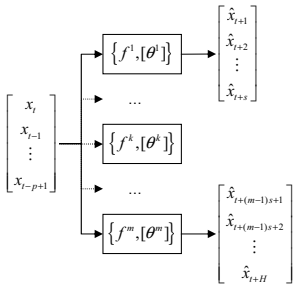

Fig. 7. MISMO approach representation

According to [9], predictions are expected to be 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 MISMO greater adaptability comes at the cost of an additional parameter s.

G. Analysis criteria and discussion

The scientific literature does not suggest the superiority of any approach upon the others ones. 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;
- implementation difficulty.

Prediction approach accuracy. In a general way, the choice of an error measure to quantify the accuracy of predictions has been much discussed (see for example [19], [20]). 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-A, the approaches can be compared if the same basic connexionist tool is used.

Computational complexity. A complexity criterion aims at assessing 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 time when learning than the single-tool approaches (iterative and parallel) since various 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 a more subjective criterion: 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 like this: the parallel approach, the iterative approach, the direct approach, the DirRec approach and finally the MISMO approach.

According to all this, 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 can not be done until tests are made. This is the aim of the remaining paper: the approaches are first compared on prediction benchmarks (section IV) and then on a real prognostic problem (section 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 form the NN3 competition which was provided to test the accuracy of computational intelligence methods (notably neural networks) in time series forecasting [21]. 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 the homogeneous population of empirical time series. Figure 8 depicts time series number 53 and 88 (taken randomly). It allows to show the diversity of data. Note that NN3 has been used in [9] for the same type of analysis as in this paper.

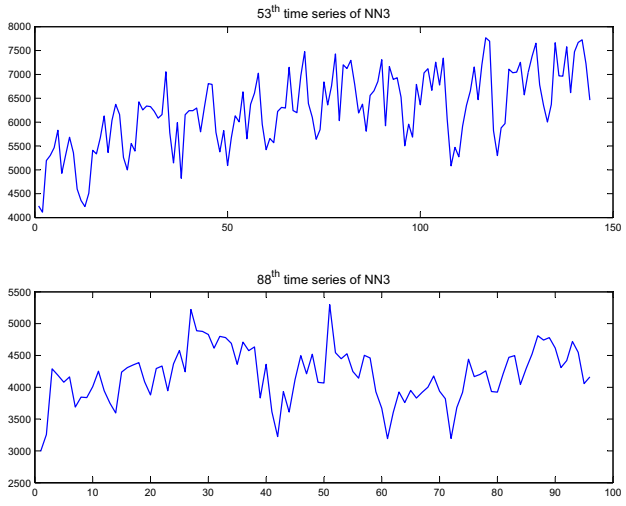

Fig. 8. Two NN3 time series

B. Simulation conditions

Basic connexionist tool. As stated in section II-A, 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 [22], [23]. 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 [24].

Horizon of prediction and set of regressors. Tests were carried out on all 111 NN3 time series, without data processing than a data normalization, and with identical initial conditions for each approach. As for the horizon of prediction (and according to NN3 competition [21]), the last 18 values of each time series where used for test ($H = 18$). In order to extract more solid conclusions 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 RMSE, μ_e and σ_e of the errors of prediction (for the 111 time series tests) (eq. 12). Since 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 before calculating the global accuracy criteria.

$$\begin{aligned} \text{RMSE} &= \sqrt{\frac{1}{H}\sum_{h=1}^{h=H}(x_{t+h}-\hat{x}_{t+h})^2} \\ \mu_e &= \frac{1}{H}\sum_{h=1}^{h=H}(x_{t+h}-\hat{x}_{t+h}) \\ \sigma_e^2 &= \frac{1}{H}\sum_{h=1}^{h=H}(x_{t+h}-\hat{x}_{t+h}-\mu_e)^2 \end{aligned} \tag{12}$$

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.

C. Results and discussion

Simulation results are given in Table I. 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), then the direct and parallel approaches. The iterative approach has the worst results. As the RMSE is an aggregated measure, it does not allow to study the dispersion of the error (couple (μ_e, σ_e)). Figure 9 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 one. Table II enables

TABLE I
RESULTS ON NN3 FROM 1 TO 5 INPUTS

1 input

Approach	RMSE	μ	σ	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	σ	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	σ	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	σ	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	σ	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

to have a closer look on these results. This table depicts the percent of time series that have been best predicted (thanks to the RMSE) with regard to the s parameter of the MISMO approach. According to this table and reminding that if $s = H$, 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 strengthens the parallel approach capacity for prediction accuracy.

As for the execution time (see Table I), the iterative and parallel approaches are equivalently the fastest ones, and thereby the most suitable ones with respect to implementation

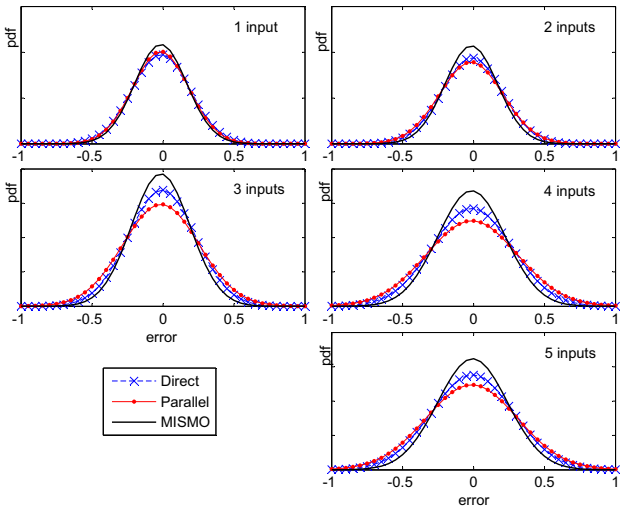

Fig. 9. NN3 error dispertion from 1 to 5 inputs

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

s parameter	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%

constraints. From this point of view, the MISMO approach is the worst one. It balances the results on accuracy and there is no way to optimize both criteria. Figure 10 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.

V. APPLICATION ON A REAL DEGRADATION DATASET

A. Data sets on a real engine health

Developments were applied on the challenge dataset of diagnostic and prognostics of machine faults from the first International Conference on Prognostics and Health Management (2009) [25]. The dataset consisted of multiple multivariate time series (26 variables) with sensor noise (see Figure 11 for an example). Each time series was from a different engine of the same fleet and each engine started with different degrees of initial wear and manufacturing variation unknown by the user. The engine was operating normally at the begining and developed a fault. The fault grew then in magnitude until system failure.

B. Simulation conditions

Among the 26 variables, experiments have been made on the fourth feature. As for the number of inputs, two regressors

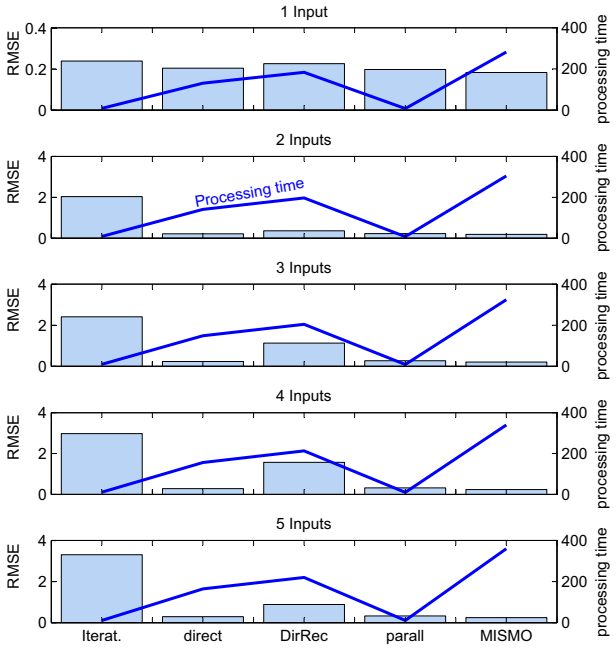

Fig. 10. RMSE versus computing time for various number of inputs

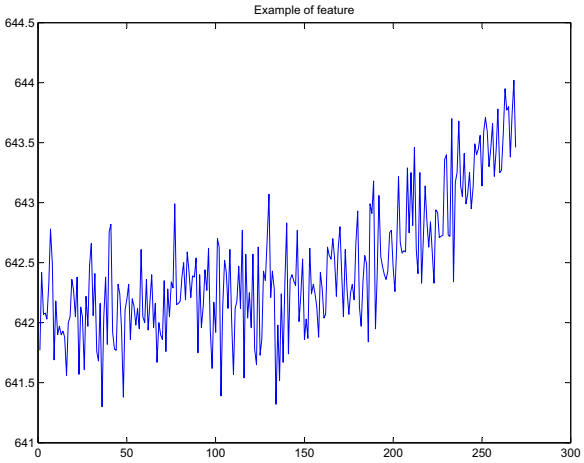

Fig. 11. Real degradation dataset - example of feature

have been used in addition to a third input which is the time index as suggested in [26]. The learning data set was composed of 40 runs from previous degraded systems. During the test, 15 distinct degradations 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 an horizon of 80 steps-ahead (from time 51 to 130).

Considering the results of section 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 DiRec 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.

TABLE III
RESULTS ON REAL DEGRADATION DATA

Approach	RMSE	μ	σ	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

Similarly to the 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. Results and discussion

Figure 12 depicts an example of 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 the parallel approach provides bests results. That can be more closely discussed by considering the whole tests on 15 degradations time series.

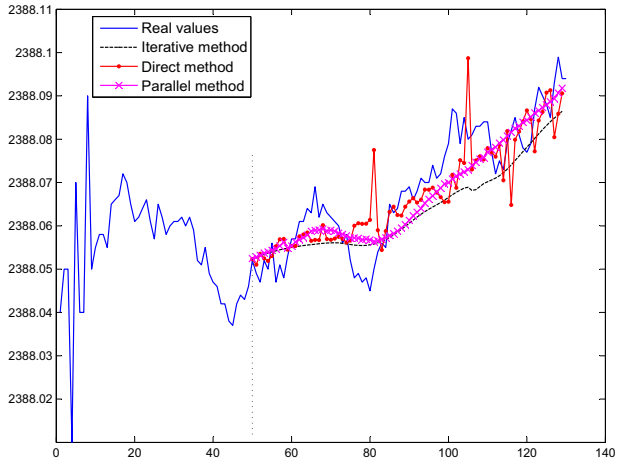

Fig. 12. Real degradation feature - example of predictions

Table III resumes the performances criteria obtained by considering 15 degradations for test. Figure 13 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 is much more long to deploy than the two others approaches (until 120 times). Following that, the most suitable approach for real world problems seems to be the parallel one since it is the one that offers the best compromise between accuracy and complexity. Those experiments confirm the NN3 tests.

VI. CONCLUSION AND WORK IN PROGRESS

The aim of this paper is to try to point out an efficient connexionist-systems-based approaches to ensure long term

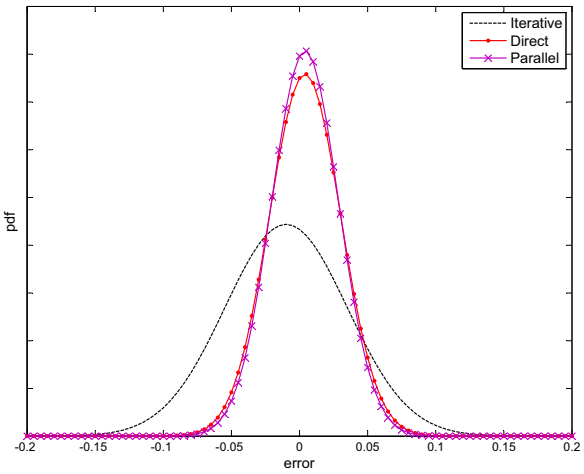

Fig. 13. Error distribution for real degradation data

predictions for prognostics. The paper emphasizes on 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 type 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 were, conclusions are similar. The approach that came out is the parallel approach since it provides a compromise in between accuracy and complexity. However, an important point must be raised. The iterative approach is the only one to be able to predict at any horizon of prediction. Indeed, in the others approaches, the practitioner must set in advance the final horizon of prediction he would like. This can be embarrassing since the time of failure is in essence uncertain. Following that, a good practice could be to implement both approaches: the parallel approach in order to get accurate predictions for a limited critical horizon, and the iterative approach in order 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 scheduled.

REFERENCES

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

[2] A. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mechanical systems and signal processing*, vol. 20, no. 7, pp. 1483–1510, 2006.

[3] A. Heng, S. Zhang, A. Tan, and J. Mathew, "Rotating machinery prognostics: State of the art, challenges and opportunities," *Mechanical Systems and Signal Processing*, vol. 23, no. 3, pp. 724–739, 2009.

[4] 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.

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

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

[7] 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.

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

[9] S. B. 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.

[10] 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.

[11] 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.

[12] 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.

[13] C. Cheng, J. Xie, K. 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.

[14] 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.

[15] V. T. Tran, B. 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, no. 5, pp. 9378–9387, Jul. 2009.

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

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

[18] B. 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.

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

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

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

[22] P. P. Angelov and D. P. 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.

[23] 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.

[24] 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.

[25] 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.

[26] 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.