Multimodal Forecasting Methodology applied to Industrial Process Monitoring

Daniel Zurita, *Student Member, IEEE,* Miguel Delgado, *Member, IEEE, Jesus A. Carino, Student Member, IEEE,* Juan A. Ortega, *Member, IEEE.*

***Abstract*— Industrial process modelling represents a key factor to allow the future generation of industrial manufacturing plants. In this regard, accurate models of critical signals need to be designed in order to forecast process deviations. In this work a novel multimodal forecasting methodology based on adaptive dynamics packaging and codification of the process operation is proposed. First, a target signal is decomposed by means of the Empirical Mode Decomposition in order to identify the characteristics intrinsic mode functions. Second, such dynamics are packaged depending on their significance and modelling complexity. Third, the operating condition of the considered process, reflected by available auxiliary signals, is codified by means of a Self-Organizing Map and presented to the modelling structure. The forecasting structure is supported by a set of ensemble ANFIS based models, each one focused on a different set of signal dynamics. The performance and effectiveness of the proposed method is validated experimentally with industrial data from a copper rod manufacturing plant and performance comparison with classical approaches. The proposed method improves performance and generalization versus classical single model approaches.**

***Index Terms*— Forecasting, Fuzzy neural networks, Industrial plants, Predictive models, Time series analysis.**

I. Introduction

Reliability and safety are becoming critical aspects in the modern industry. In this regard, the industrial sector has made a considerable effort to integrate process monitoring approaches during the last decade [1]. However, providing information regarding future condition of industrial process behavior is becoming critical in order to gain reaction time for the correction of undesired process deviations [2]. Thus, it is being required, in order to work towards the next generation of industrial monitoring approaches, the proposal of accurate forecasting models to be applied over critical process' signals, and drive those models into a specific time horizon in order to obtain future behaviors [3].

The literature published to date shows that forecasting of industrial manufacturing processes, in terms of critical signals evolution for supervision purposes, is still a novel field for research, in which performing methodologies are expected [4]. Indeed, there are two main challenges related to industrial process forecasting. First, the consideration of suitable procedures to deal with highly non-linear signal behaviors, as is the case of most industrial processes, where the correlation of the objective signal with the rest of process information quickly decreases within a short period of time [5]. Second, the assessment and exploitation of such auxiliary information related with the objective signal, which is required to enhance the forecasting performance avoiding computational complexity and model overfitting [6].

For instance, Su *et al.*, in [7], propose the use of an Adaptive Neuro Fuzzy Inference System (ANFIS), to predict the evolution of a non-linear time series. In such work, a non-linear input selection method based on an adaptive expectation method is implemented to select the best suitable inputs for the ANFIS model. However, the proposed input selection method based on single input evaluation over-adapts the set of inputs to the training data, resulting in a clear risk of generalization decreases. The ANFIS based models fuse the parametric adaptability of neural networks and the generalization capabilities of fuzzy logic [8]. Thus, ANFIS based forecasting offers a very reliable and robust condition predictor, since it can capture non-linear input relations quickly and accurately [9]. Indeed, ANFIS based modelling is one of the most used methods for industrial process modelling, but the risk to get trapped in a local minima during the convergence procedure must be considered. In this regard, Zamani *et al.*, in [10], propose the use of an ANFIS modelling for complex non-linear time series forecasting. In such method, a single ANFIS model to forecast a complex non-periodic gas concentration signal is used. Indeed, the single-model approach corresponds to the easiest procedure used in literature to handle complex signal's dynamics, since it corresponds to the use of the whole raw data inputs during the training of algorithm. However, in such approaches the modelling scheme is not able to learn the variability of the signal and the architecture undergoes a loss of performance leading the system to an over-fitted response.

In order to deal with such problems, recent studies pointed out that splitting the target signal in order to be modelled in different modes is a suitable approach when dealing with non-linear time series. Indeed, different multi-scale signal decomposition approaches are being proposed [11]. Hooshmand *et al.*, in [12], proposes the combination of Wavelet Package Decomposition (WPD), and ANFIS, to perform an electric load forecasting. In such approach, WPD is used to extract high and low frequency modes of the signals. Then, a dedicated ANFIS model for each set of frequencies is

This research was supported in part by the Spanish Ministry of Education, Culture, and Sport under the grant FPU13/00589. The authors also wish to acknowledge financial support from the Spanish Ministry of Economy and Competitiveness (TRA2016-80472-R).

D. Zurita, M. Delgado, J.A. Carino, and J.A. Ortega, are with the MCIA Research Center, Department of Electronic Engineering, Technical University of Catalonia (UPC), Terrassa, 08222, Spain. E-mail: [daniel.zurita, miguel.delgado, jesus.carino, juan.antonio.ortega] @mcia.upc.edu.

considered. The authors state that decomposing the signal outperforms classical single-model approaches. However, the main limitation of such multi-scale decomposition is that the configuration of the filters must be specified by the user, which requires a deep knowledge of the process responses.

Alternatively, Empirical Mode Decomposition (EMD), offers an adaptive capability during the signal decomposition process, resulting in a collection of Intrinsic Mode Functions (IMF), and a residue. Wei, in [13], propose the combination of EMD and ANFIS modelling. The author states that the combination of EMD and ANFIS improves the resulting performance of a time series forecasting in signals with high variability, that is, multiple oscillation modes. However, as presented in [14], [15], the IMF are directly decomposed and modelled, that is, one forecasting model is required for each signal partitioning, which represents a high computational-burden strategy and, moreover, the simplicity of modelling some of the IMF can lead the corresponding model and, then, the global forecasting performance, to an intense overfitting.

In this work, a novel forecasting methodology for industrial process' signal modelling is proposed. In this regard, contribution of this work lies on the validation of a multimodal forecasting approach, in which the outcome estimation is carried out from the combination of multiple model outcomes, that manage different signal dynamics while preserving generalization capabilities. Novelties of this work include, first, the signal decomposition of the target signal by EMD and the proposal of an adaptive dynamics packaging procedure in order to define the number of required models. Second, a non-linear mapping procedure of the available auxiliary signals in order to reduce the dimensionality of the ANFIS convergence problem. And, third, the combination of multiple model outputs to obtain the global forecasting outcome. Indeed, the main contribution of the proposed method consists on relating each one of the dynamics of the signal with its relative significance with the contribution of each dynamic to the modelling error. Therefore, it is proposed to forecast industrial time series considering a relation among three factors, the dynamics, their significance and their associated modelling error. Furthermore, the way auxiliary signals are combined represents a step forward of the classical optimization based input selection approaches. Note that this study highlights the necessity of analyzing signal dynamics in multimodal modelling forecasting approaches. The feasibility of the proposed method is validated over a real case study, a copper rod manufacturing process, and the results are compared with classical forecasting approaches.

II. Definition of the Method

The proposed method is shown in Fig. 1. Two main procedures are considered to enhance the modeling performance: (i) the addition of an adaptive processing step that decompose, analyzes and packages signal's dynamic modes, and (ii) the codification of the auxiliary information in regard with the process operating condition. Thus, a three-step multimodal forecasting method is proposed, in which different sets of IMF are packaged and modeled separately with dedicated ANFIS models, Step 1, process operating information is codified by means of SOM, Step 2. Finally, considering the superposition properties of the resulting IMF of the EMD analysis, the outputs of the models are combined to obtain the forecasting estimation, Step 3.

A. *Adaptive dynamics packaging*

Considering an objective signal decomposition in a set of IMF, the resulting accuracy of a single model for a joint subset of IMFs exhibit, in general, a non-linear decrease with the number of IMF considered, as illustrated in Fig. 2. Indeed, the increase of the modelling error is due to the limitation of an ANFIS model to define proper weighting matrices dealing with wideband signals and non-linear relations among inputs [11].

In this regard, the aim of the proposed method is the analysis of a coherent signal decomposition strategy by analyzing the significance of each IMF in order to package them in different sets for modelling performance enhancement. Indeed, such coherent packaging of the IMF represents the multimodal approach proposed to increase modelling performance while optimizing the number of required models, which allows a general modelling approach, instead of the use of a single model for every IMF that leads in a huge computational burden and in a lack of generalization.

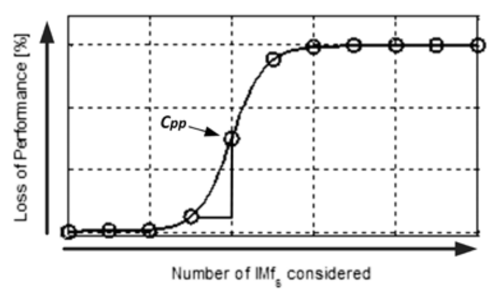


Fig. 2. Representation of the effect over the resulting accumulated error of the number of IMF considered for a single model. The critical performance point, C_{pp}, represents an allowable down limit in regard with the modelling performance. This curve is obtained by the successive evaluating of the IMFs extracted from multi-sinus signal of 11 frequencies in a classic 1-input 1-output ANFIS structure at a forecasting horizon of 10 samples.

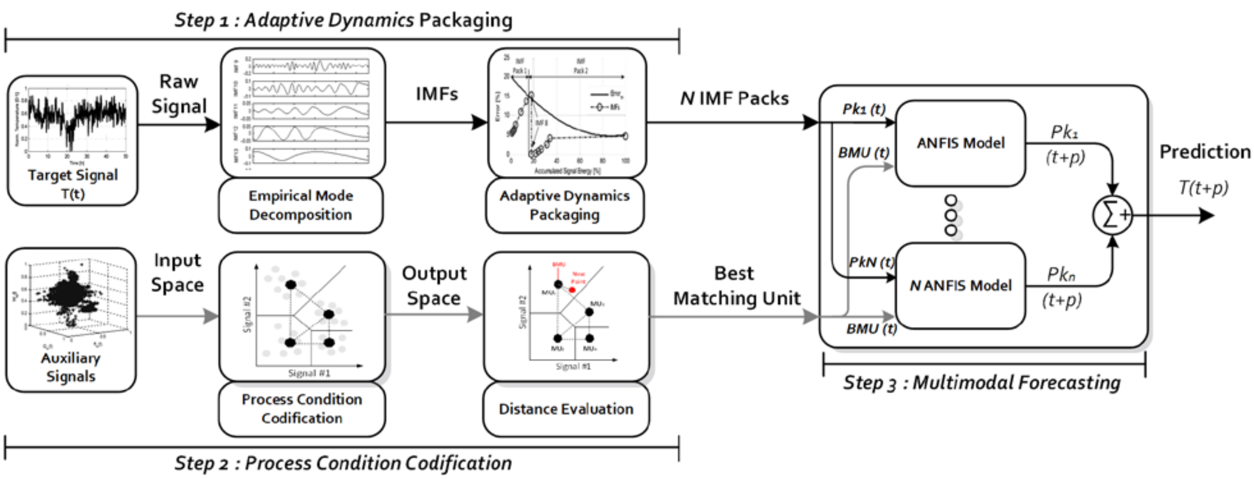


Fig. 1. Block diagram of the design procedure of the proposed multimodal modelling method. The procedure is divided in three steps: (i) Adaptive dynamics packaging, (ii) Process condition codification, and (iii) Multimodal forecasting.

Then, the problem turns into the identification of the optimum number of IMF that should form each package. As seen in the Fig. 2, it is expected a point in the number of considered IMF in which the accumulated error increases suddenly. This point is considered to be the Critical Performance Point, C_{pp}, and it is considered to define the end of an IMF package to be considered by a single model. Thus, the developed method proposes the identification of successive C_{pp} during the required packaging iterations till all IMF take part of one package to be modelled. Such C_{pp} is mathematically described by the proposal of an error threshold function, E_{TH}. That is, the E_{TH} defines the allowable model error curve in regard with the significance, in terms of relative energy contents of the considered IMF compared with the original signal, Re_j, and the relative modelling error achieved by the combination of IMFs evaluated in an auxiliary model, Er_j. Considering this, E_{TH} is defined as a decreasing second order function, as shown in Eq. (1).

$$\mathrm{E_{TH}}(Re_i, Er) \leq A \cdot Re_i^2 + B \cdot Re + C \qquad (1)$$

The industrial applicability has been taken into consideration for the mathematical definition of the error function, where, generally, low frequency modes represent long-term process behaviors. Therefore, this function has been designed to be permissive in terms of error with low-significant modes, which in most cases correspond to the higher frequencies contents of the signal under analysis, and restrictive with the high-significant modes, that usually represent the low frequencies and main trends of the signal. Note that the significance is quantified by the amount of energy that accumulates a certain IMF in regard with the original signal.

Thus, taking into account a specific application, the parameters A, B and C, used to define the error threshold function, would be identified by means of interpolation, as it is shown in Fig. 3. Three points are needed to allow the regression, that are: (i) the maximum allowed error for the low-significant modes, LW_{MAX}, (ii) the maximum allowed error for the high-significant modes, HG_{MAX}, and (iii) the smoothing factor, S_m, which fixes the decay of the curve.

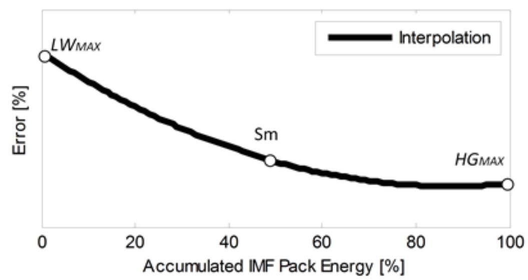


Fig. 3. Example of a resulting error threshold curve in regard with the accumulated energy, R_e, of the IMF package. The curve is described by the interpolation of a second order function within the three points defined, LW_{MAX}, S_m and HG_{MAX}].

Thus, in order to carry out the IMF packaging procedure, the modelling errors must be estimated and, then, compared with the predefined error threshold curve. For this procedure, a simple model is proposed, that is, a classical time series 1-input ANFIS structure, using current value of the objective signal as an input of the model, and the predicted value as the output. This approach represents a trade-off between simplicity and performance analysis. Then, when the model error surpasses the corresponding threshold, the last intrinsic mode function is removed from the model, and the fist IMF package is closed. Then, the procedure starts again with a new model consideration, and the iterative addition of the rest of the IMF till the N packs are formed. The objective is to obtain N IMF packages of which dynamic combination is affordable by a simple ANFIS model.

In order to carry out the IMF packaging procedure an iterative process is defined as follows: the EMD is applied to extract the k different IMF. Let $Pk_j(t)$ be the j-*th* package of IMF, where $j \epsilon [1..N]$ and contains IMFs from ki_j to kf_j, then,

1. In each iteration j, $Pk_j(t)$ is formed by Eq. (2). Note that in first iteration the method initializes by $ki_1 = kf_1 = 1$.

$$Pk_j(t) = \sum_{k=ki_j}^{kf_j} IMF_k(t) \qquad (2)$$

2. $Pk_j(t)$, is modelled by a 1 input ANFIS model in order to obtain the predicted output, $\widehat{y_j}(t)$, at a time horizon p, Eq. (3). The model uses as inputs the current value of the signal, $Pk_j(t)$.

$$\hat{y_j}(t) = ANFIS\left(Pk_j(t-p)\right) \quad (3)$$

3. The model performance is evaluated and the prediction error, E_r, is calculated by Eq. (4).

$$E_r = \frac{1}{L}\sum_{t=1}^{L}\frac{|y_j(t) - \hat{y_j}(t)|}{|y_j(t)|} \quad (4)$$

4. The significance in terms of relative energy, Re_j, of $Pk_j(t)$ versus the complete signal $x(t)$ is obtained by Eq. (5).

$$Re_j = \frac{\sum|Pk_j(t)|}{\sum|x(t)|}\cdot 100\ \% \quad (5)$$

5. The error threshold function evaluated in regard with both calculated points, *Re* and *Er*.

If the performance is under the allowable error defined in Eq. (1), another intrinsic mode function should be added to the model, $kf_j = kf_j + 1$. Then, steps from 1 to 5 are repeated till the consideration of the *k* IMF. If the point is above the curve means that $Pk_j(t)$ should be closed with the IMF added in the previous iteration. Therefore, the number of IMF is decreased by $kf_j = kf_j - 1$; and the initial intrinsic mode function of the next package, $Pk_{j+1}(t)$, is prepared by $ki_{j+1} = kf_{j+1} = kf_j + 1$. Then, $j=j+1$ and the algorithm returns to step 1. At the end of this procedure, all the IMF will be distributed in *N* packages.

B. Process condition codification

Generally, dealing with an industrial process, a set of auxiliary signals are complementary available with the target signal to be modelled, that all together define the process condition. These signals present a great potential to be informative enough to the modelling process However, the direct introduction of such signals in the model algorithm may cause an unnecessary growth of the number of model inputs, increasing model complexity and the amount of data required to achieve a proper convergence during the learning phase. This problem is critical dealing with ANFIS models, since modelling efforts are focused to adapt the membership functions to the input distributions [16]. For this reason, auxiliary inputs must be pre-processed to remove non-significant and redundant information.

In this way, the feature reduction process has been typically implemented with linear techniques such as Principal Component Analysis (PCA) [17]. However, PCA technique has been discussed by many authors emphasizing its limitation dealing with large data sets, because it seeks for a global structure of the data [18]. The information contained in a *D*-dimensional space mostly has a nonlinear structure. Concerning with this problems, manifold learning methods have been applied in the last years to preserve this information [19]. Among them, Self-Organizing Maps (SOM) is the most used, which is based on developing a neural network grid to preserve most of the original distances between feature vectors representations in the input space [20]. Such space is initially predefined as a regular D-dimensional grid [21] then, SOM adapts this grid to data distribution defined by the auxiliary signals. In Fig. 4, the topology preservation mapping is illustrated. Prior to the training, the grid is defined, Fig. 4a. During the training, Fig. 1b, the grid successively adapts itself in order to preserve as much as possible the topology described by the original data. Finally, the resulting grid is evaluated over a new data, Fig. 4c. Thus, for a new data point, the Euclidean distance, to each neurons in the *D*-space is calculated. The neuron with the shortest distance is considered to be the Best Matching Unit, BMU. All the coordinates of the point are mapped in the number of BMU, providing to the SOM the capability of data codification. Indeed, SOM can be seen as a neural network that non-linearly discretizes the input data space and codifies such partition into the BMU number.

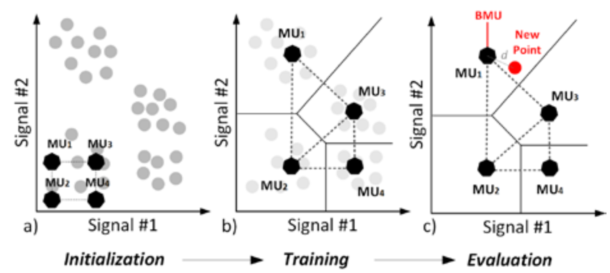


Fig. 4. SOM procedure to codify the input space.

Then, from the proposed method point of view, the available auxiliary signals are presented to the trained SOM, and a corresponding BMU is obtained. As seen in Eq. (6), the BMU represents a discretized signal that summarizes process condition by the auxiliary signals.

$$BMU(t) = SOM\left(Aux_1(t), \dots, Aux_q(t)\right) \quad (6)$$

C. Multimodal forecasting by Neuro Fuzzy Inference

The resulting *N* ANFIS models are considered, that corresponds to the *N* packs of IMF found. For each ANFIS model, the proposed forecasting follows Eq. (7). The proposed structure for the *N* resulting models corresponds to a 3 input – 1 output scheme. The considered inputs are: (i) the current value of the IMF package, $P_k(t)$, (ii) *a* past value of the signal in order to have a reference of its tendency, $P_k(t-n_1)$, and (iii) the codification of the auxiliary signals condition, *BMU (t)*. The considered model output is the resulting signal at *p*, $P_k(t+p)$.

$$P_k(t+p) = ANFIS(\,P_k(t), P_k(t-n_1), BMU(t)\,) \quad (7)$$

In order to compare the proposed method performance with classical approaches, a set of standard metrics have been considered. The Root Mean Squared Error (RMSE), reflects the standard deviation of the prediction error. The Mean Absolute Error (MAE), measures the average error. The Mean Absolute Percentage Error (MAPE), reflects the average deviation of each observation divided by the signal amplitude, it exhibits the percentage precision of the modelling. Such performance metrics are defined in Eq. (8) to (10), where *L* corresponds to the number of samples.

$$RMSE = \sqrt{\frac{\sum_{t=1}^{L}\left(y(t) - \hat{y}(t)\right)^2}{L}} \quad (8)$$

$$MAE = \frac{\sum_{t=1}^{L}|(y(t) - \hat{y}(t))|}{L} \quad (9)$$

$$MAPE = \frac{\sum_{t=1}^{L}\left|\frac{(y(t) - \hat{y}(t))}{y(t)}\right|}{L}\cdot 100\% \quad (10)$$

III. COMPETENCY OF THE METHOD

The proposed method is validated using industrial data collected from a Spanish metallurgy company, La Farga, specifically, from a high purity Copper Rod Manufacturing Process (CRMP). Such industrial system represents a challenging scenario for process forecasting due to the non-stationary operating conditions and the non-linear relation among the process signals.

A. Experimental plant definition

The copper rod manufacturing process is presented in Fig. 5. CRMP process is divided in five main elements: (i) the shaft furnace, is a vertical natural gas fired furnace in charge of melting the input high purity copper cathodes. (ii) The holding furnace, acts as a lung for the copper melting process which aim is to provide a constant flow of copper to the rest of the process. (iii) The tundish is a ceramic valve that controls the melted copper flow to the rest of the process. (iv) The casting wheel is in charge of solidifying the melted copper by a heat extraction process. It uses a water-cooled steel band that encloses the casting cavity in which the molten copper solidifies to form a raw rod. Both casting wheel and the steel enclosure are refrigerated by means of a water cooling circuit. (v) The roughing mill, reduced the diameter of the raw copper rod to meet the specified diameter conditions fixed by the plant operators. Finally, copper rod is coiled and packed giving the final manufacturing product.

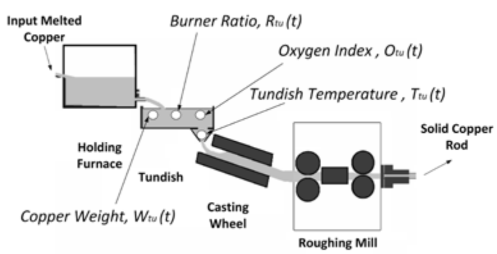


Fig. 5. Diagram of the copper rod manufacturing process [22].

Indeed, the manufacturing process implies the transformation of the melted input copper in a solid copper rod. Such transformation is based on a controlled solidification process. Indeed, the solidification of the copper is a critical aspect within the manufacturing process, in which the heat must be properly extracted from the copper bar. The objective of this application is to forecast the tundish temperature, $T_{tu}(t)$. This magnitude is critical in the CRMP since it is the last measurement of the melted copper before starting the solidification procedure. Non-expected variations of this temperature imply imperfections in the final product due to non-uniformities in the copper density. The modelling and forecasting of such temperature allows reducing the affectation of such deviations to the next manufacturing batch. Available information in regard with the considered part of the process is: the oxygen level, $O_{tu}(t)$, the weight of the melted copper, $W_{tu}(t)$, and the ratio of air/gas of the burner, $R_{tu}(t)$. All signals and their description are shown in Table 1.

TABLE 1. PROCESS SIGNALS AND THEIR DESCRIPTION

Avb.	Description
T_{tu} (t)	Tundish temperature [ºC]
O_{tu} (t)	Oxygen concentration in the melted copper [ppm]
W_{tu} (t)	Weight of the melted copper inside the tundish [kg]
R_{tu} (t)	Ratio of air vs gas from the burner of the tundish [%]

All signals are acquired synchronously, and are automatically stored in a standard SQL database at a period of 10 seconds, a sampling frequency, f_s, of 0.1 Hz. The available data set correspond to 96 hours of consecutive plant operation. The target signal, the tundish temperature among this operating time is shown in shown in Fig. 6. In this regard, the first 48h are used for training purposes, while the remaining 48h are used for testing. The forecasting horizon, p, represents one of the most critical parameters to be defined. In this application the forecasting horizon is fixed by the application requirements of detecting deviations with high resolution within one manufactured element time. Therefore, p is related with the copper rod manufacturing time, that is 15 minutes, that is, p = 90 samples.

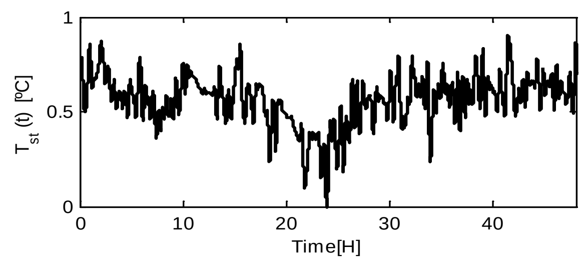


Fig. 6. Tundish temperature, $T_{tu}(t)$, considered in the training set of 48 hours of process operation.

B. Target Signal Decomposition and Packaging

Considering the continuous operation requirements of the proposed method, the EMD is approached by means of the utilization of a time based buffer. The corresponding time window is selected in regard with the lowest dynamic modes contained in the objective signal, that is a temporal window of 4.5 hours (1620 samples). Thus, the EMD is applied to $T_{tu}(t)$*and a* total number of k = 9 IMF are obtained. An example of a resulting IMF decomposition is shown in Fig. 7.

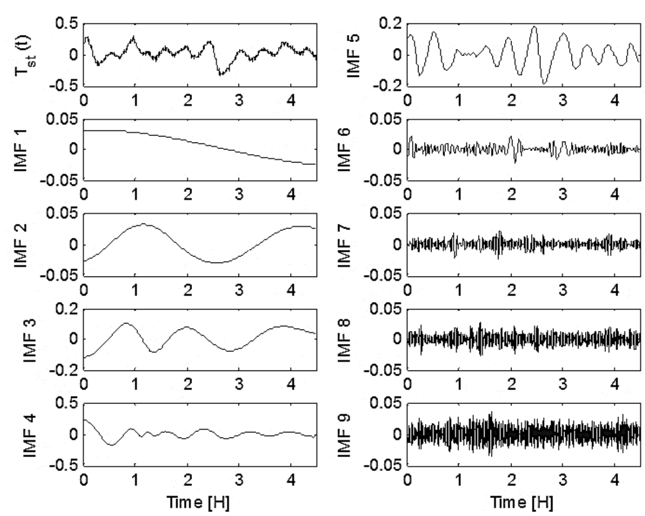


Fig. 7. Intrinsic mode functions extracted from the tundish's temperature of the training set during a time window processing.

Following the method, the IMF are successively evaluated over an error threshold curve to find the optimum number of N IMF packages. The E_{TH} has been defined by fixing the low significance error, LW_{MAX}, at *20%*, the smooth parameter, S_m, at *7.5%*, and the high significance dynamics error, HG_{MAX}, at *5%*, that results in $A= 0.002$, $B= 0.35$ and $C= 20$ following Eq. (1).

The procedure to form the first package of IMFs, $Pk_{j=1}$, starts by evaluating IMF_1 in terms of relative energy and error, Eq. (4), versus E_{TH}. As the threshold is not met, consecutive IMF are successively added to $Pk_{j=1}$. The addition of IMF_4 in $Pk_{j=1}$, shows an error value over the admissible threshold. Thus, $Pk_{j=1}$ is formed by the three first IMF, from IMF_1 to IMF_3. Then, the second package, $Pk_{j=2}$, starts with IMF_4 and successively evaluates all modes till IMF_9. Since the error threshold is not reach anymore, $N=2$ packages. This analysis is shown in Fig. 8. The resulting packages, $Pk_j(t)$, are shown in Fig. 9.

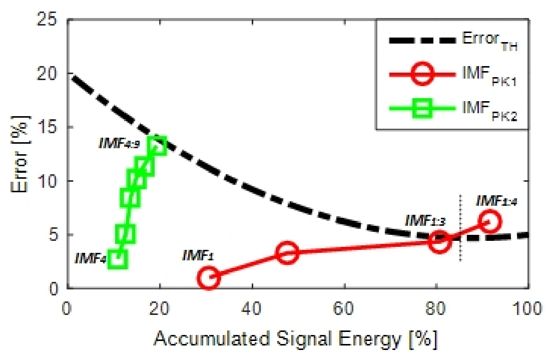


Fig. 8. Evaluation of the training set versus the E_{TH} to find the number N of IMF packages.

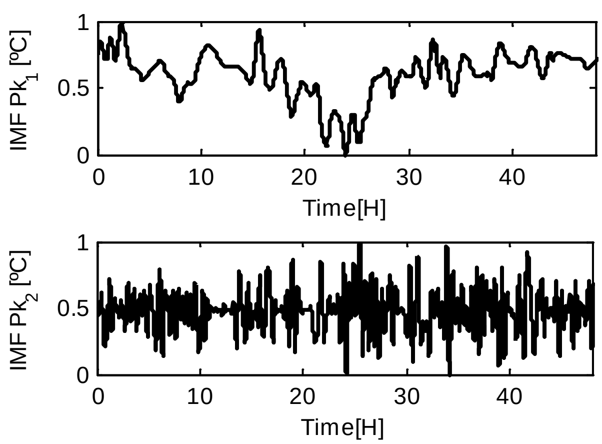


Fig. 9. Resulting IMF packages to be modelled. (a) Package j=1 containing the summation of IMFs from 1 to 3. (b) Package j=2 containing the summation of IMFs from 4 to 9. It should be noticed that a pack of IMFs is the direct combination off all the IMFs that form the package.

C. Mapping of Auxiliary Signals

Fig. 10(a) shows the input space formed by the auxiliary signals, and Fig. 10(b) the resulting 3-dimensional SOM grid. The SOM grid has been configured by a hexagonal distribution 15x15, that is, a total of 225 units. The SOM has been initialized and trained by a batch algorithm and a total amount of 100 epoch were performed.

As it can be seen, the data distribution process presents a central area with high density of data which corresponds to the main operating condition. However, there is a dispersion in each axis around the center, which means deviations from the nominal values reflected in the auxiliary signals.

The resulting set of BMU for the training data is shown in Fig. 11. As it can be seen, most of the data presents a BMU value from 50 to 120. This interval corresponds to the central cluster identified as the most common operating condition. Values under and over this interval are considered as the variations over the normal operating condition. Indeed, the training set exhibits a variation of the temperature around the 22^{th}-28^{th} hour of operation that match with the behavior seen in Fig. 6, which confirms the relation between the tundish temperature and the process operating condition codified by means of the auxiliary signals.

D. Ensemble ANFIS based forecasting

In order to train the N = 2 models, the n_1 value for the third input parameter must be selected. For this aim, the autocorrelation analysis is proposed [23]. In this regard, such analysis compares the correlation of a target signal with the same signal but with an iterative delay added. It is used to show periodicities and oscillation modes in the analysed signal [24]. As shown in Fig. 12, the resulting signal from Pk_1 shows a smoother autocorrelation decrease indicating that the loss of performance is smoother in regard with the forecasting horizon.

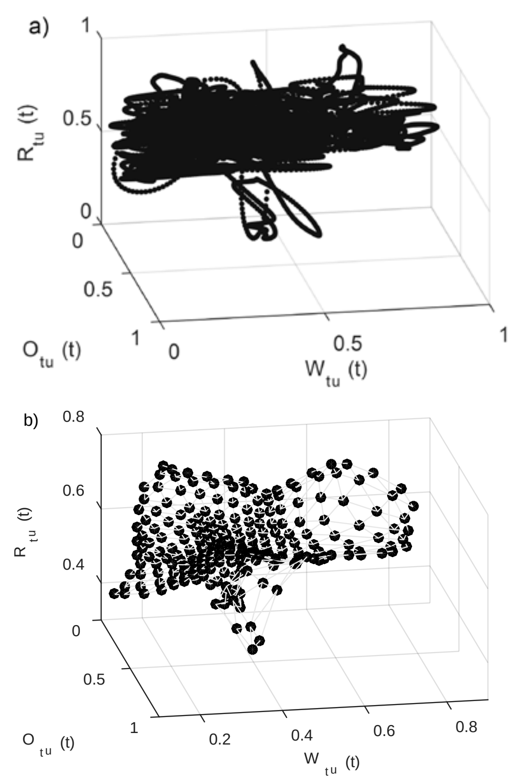


Fig. 10. a) SOM input data space of training set: x) Tundish's Oxigen value in ppm, y) Tundish burner's ratio in % and z) Copper weight in the tundish in Kg. b) Neurons of the SOM grid after the training procedure.

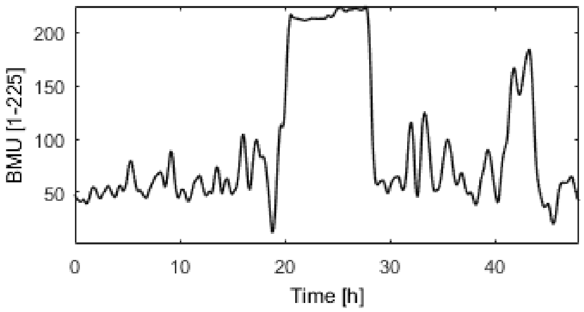


Fig. 11. Resulting best matching units for the training set.

Otherwise, the resulting signal from Pk_2 exhibits a sudden drop of autocorrelation till sample 60th. This response is characteristic of high frequency modes, that are included in the Pk_2. That is, such difference in the autocorrelation response between Pk_1 and Pk_2 relies on the predominant low and high frequency contents, respectively. The temporal values that show an autocorrelation value over 0.4 have been considered as proper n_1 delays. A Genetic Algorithm (GA) [25], has been applied to find such delays for each package, using the forecasting performance of the ANFIS as cost function. The resulting delays are n_1=11 for $Pk_{j=1}$ and n_1=5 for $Pk_{j=2}$, that corresponds to a delay of 110 and 50 seconds respectively.

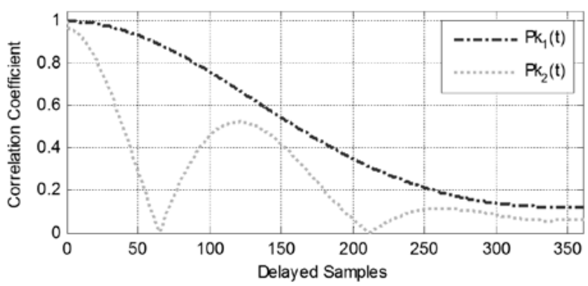


Fig. 12. Autocorrelation analysis between the IMF packages (Pk_1 and Pk_2), and their successive delayed signals.

Finally, in order to complete the ANFIS design, each input is normalized with the min-max scalling method [26], obtaining input signals in the range 0 to 1. The inputs are fuzzified by means of three generalized bell-shaped membership functions. The model is trained for 15 epochs by means of the classical hybrid learning algorithm, which is the combination of the least-squares method and the backpropagation gradient descent method. A k=4 k-fold cross validation method has been used in order to perform both training and validation test [27]. Such procedure has been configured with a window of 48h with an overlapping of 12 hours. The forecasting performance during the training can be seen in Fig. 13, and the corresponding validation in Fig.14(a). The performance' metrics of the model are shown in Table 2. The results show that the proposed modelling structure fits significantly both training and validation data sets with a MAPE error lower than 10%. Therefore, this performance represents a competitive forecasting, it should be noted that most of the error is located at higher frequency components contained in the $Pk_2(t)$. The low values shown by the RMSE implies significant generalization capabilities, that is, a smoothed response versus the outliers. Furthermore, the model exhibits low variations during the validation procedure reflected in low standard deviation achived after the cross validation procedure.

E. *Comparison with other methods*

The performance of the proposed method, M_1, has been compared with three approaches found in the literature, a ANFIS with a genetic algorithm, M_2. A dedicated ANFIS model for each IMF resulting of the EMD decomposition, M_3, and finally, a Neural Network modelling method, M_4.

Therefore, M_2 consists on a GA-ANFIS structure that uses an input selection method based on a GA in order to select the most suitable inputs. According to the literature, the cost function is usually based on the MAPE estimation of the model against the validation data set [28], [29]. In this study the GA has been configured to select the best inputs from [$T_{tu}(t-n_1)$, $T_{tu}(t-n_2)$, $R_{tu}(t)$, $O_{ut}(t)$ *and* $W_{tu}(t)$]. Note that the current value $T_{tu}(t)$ is always introduced as an input of the model. The chromosomes of the GA are configured in regard with the kind of input. For the first two, the past values, the limits of the GA have been configured to vary between 1 and 90 samples, for the rest of signals, binary inputs are used in order to incorporate or discard the signal as an input of the model. After the application of the GA, the best selected inputs and the structure of the final model is defined in Eq. (11).

$$\widehat{T_{tu}}(t+p) = GANFIS(T_{tu}(t), T_{tu}(t-8), R_{tu}(t)) \qquad (11)$$

In order to implement the M_3, a multi-model approach, the tundish's temperature is decomposed in IMF. A dedicated ANFIS model is generated for each IMF, which means a total number of 14 models [14], [15]. The model for each IMF uses as inputs the current value of the *k-th* IMF, $IMF_k(t)$, and the best delay found in the correlation of the target signal, $IMF_k(t-n1)$. Accordingly, the model is defined as in Eq. (12).

$$\widehat{T_{tu}}(t+p) = \sum_{k_i=1}^{k} ANFIS(IMF_k(t), IMF_k(t-n_1)) \qquad (12)$$

Finally, to implement M4, a NN based model has been applied using both target signal and auxiliary signals as inputs [30], as shown in Eq. (13). The NN is trained by using the training data and the class information of each observation. In this regard, a multi-layer NN has been configured with two hidden layer, each one is composed of 10 hidden neurons. The neurons have been configured with a sigmoid activation function, which is usefool in order to smooth the network response.

$$\widehat{T_{tu}}(t+p) = NN(Ttu(t), Otu(t), Wtu(t), Rtu(t)) \qquad (13)$$

As a result, M_2, exhibits the worst results for both training and validation datasets with a MAPE greater than 15% in both datasets. As expected, by means of a single model, it is difficult to forecast the behavior of such a non-linear time series. Furthermore, as the cost function is made in regard with the validation set, the model shows an overfitted response for the validation. This fact, causes the model to approximate better the validation set than the training set, that is, MAPE in validation is lower than in the training. In regard with M_3, it performs slightly better during the training procedure. As it can be seen in the table, it only presents a MAPE of 8%. However, the model exhibits a loss of performance when dealing with the forecasting of the validation set, as can be appreciated in Fig. 13(c). This means that the model suffers a lack of generalization by predicting each IMF individually compared with the proposed method. M_4 achieve a smooth response that is able to follow the target signal in its mean value, however, further resolution is missing in the extrema values. Even tough, it presents better performances than M_2 by mixing together all the significant signals during the modelling. However, the performances are far away from the decomposition approaches. It should be noticed that in terms of error statistics, both M2 and M4 present a more stable response since they are concentrated on following the mean value of the signal, such fact is reflected in a low error deviation in the iterations. However, M3 presents a more unstable response due to an overfitting phenomenon during the training procedure in every fold evaluated.

IV. CONCLUSIONS

This work presents a forecasting methodology applied to a copper rod manufacturing process based on a novel multimodal approach. There are four important aspects in this new method. The first one is the consideration of empirical mode decomposition as adaptive non-stationary signal analysis. The proposed dynamics packaging allows to group together reduced sets of intrinsic mode functions avoiding loss of modelling performance.

TABLE 2. PERFORMANCE OF THE STATISTICAL ERROR METRICS BY 4-FOLD CROSS VALIDATION.

	M1 - Prop. Method			*M2 GA ANFIS*			*M3 EMD ANFIS*			*M4 NN*		
	Trn.	*Val.*		*Trn.*	*Val.*		*Trn.*	*Val.*		*Trn.*	*Val.*	
	$\bar{x}$	$\bar{x}$	σ	$\bar{x}$	$\bar{x}$	σ	$\bar{x}$	$\bar{x}$	σ	$\bar{x}$	$\bar{x}$	σ
RMSE	0.055	0.076	0.005	0.106	0.129	0.011	0.052	0.097	0.006	0.104	0.128	0.011
MAE	0.307	0.259	0.03	0.283	0.321	0.05	0.214	0.312	0.1	0.289	0.331	0.06
MAPE	7.95%	9.32%	1.5%	26.25%	21.52%	2.36%	7.18%	17.75%	5.89%	13.40%	19.84	2.21%

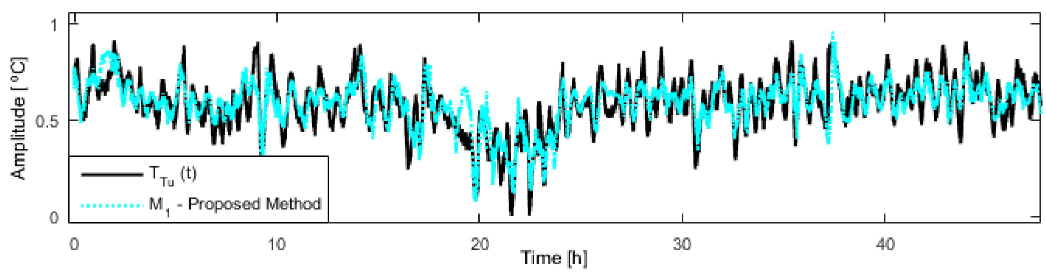
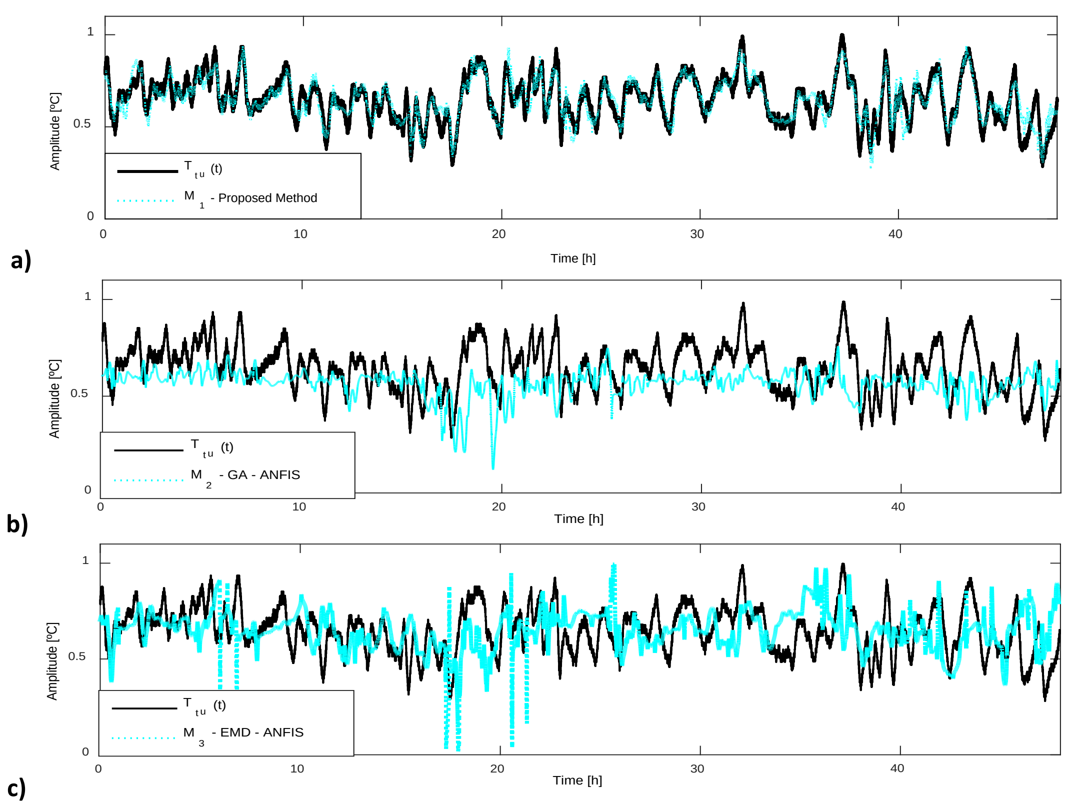


Fig. 13. – Example result of the he proposed method, M_1, during the first fold iteration in the training procedure.



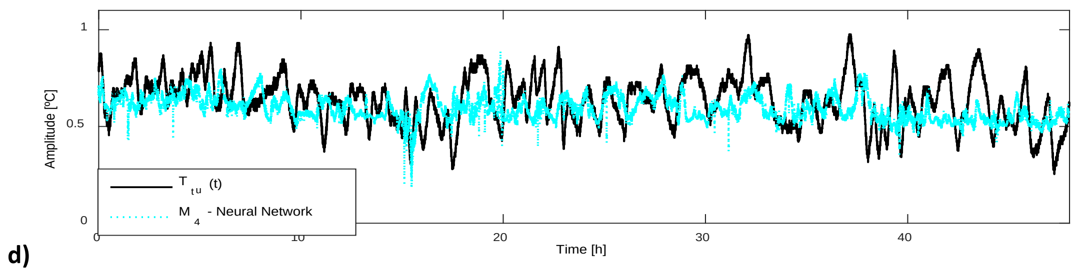


Fig. 14. – Example result of the fourth fold iteration during the validation procedure. Results of the forecasting models: (a) Output of the proposed method, M_1. (b) Comparative of the M_2 - GA-ANFIS. (c) Comparative of the M_3 - EMD-ANFIS a. (d) Comparative of the M_4 – NN.

The second, is the exploitation procedure of available auxiliary process signals. A self-organizing map codification is proposed to project the process operating condition in just one-dimensional variable to be added to the forecasting model. It has been checked how the codification of the auxiliary signals helps to learn the different process behaviors and control actions improving with it the generalization capabilities of the modelling. Third, is the multimodal ANFIS structure, in which the resulting forecasting outcomes of the models are combined to obtain the final forecasted signal value.

Industrial data from a copper rod manufacturing process has been considered, which represents a significant experimental scenario for validation, including a sets of 96-hours of operation for training and validation. Under these experimental conditions, the proposed methodology shows excellent performances, mainly, in terms of low modelling error and generalization capabilities. Additionally, a comparative analysis with state of the art methods has been carried out. The performance of the proposed method exhibits a decrease of 50% of mean absolute percentage error compared with single model approaches.

It should be noted that the proposed methodology could be applied in multiple complex industrial processes. The potentiality of the adaptation and codification capabilities of the proposed method allows an extensive applicability.

V. References

[1] S. Yin, X. Li, H. Gao, and O. Kaynak, "Data-Based Techniques Focused on Modern Industry: An Overview," *IEEE Trans. Ind. Electron.*, vol. 62, no. 1, pp. 657–667, Jan. 2015.

[2] Z. Ge and J. Chen, "Plant-Wide Industrial Process Monitoring: A Distributed Modeling Framework," *IEEE Trans. Ind. Informatics*, vol. 12, no. 1, pp. 310–321, Feb. 2016.

[3] K. Zhou, "Industry 4.0: Towards future industrial opportunities and challenges," in *2015 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, 2015, pp. 2147–2152.

[4] S. Yin, S. X. Ding, X. Xie, and H. Luo, "A Review on Basic Data-Driven Approaches for Industrial Process Monitoring," *IEEE Trans. Ind. Electron.*, vol. 61, no. 11, pp. 6418–6428, Nov. 2014.

[5] F. Guo, L. Lin, and C. Wang, "Novel continuous function prediction model using an improved Takagi–Sugeno fuzzy rule and its application based on chaotic time series," *Eng. Appl. Artif. Intell.*, vol. 55, pp. 155–164, 2016.

[6] W. Wang, W. Pedrycz, and X. Liu, "Time series long-term forecasting model based on information granules and fuzzy clustering," *Eng. Appl. Artif. Intell.*, vol. 41, pp. 17–24, May 2015.

[7] C.-H. Su and C.-H. Cheng, "A hybrid fuzzy time series model based on ANFIS and integrated nonlinear feature selection method for forecasting stock," *Neurocomputing*, vol. 205, pp. 264–273, 2016.

[8] J. S. R. Jang, "ANFIS: adaptive-network-based fuzzy inference system," *Syst. Man Cybern. IEEE Trans.*, vol. 23, no. 3, pp. 665–685, 1993.

[9] M. V. V. N. Sriram, N. K. Singh, and G. Rajaraman, "Neuro fuzzy modelling of Basic Oxygen Furnace and its comparison with Neural Network and GRNN models," in *2010 IEEE International Conference on Computational Intelligence and Computing Research*, 2010, pp. 1–8.

[10] H. A. Zamani, S. Rafiee-Taghanaki, M. Karimi, M. Arabloo, and A. Dadashi, "Implementing ANFIS for prediction of reservoir oil solution gas-oil ratio," *J. Nat. Gas Sci. Eng.*, vol. 25, pp. 325–334, 2015.

[11] Y. Ren, P. N. Suganthan, and N. Srikanth, "A Comparative Study of Empirical Mode Decomposition-Based Short-Term Wind Speed Forecasting Methods," *IEEE Trans. Sustain. Energy*, vol. 6, no. 1, pp. 236–244, Jan. 2015.

[12] R.-A. Hooshmand, H. Amooshahi, and M. Parastegari, "A hybrid intelligent algorithm based short-term load forecasting approach," *Int. J. Electr. Power Energy Syst.*, vol. 45, no. 1, pp. 313–324, 2013.

[13] L.-Y. Wei, "A hybrid ANFIS model based on empirical mode decomposition for stock time series forecasting," *Appl. Soft Comput.*, vol. 42, pp. 368–376, 2016.

[14] T. Xiong, Y. Bao, and Z. Hu, "Interval forecasting of electricity demand: A novel bivariate EMD-based support vector regression modeling framework," *Int. J. Electr. Power Energy Syst.*, vol. 63, pp. 353–362, Dec. 2014.

[15] J. J. Wang, W. Zhang, Y. Li, J. J. Wang, and Z. Dang, "Forecasting wind speed using empirical mode decomposition and Elman neural network," *Appl. Soft Comput.*, vol. 23, pp. 452–459, Oct. 2014.

[16] V. S. Kodogiannis, T. Pachidis, and E. Kontogianni, "An intelligent based decision support system for the detection of meat spoilage," *Eng. Appl. Artif. Intell.*, vol. 34, pp. 23–36, Sep. 2014.

[17] J. Zhu, Z. Ge, and Z. Song, "Distributed Parallel PCA for Modeling and Monitoring of Large-Scale Plant-Wide Processes With Big Data," *IEEE Trans. Ind. Informatics*, vol. 13, no. 4, pp. 1877–1885, Aug. 2017.

[18] K. Pearson, "LIII. On lines and planes of closest fit to systems of points in space," *Philos. Mag. Ser. 6*, vol. 2, no. 11, pp. 559–572, Nov. 1901.

[19] Chunhui Zhao and Youxian Sun, "Multispace Total Projection to Latent Structures and its Application to Online Process Monitoring," *IEEE Trans. Control Syst. Technol.*, vol. 22, no. 3, pp. 868–883, May 2014.

[20] T. Kohonen, "The self-organizing map," *Proc. IEEE*, vol. 78, no. 9, pp. 1464–1480, 1990.

[21] L. Weihua, Z. Shaohui, H. Guolin, W. Li, S. Zhang, and G. He, "Semisupervised Distance-Preserving Self-Organizing Map for Machine-Defect Detection and Classification," *Instrum. Meas. IEEE Trans.*, vol. 62, no. 5, pp. 869–879, 2013.

[22] O. Rentz, "Report on BAT in German Copper Production (Final Draft).," University Karlsruhe (DFIU)., 1999.

[23] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis : forecasting and control.* .

[24] M. Ceci, R. Corizzo, F. Fumarola, D. Malerba, and A. Rashkovska, "Predictive Modeling of PV Energy Production: How to Set Up the Learning Task for a Better Prediction?," *IEEE Trans. Ind. Informatics*, vol. 13, no. 3, pp. 956–966, Jun. 2017.

[25] M. Mitchell, *An introduction to genetic algorithms*. MIT Press, 1996.

[26] J. Han, M. Kamber, and J. (Computer scientist) Pei, *Data mining : concepts and techniques*. Elsevier/Morgan Kaufmann, 2012.

[27] C. Bergmeir and J. M. Benítez, "On the use of cross-validation for time series predictor evaluation," *Inf. Sci. (Ny).*, vol. 191, pp. 192–

213, May 2012.

[28] K. Kampouropoulos, J. J. Cardenas, F. Giacometto, and L. Romeral, "An energy prediction method using Adaptive Neuro-Fuzzy Inference System and Genetic Algorithms," in *Industrial Electronics (ISIE), 2013 IEEE International Symposium on*, 2013, pp. 1–6.

[29] Z. Hu, Y. Bao, T. Xiong, and R. Chiong, "Hybrid filter–wrapper feature selection for short-term load forecasting," *Eng. Appl. Artif. Intell.*, vol. 40, pp. 17–27, Apr. 2015.

[30] M. Khashei and M. Bijari, "Hybridization of the probabilistic neural networks with feed-forward neural networks for forecasting," *Eng. Appl. Artif. Intell.*, vol. 25, no. 6, pp. 1277–1288, 2012.

BIOGRAPHIES

Daniel Zurita (S'13) received his received the M.S. and Ph.D. degrees in Electronics Engineering and the Ph.D. degree from the UPC, Barcelona, Spain in 2013 and 2017 respectively. In 2013, he joined the MCIA Center, where he is currently a researcher. His research interests include fault diagnosis and prognosis in electric machines, industrial process monitoring, fault detection algorithms, machine learning and signal processing methods.

Miguel Delgado (M'03) received the M.S. and Ph.D. degrees in Electronics Engineering and the Ph.D. degree from the UPC, Barcelona, Spain in 2007 and 2012 respectively. From 2004 to 2008 he was a Teaching Assistant in the Electronic Engineering Department of the UPC. In 2008 he joined the MCIA Center, where he is currently a research assistant. His research interests include fault detection algorithms, machine learning, signal processing methods and embedded systems.

Jesus A. Carino (S'13) received the M.S. degree in electrical engineering from the University of Guanajuato, Guanajuato, México, in 2012. Currently, he is working toward the Ph.D. degree under CONACyT scholarship program at the MCIA Center in UPC, Terrassa, Spain. His research interests include digital signal processing on FPGAs for applications in mechatronics, fault diagnosis in electric machines, fault detection algorithms and pattern recognition.

Juan A. Ortega (M'94) received the M.S. Telecommunication Engineer and Ph.D. degrees in Electronics from the Technical University of Catalonia (UPC) in 1994 and 1997, respectively. In 1994, he joined the UPC Department of Electronic Engineering. Since 2001 he belongs to the Motion Control and Industrial Applications research group. His research activities include: motor diagnosis, signal acquisition, smart sensors, embedded systems and remote labs.