

© 2016 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Industrial Process Monitoring by means of Recurrent Neural Networks and Self Organizing Maps

Daniel Zurita, Enric Sala, Jesús A. Carino, Miguel Delgado, Juan A. Ortega

MCIA Research Center, Department of Electronic Engineering

Technical University of Catalonia (UPC)

Rbla. San Nebridi nº22, Gaia Research Building, 08222 Terrassa, Spain

{daniel.zurita, enric.sala, jesus.carino, miguel.delgado, juan.antonio.ortega}@mcia.upc.edu

Abstract—Industrial manufacturing plants often suffer from reliability problems during their day-to-day operations which have the potential for causing a great impact on the effectiveness and performance of the overall process and the sub-processes involved. Time-series forecasting of critical industrial signals presents itself as a way to reduce this impact by extracting knowledge regarding the internal dynamics of the process and advice any process deviations before it affects the productive process. In this paper, a novel industrial condition monitoring approach based on the combination of Self Organizing Maps for operating point codification and Recurrent Neural Networks for critical signal modeling is proposed. The combination of both methods presents a strong synergy, the information of the operating condition given by the interpretation of the maps helps the model to improve generalization, one of the drawbacks of recurrent networks, while assuring high accuracy and precision rates. Finally, the complete methodology, in terms of performance and effectiveness is validated experimentally with real data from a copper rod industrial plant.

Keywords— *Artificial intelligence, Condition monitoring, Feature extraction, Machine learning, Recurrent neural networks, Self-Organizing maps.*

I. INTRODUCTION

Ensuring reliability and robustness has long been one of the main concerns in industrial manufacturing because of the ever-increasing demand for improvements over the cost and quality of the processes' outcome product. Due to these performance requirements, necessary in order to keep up with competition, the current economic scheme is incentivizing a response from the manufacturing enterprises to said conditions, such as the adoption of intelligent monitoring strategies [1]. These new approaches are aimed to improve the control of the plants by means of extracting actionable knowledge from the internal mechanics of the machines [2]. A deviation from the nominal working parameters implies a divergence from the optimal production specification and a misalignment from the best product quality point, effectively causing a critical loss of potential earnings [3]. Thus, the manufacturing industry as a whole, and more so in Europe, is striving towards the Zero Defects Manufacturing (ZDM) paradigm [4].

In this regard, the current main approaches are using fault diagnosis methodologies [5]. However, in order to prompt a step forward towards the next generation of factories, the exploitation of forecasted information is becoming a critical need, because it provides the sufficient insight into the process to be able to

understand the effects of deviations a step ahead [6, 7]. In particular, the modeling of industrial processes aimed at forecasting of future conditions is considered as one of the key factors in ZDM paradigm, which is the foundation of the Industry 4.0 [8]. Accordingly, one of the current challenges of industrial condition monitoring is finding useful prediction schemes capable of rendering the future condition of the process or machines under monitoring, i.e. prognosis approaches [9, 10], which are the objective of industrial forecasting methods.

Nevertheless, the topic of modeling in the industrial sector has been widely studied in the literature because of its potential. Nevertheless, due to the challenging complexity of manufacturing plants, combined with the sophistication of data-driven approaches, it is mostly studied at the machine level [11]. This means that these monitoring solutions are focused and tuned to keep track of the operating point of individual machines, which may lead to reaching a local minima on the overall process. Meanwhile, the development of a monitoring methodology that instead focuses on the whole process could further improve the accuracy of the monitoring system by being able to distill interactions among machines and also discovering operation patterns not perceptible at a machine level [12].

On the one hand, during recent years the methods employed to achieve these tasks have underwent changes from the early parametric models, in fact, the trend has been shifting towards data-driven methodologies that take advantage of the prevalence of a wealth of data and sensors in factories. The increasing adoption of Big Data analysis [13] and automatic control schemes [14], among others, are developments from this evolution. Some of the most widely used methods of forecasting are Neural Networks and Adaptive Neuro-Fuzzy Inference Systems because of their ability to adapt to complex signals, discover periodic or repeating patterns and their relative ease of configuration. However, manufacturing processes often include non-periodic, non-linear components, in addition to the added difficulty of establishing the dynamics of the signal. This is why a modeling method with built-in dynamics support, such as Recursive Neural Networks (RNN), could overcome this difficulties and facilitate the process while improving its accuracy [15].

On the other hand, the classical feature selection stage in order to identify the best set of auxiliary signals represents an overfitting approach usually neglected. The cost function related with feature selection techniques is commonly analyzed in terms

of forecasting performance, thus, the generalization capabilities of the original set of data is lost. Concerning with this problem manifold learning approaches has been applied in the last years to preserve as much as possible the original information in a lower dimensional space [16]. Among them, Self-Organizing Maps (SOM) is the most used, which is based on developing a neural network grid to preserve most of the data topology defined by the original feature vectors. Instead of using the classical SOM approaches, in which the lower dimensional output vector is used, in this work, the SOM is applied to codify the most significant combination of signal values to get the sub-process operating condition.

The methodology presented in this paper shows that the application of a high level modeling implementation based on the data collected from the industrial database, allows to forecast the critical signals evolution of an industrial process, hence providing the system's response up to the forecast horizon selected. The methodology takes advantage of the capability of SOM for compressing the information while maintaining the topographical properties of the data. SOM is used to extract the operating conditions from a set of auxiliary signals that conforms a sub-process. This information calculated for the different sub-processes that affect the critical signal is introduced into a RNN model. This information helps the model to improve generalization in order to overcome one of the limitations of RNN, while assuring high ratios of accuracy and precision characteristic of such modeling approaches.

This paper is organized as follows. Section II presents the theoretical aspects of SOM for information compression, and RNN for time series modeling. Section III presents the architecture of the proposed method. In Section IV the experimental results and the method capabilities are verified by means of real industrial data. Finally, the conclusions are described in Section V.

II. THEORETICAL APPROACH

A. Self-Organizing Maps

The Self-Organizing Map (SOM) learning rule, also called as Kohonen map, was proposed by Kohonen [17] in 1990 to build topology preserving mappings. The main important aspect is that SOM is based on neural networks. In this sense, SOM corresponds to a neural network grid trying to preserve the topological properties of the input space. The output space (the mapped space or latent space), is predefined as a regular d-dimensional grid. Generally, a two or three-dimensional grid is predefined and suitable for most of applications [18, 19].

Each point of the SOM grid represents a neuron or a Matching Unit (MU). For each neuron, MU_s , a D-dimensional weight vector w_{v-j} is defined for the v -th data of the j -th neuron. The weights represent the coordinates of neurons position in the input space. Mapping is performed by assigning each data point x_{v-j} in the input space to one of these neurons, namely the one whose weight vector is closest to the point, which is called the Best Matching Unit (BMU). The position vector y_{v-j} in the output space is then given by the grid position of this neuron. The used error function corresponds as follows:

$$E_{SOM} = \sum_j \sum_{i \in s_{y-i}} (w_{v-j} - y_{v-i})^2 \quad (1)$$

with s_{y-i} being the set of data points which have neuron i as closest neuron. E_{SOM} expresses the average squared distance from a point to its representative. The minimization of E_{SOM} represents the objective of the operation, and is performed with respect to the weight vectors w_{v-j} . The classical gradient descent approach leads to the following updating rule:

$$w_{v-j}^{(t+1)} = w_{v-j}^{(t)} - \alpha^{(t)} (\nabla E_{SOM}^{(t)})_{v-j} \quad (2)$$

However, this algorithm clusters the input space, but is not useful for mapping since the learning rate, α does not depend on the output space. For this reason the learning rate is replaced by a neighborhood function Nhf_{wn} which explicitly depends on the mapped space:

$$Nhf_{wn}^{(t)} = \begin{cases} \alpha^{(t)} & \text{if } i \in N_{wn}^{(t)} \\ 0 & \text{if } i \notin N_{wn}^{(t)} \end{cases} \quad (3)$$

where $N_{wn}^{(t)}$ is the set of all neurons within a certain range of the BMU (the nearest element to the presented data point) in the output space. During training, this range and $\alpha(t)$ are decreased monotonically. Since neighboring neurons in the output space will be neighbors in the input space, the mapping preserves a certain range of local topology.

Typically, the performance of SOM map is evaluated by average quantization error and topographic error in terms of the training data. The average quantization error means the average distance from each data vector to its BMU, and the topographic error means the percentage of data vectors for which the BMU and the second-BMU are not adjacent units.

B. Recurrent neural networks

Artificial Neural Network (ANN) are data-driven self-adaptive information processing method inspired in biological systems, and represents the most commonly found data-driven technique in the prognosis literature [20,21]. It is composed by a number of interconnected processing elements (neurons) working at the same time to solve a specific problem. The systems based in NN for pattern recognition are configured for a specific application, such as diagnosis, through a learning process. NN represents a non-linear, multivariate and non-parametric algorithm [22].

ANN are constructed by several neurons and these neurons are structured in different layers. The basic processing unit, the neuron, produces an output by forming a weighted sum of a number of inputs, x_v . The basic operation is that, if the summed value is higher than a threshold, S_{ne-j} , then the neuron generates an output pulse, y_j , which is sent to neighboring neurons and propagated around the different layers of the net. The characteristic behavior of the neuron can be described as shown in equation (4) and (5), for an ANN with an input vector $x_v = \{e_1, e_2, \dots, e_D\}$, where w_{ij} is the i -th weight of the j -th neuron.

$$S_{ne-j} = \sum_{i=1}^D w_{ji} \cdot f_{e-i} \quad (4)$$

$$y_j = f(S_{ne-j}) \quad (5)$$

However, ANN present a strong limitation when dealing with time series forecasting problem. ANN assume that exists an independence between training data and test data. After the evaluation of one sample, the entire state of the network is lost. This has no consequences if the data is independent like in classification approaches, but is a critical drawback in time series modeling since there is a strong relation between time and the data introduced.

To overcome this problem, Recurrent Neural Networks (RNN) are connectionist models with the ability to capture the dynamics of data in the networks nodes [23]. The classic layer recurrent network model is shown in Fig. 1, in which the neurons of the hidden layer present a feedback loop in which prior states of the network are considered in the evaluation of a new sample. In this regard, this feedback loop provides the network the capability to retain information from previous network states, that is, to have memory and relate the output of the network to the input sequence introduced. This characteristic made RNN a powerful modeling method for time series modeling since it maintains all the main aspects of ANN, and includes the ability to capture time dynamics in complex signals. The initial research in RNN was made by Hopfield et al. in 1982 [24] with the design of Hopfield networks that have applications in pattern recognition.

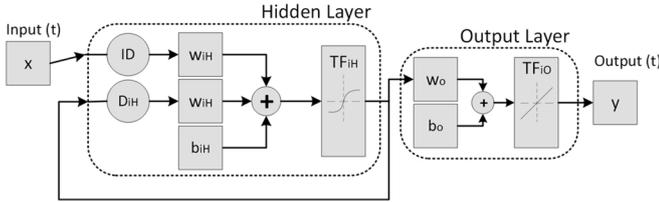


Fig. 1 Main diagram of a Recurrent Neural Network, in which the hidden layer presents a feedback loop over the output of the same layer.

This diagram present implications over the classic ANN approach. Although, the basic functioning of the neurons of the network remain inalterable, the concept of layer delay, D_{IH} , appears, that is directly related with the ability of the network to retain past states of itself. Note that additional input delays ID can be specified in this architecture. The RNN is trained by means of the backpropagation through time algorithm (BPTT).

III. ARCHITECTURE OF THE METHOD

The proposed method deals with the monitoring of industrial processes in which the critical signal that wants to be modeled presents a complex behavior, that means presenting a non-periodic response and non-linear relations among time. Furthermore, it is considered that the target signal is affected in some measure by other surrounding sub-processes of the industrial plant.

The method takes advantage of the modeling capabilities of RNN for non-linear time series forecasting in order to improve the accuracy in the modeling. However, the main problem of

such complex modeling methods is the risk of overfitting to the training signal. In order to avoid this problem and increase generalization, the operating point of the surrounding sub-processes, represented by the auxiliary signals presented in the industrial database, is codified by means of a dedicated SOM. Therefore, during the online procedure the corresponding BMU of each sub-process is introduced as an auxiliary input to the RNN model. It should be noticed that the evaluated BMU corresponds to a temporal signal that summarizes the operating point of the process, and not to a static measure. This means that that the BMU evolves with time and as a result, it presents a characteristic dynamic related to the process behavior. Fig. 2 illustrates the procedure of the proposed modeling method.

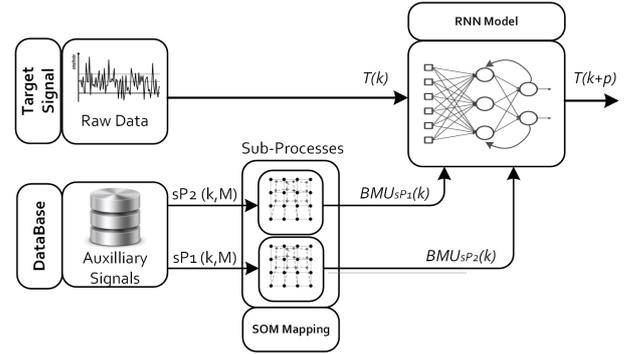


Fig. 2 Block diagram of the training procedure from the proposed method when modeling the target signal $T(t)$ to obtain its future behavior, $T(t+p)$, at a certain time horizon p . Note that for this example, two sub-processes that affect $T(t)$ are available and mapped with a dedicated SOM. The BMUs of each SOM are introduced as inputs to the RNN model.

A. Process condition codification

Due to the characteristics of industrial processes, there are auxiliary signals that are highly correlated with the target signal. Those signals have a great potential to be included in the modelling in order to increase adaptability and generalization. In this work, a SOM based topology preservation approach is proposed, in which the data is projected to a lower dimension seeking that the samples maintain the topographic distribution from the higher dimension space. Thus, the SOM mapping codification is a two-step procedure. The first phase deals with the compression of the D -dimensional data space (usually $D > 2$), into a d -dimensional grid formed by the SOM's neurons. Usually, for visualization purposes, $d \leq 2$. This step is achieved after the initialization and the training of the network with a training set. The training is the procedure to adapt the input SOM grid to the data space in order to maintain as much as possible the data variance, the information, by an adaptive analysis of the data distribution in the input space. The second step compresses the grid information into the BMU. After the training, the SOM network can be evaluated regarding an input in order to detect which is the associated BMU to this input. Therefore, each BMU represents the mapping of a non-uniform region of the data space.

For the method, a dedicated SOM will be generated for each sub-process identified that affects the target signal. Then, SOM model is fed with a selection of the available auxiliary signals that are correlated with the target signal and the BMU is introduced as input of the model.

B. Forecasting by means of Recurrent Neural Networks

A single RNN model is considered for the method in order to forecast the target signal. And therefore, it should be designed and trained. The RNN model follows the equation given in Eq. 6. For a target signal, $T(t)$, and two sub-processes $sP1$, and $sP2$, the model presents a single output, which is the forecasted signal at a time horizon p , $T(t+p)$. The considered inputs of the model are: (i) The current value of the signal to be forecasted, $T(t)$. (ii) The first sub-process operating point by means of the BMU of the SOM grid for the current value of the signals, $BMU_{sP1}(t)$. (iii) The BMU of the second SOM grid for $sP2$, $BMU_{sP2}(t)$.

$$T(t+p) = f(T(t), BMU_{sP1}(t), BMU_{sP2}(t)) \quad (6)$$

Note that since the recurrent architecture presents a certain memory from pasts states of the net, there is not needed to introduce past inputs of $T(t)$ as inputs of the models, information regarding temporal evolution is stored inside the network. However, the number of delays in the feedback loop of the hidden layers, D_{IH} , is a direct measure of the inherent memory capability of the net and must be configured in regard of the application. Hence, a D_{IH} between 2 and 5 samples is recommended for industrial signals in which do their nature, long term past values does not affect the output.

C. Performance Evaluation

In order to evaluate the performance and the accuracy of a model, statistical metrics have been used, which are: RMSE, MAE and MAPE, which are defined below [25]. The RMSE is one of the most used performance metrics for the development of forecasting models, is a measure of the standard deviation of the differences between predicted values and observed values. It is useful in order to analyse the global behaviour of the model, but is very sensitive with the amplitude. In this regard the MAE error is used to evaluate the forecast since it is less sensitive to outliers. Finally, the MAPE error helps to determine the mean deviation of each sample normalized by the amplitude, so it helps to unify the scale and can be used to compare the errors of signals with different levels of amplitudes, such as the IMFs.

$$RMSE = \sqrt{\frac{\sum_{t=1}^L (y(t) - \hat{y}(t))^2}{L}} \quad (7)$$

$$MAE = \sqrt{\frac{\sum_{t=1}^L |y(t) - \hat{y}(t)|}{L}} \quad (8)$$

$$MAPE = \frac{\sum_{t=1}^L \left| \frac{y(t) - \hat{y}(t)}{y(t)} \right|}{L} \cdot 100\% \quad (9)$$

IV. EXPERIMENTAL VALIDATION

The method is validated using industrial data collected from the Spanish metallurgy company, La Farga, specifically from a high purity copper rod manufacturing plant. A deviation in a primary signal, the casting wheel refrigeration index, $R_{ind}(k)$, affects the proper functioning of the process and, therefore, the final manufactured product's quality. Such industrial system represents a complicated scenario for process forecasting due to the non-stationary operating conditions and the non-linear relation among signals.

A. Experimental plant definition

The production process, represented in Fig. 3, proceed as follows: first, high purity copper cathodes are melted by natural gas fired burners arranged in rows around the shaft furnace. Second, the copper flows from the shaft furnace via a gas-fired launder to a second furnace, the holding furnace, which acts as a lung. The holding furnace, which is also fired with natural gas, serves as a buffer to provide a constant flow to the rest of the process and, if required, can be used to increase the temperature. Third, the molten copper flows from the holding furnace via another gas-fired launder to a tundish with a ceramic valve, which feeds the casting wheel. A water cooled steel band encloses half of the casting wheel, forming the casting cavity in which the molten copper solidifies to form a raw rod by means of a heat extraction process. In this regard, acetylene, burnt with air, produces a soot dressing for the casting wheel and steel band facilitating heat transfer between the cooper and the steel band. Fourth, after being beveled and shaved, the cast bar is moved to a rolling mill consisting of a roughing section and one finishing section, which reduces the bar to its final diameter. Finally, the copper bar is cooled to proceed with the coiling and packaging, where the copper rod is strapped and covered with a polyethylene film, giving with it the final manufactured product, the copper rod.

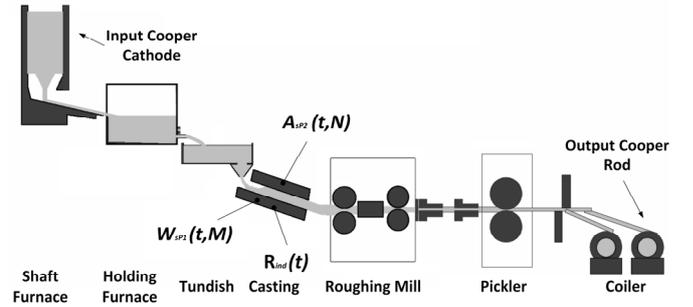


Fig. 3. Diagram of the main elements of the copper rod manufacturing plant [26]. Target signal: $R_{ind}(t)$ – Refrigeration index of the casting wheel measured in °C. Sub-process that affect the target signal: $W_{sP1}(t,M)$ – Water refrigeration process of the casting wheel. $M=9$ signals in $sP1$. $A_{sP2}(t,N)$ – Acetylene panting process of the casting wheel. A total amount of $N=9$ signals are available.

The objective of this application is to forecast the refrigeration index of the casting wheel, $R_{ind}(t)$. This index is an indirect measure of the effectiveness of the refrigeration among time. This magnitude is critical in the manufacturing process because represents the temporal heat extraction from the melted cooper during the casting procedure. Deviations in the refrigeration imply imperfections in the manufactured copper rod due to non-uniformities in the copper density. The forecasting of such index allows the corresponding actions to avoid the affectation to the next manufacturing batch.

Available information in regard with the heat extraction process can be divided in two sub-processes, the water refrigeration sub-process, W_{sP1} , with 9 auxiliary signals and the acetylene painting sub-process, A_{sP2} , with other 9 auxiliary signals. In one hand, the matrix of auxiliary signals from W_{sP1} defined as $W_{sP1}(t, 1:M)$, include the following auxiliary signals: the temperature of the casting wheel [$W_{sP1}(t, 1)$], measures of flow and pressure in in the interior of the wheel $W_{sP1}(t, 2:3)$,

same measures in the exterior [$W_{SP1}(t, 4:5)$], lateral [$W_{SP1}(t, 6:7)$] and center of the wheel [$W_{SP1}(t, 8:9)$].

On the other hand, regarding to the signals from the acetylene process, the matrix $A_{SP2}(t, 1:N)$ includes the temperature of the steel band $A_{SP2}(t, 1)$, and other signals regarding the pressures and flows from the acetylene painting elements measured in different positions the steel band, which are the exterior [$A_{SP2}(t, 2:3)$], interior [$A_{SP2}(t, 4:5)$], central [$A_{SP2}(t, 6:7)$] and lateral positions [$A_{SP2}(t, 8:9)$].

All signals are acquired synchronously, and are automatically stored in a standard SQL database at a period of 10 seconds, that corresponds to a sampling frequency, f_s , of 0.1 Hz. Two data sets corresponding to 50 hours of plant operation are used in this work, the first set, shown in Fig. 4.a, is used for training purposes, while the second test Fig. 4.b is used for testing the suitability and the generalization capabilities of the method.

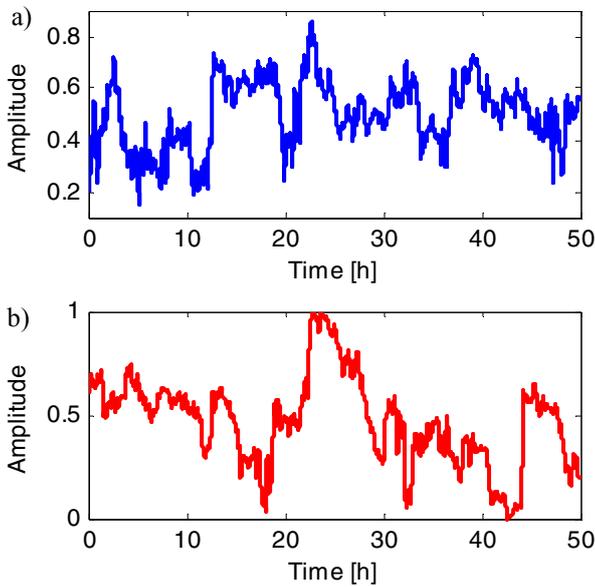


Fig. 4. Acquired refrigeration index, $R_{ind}(t)$, from the manufacturing plant in two periods of 50 hours of operation. a) Corresponds to the training set, while b) to the validation set. Notice that data has been normalized in the [0-1] interval.

B. Correlation based signal selection

Since all signals available have been selected, it is necessary to perform a signal selection approach in order to remove redundant or non-correlated information from the analysis. To do so, the correlation coefficient, $C_{coef}(n)$, is proposed. The C_{coef} is extracted from Eq. (10) as the ratio between the covariance matrix of the n -th signal of the sub-process variables $Aux_{in} = \{W_{SP1}(t, 1:9), A_{SP2}(t, 1:9)\}$ and the target signal $R_{ind}(t)$ and their individual dispersion σ .

$$C_{coef}(n) = \frac{cov(Aux_{in}(t, n), T(t))}{\sigma(Aux_{in}(t, n)) \cdot \sigma(T(t))} \quad (10)$$

If an auxiliary signal from any sub-process achieves a C_{coef} lower than a predefined threshold of significant correlation, fixed at 10%, the signal is discarded from the analysis. Table 1 shows the correlation coefficients for all signals of both subprocesses. Note that after the analysis the following signals

are discarded: (i) From the water cooling sub-process, [$W_{SP1}(t, 2)$, $W_{SP1}(t, 6)$, $W_{SP1}(t, 7)$], and from the acetylene sub-process, [$W_{SP1}(t, 5)$, $W_{SP1}(t, 7)$, $W_{SP1}(t, 9)$] are discarded, that gives 6 useful signals for each sub-process.

Table 1. Correlation achieved for each signal analyzed for both subprocesses in (%).

W_{SP1}	1	2	3	4	5	6	7	8	9
C_{coef}	37.5	1.1	15.2	20	39.4	0.9	2.25	10.3	22.8
A_{SP2}	1	2	3	4	5	6	7	8	9
C_{coef}	42.2	19.8	45.7	32.6	4.31	28.5	1.18	39.6	1.29

C. Process condition codification

After useful information has been selected, the resulting reduced auxiliary signal inputs from each sub-process will be introduced in a SOM for mapping the operating point of each sub-process. This dedicated SOM will adapt the internal grid to the topology of this input data space. Therefore, the SOM grid has been configured with a 2D hexagonal connection grid with a size of [15x15], which fixes a total amount of 225 neurons to cover the whole space. Prior to the training, the input signals have been normalized in order to have mean zero and standard deviation equal to 1. Then, the SOM is initialized and trained with a batch algorithm and a total amount of 100 epoch were performed.

Typically, the performance of SOM map is evaluated by average quantization error and topographic error in terms of the training data. For this training set, the average quantization error is equal 1.081 to and the topographic error is 0.045 for W_{SP1} , and 0.977 for quantization error and 0.086 of topographic error for A_{SP2} .

The obtained BMU for both test set is shown in Fig. 6. Theoretically, any process variation should be reflected in a certain region of the input data space covered by the SOM neurons.

The resulting SOM mapping can be seen by means of the U-Matrix in Fig. 5.a) for W_{SP1} and Fig. 6. A) for A_{SP2} . The U-Matrix represents the distances between the neurons of the trained SOM. The dark areas represent regions of the map in which data is concentrated (low distance between neurons), and the light areas represent the frontiers of the clusters (high distances between neurons). In the figure, the dark area in the center correspond to the normal operation point of the process. Indeed, data in the light area correspond to those points that are far from the center and are identified as punctual deviations from the nominal conditions.

It can be appreciated in Fig. 5. A) how the water cooling process present a more disperse behavior since there are not high density areas presented among the data, however, it can be appreciated a consistent operating point in the middle of the U-matrix. This fact is also reflected in the BMU achieved in the number of jumps over different number of neurons. In the opposite, the acetylene painting process presents a more stable behavior since 5 big operating points can be identified in the u-matrix of Fig. 6. a), this results in a more stable behavior of the obtained BMU as can be seen in Fig. 6. b).

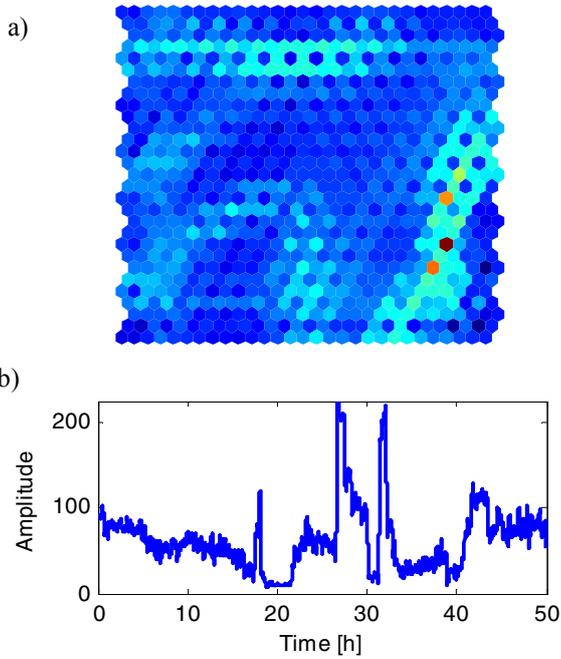


Fig.5. Results from the SOM of the training set applied to the water cooling sub-process W_{sp1} . A) U-matrix representing the inter-distances between neurons, b) BMU extracted with the temporal signal, $BMU_{Wsp1}(t)$.

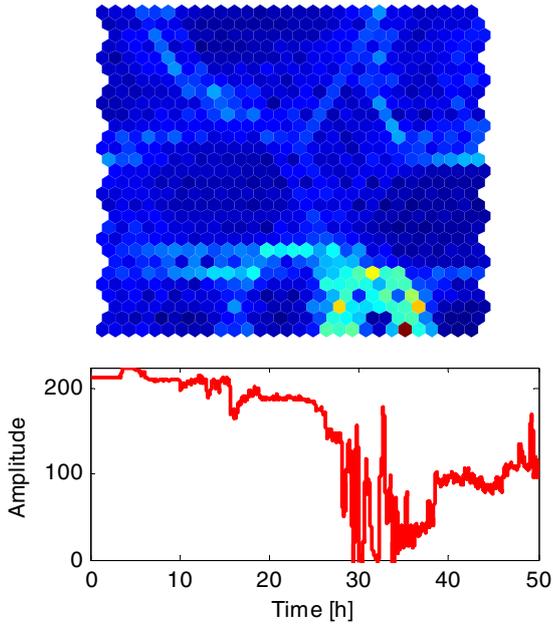


Fig.6. Results from the SOM of the training set applied to acetylene painting sub-process A_{sp2} . a) U-matrix representing the inter-distances between neurons, b) BMU extracted with the temporal signal, $BMU_{Asp2}(t)$.

D. Recurrent neural network based modeling

Once the information regarding the operating point of each sub-process is summarized into the BMU, it is time to design and train the recurrent layer neural network. Prior to the design, the prediction horizon p of the models should be selected. For this concrete experiment, the horizon has been configured to detect deviations in the actual cooper rod under production, before it affects the next manufactured unit. Considering the

manufacturing time of one cooper rod, the prediction horizon is configured to be 900 seconds ($p=90$ samples).

The RNN model inputs been configured according to the structure defined in section III. For this application, the total amount of inputs to the model is three, and consists on: (i) the current value of the refrigeration index, $R_{ind}(t)$; (ii) the best matching unit of the SOM mapping after the training procedure for the water cooling sub-process, $BMU_{Wsp1}(t)$, and (iii) the BMU for the acetylene painting sub-process $BMU_{Asp2}(t)$. The output of the model is the refrigeration index at the time horizon p , $R_{ind}(t+p)$. Note that each input is normalized with the min-max method in order to have a range from 0 to 1.

For this application, a multi-layer NN has been configured with one hidden layer which is composed of 10 recurrent hidden neurons. The neurons have been configured with a sigmoid activation function, which is recommended for smoothing the response in time series forecasting. The network is trained by means of backpropagation trough time algorithm (BPTT). Once the network has been defined and configured, it is trained with the training set during 100 epoch, achieving a performance in training of in terms of MSE of 0.0042.

In this regard, the forecasting performance during the training can be seen in Fig. 7, and the forecasting performance against the test set in Fig. 8. It can be appreciated how the model is able to capture the dynamics of the signal with almost no error. Furthermore, the generalization capabilities of the model are demonstrated by the minimal loss of performance while modeling the validation set. Table 2 show the error metrics of both datasets. It should be appreciated the low achieved RMSE, that means that the model exhibits a robust response with no overvalues.

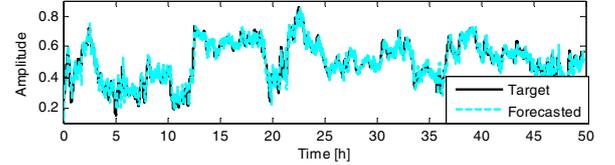


Fig. 7 Training result of the RNN model with prediction horizon, $p=90$.

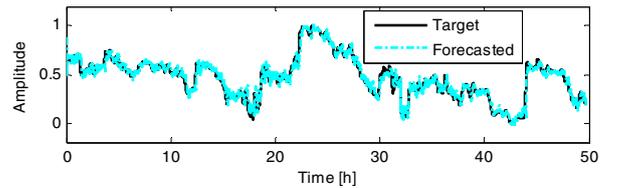


Fig. 8 Validation result of the RNN model with prediction horizon, $p=90$.

Table 2. Performance evaluation of the RNN model with the training and the validation sets.

	Training	Validation
RMSE	0.036	0.057
MAE	0.167	0.431
MAPE	6.17%	13.51%

E. Comparison with other methods

In order to contrast the improvement of the proposed method, the results are compared with the G-ANFIS method, which is a state of the art modeling methodology used to forecast industrial signals [27]. The particularity of this method is that it uses a GA based input selection method in regard with the signal, past inputs and the available auxiliary signals to select those signals that reduce the error function of the model.

In this study the GA has been configured to select the best inputs from $[R_{ind}(t-n_1), W_{sP1}(t, 1:9), \text{ and } A_{sP2}(t, 1:9)]$, note that the current value $R_{ind}(t)$ 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, for the rest of signals, binary inputs are used in order to incorporate (1) 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).

$$R_{ind}(t+p) = G_ANFIS(R_{ind}(t), R_{ind}(t-8), W_{sP1}(t, 1:3), A_{sP2}(t, 6)) \quad (11)$$

Table 3. Performance evaluation of the proposed method versus the G-ANFIS model with the training (*Trn.*) and the validation (*Val.*) data sets.

Data	SOM+RNN		G-ANFIS	
	Trn.	Val.	Trn.	Val.
RMSE	0.036	0.057	0.098	0.305
MAE	0.167	0.431	0.328	0.669
MAPE	6.17%	13.51%	12.28%	45.51%

Table 3 shows the results of the comparison in terms of performance metrics. In can be seen in terms of RMSE, the proposed method outperforms the classic model. It is also important the MAPE value, since the proposed method for both training and validation sets the MAPE is similar (6-13%), for the classical ANFIS the MAPE is highly increased (from 12% up to 45%). The results indicate the suitability of the proposed method in regard with a classical ANFIS model with a GA based input selection. Furthermore, as the complexity of the model increases with the number of inputs, the SOM compression improves the generalization and in regard with using single isolated inputs.

V. CONCLUSIONS

In this work, a method for time series forecasting has been presented and validated by means of industrial data from a copper manufacturing plant. The method combines the adaptability and performance of the Recurrent Neural Networks with the information compression of Self-Organizing Maps in order to achieve better performance and generalization in critical industrial signal modeling.

It has been seen how the SOM is able to capture each sub-process behavior and compress this information in the BMU indicator. Furthermore, it has been seen how each sub-process may present a certain dynamic and might be more or less

correlated with the target signal to be forecasted. Furthermore, it has been seen from the cooper manufacturing point of view, that the water cooling procedure presents a less stable behavior due to the variations in the cooper temperature and the response of the control loops of the factory.

The results shows that RNN are suitable methods to be applied in industrial process modeling. This means that the introduction of the recurrent loop improves the performance of the modelling versus time series in which input data to the neural network depends of the time value. It should be noticed that the experimental validation has been oriented to test suitability of the method in industrial processes, therefore, no further signal pre-processing steps has been introduced.

It should be noted that this methodology can be applied to monitor and forecast time series of interest in different industrial applications in real time due to the fact that the BMU corresponds to a continuous signal that can be calculated at every step. Future work can be concentrated on implementing more complex recurrent structures such as the nonlinear autoregressive network with exogenous inputs in which not only the hidden layer presents a recurrent loop, but the output of the net is introduced as a recursive input. This may lead to study the stability of the net to assure the proper functioning.

VI. ACKNOWLEDGEMENT

The authors would like to thank the support and the access to the industrial database provided by La Farga under the project MONIPRO IDI-20140898, founded by MINECO-EEA GRANTS framework. This work was also supported in part by the Spanish Ministry of Education, Culture, and Sport under the grant FPU13/00589.

REFERENCES

- [1] S. Yin and O. Kaynak, "Big Data for Modern Industry: Challenges and Trends [Point of View]," *Proceedings of the IEEE*, vol. 103, pp. 143-146, 2015.
- [2] R. K. Singleton, E. G. Strangas, and S. Aviyente, "Extended Kalman Filtering for Remaining-Useful-Life Estimation of Bearings," *IEEE Transactions on Industrial Electronics*, vol. 62, pp. 1781-1790, 2015.
- [3] D. Zurita, J. A. Carino, E. Sala, M. Delgado-Prieto, and J. A. Ortega, "Time series forecasting by means of SOM aided Fuzzy Inference Systems," in *Industrial Technology (ICIT), 2015 IEEE International Conference on*, 2015, pp. 1772-1778.
- [4] O. Myklebust, "Zero Defect Manufacturing: A Product and Plant Oriented Lifecycle Approach," *Procedia CIRP*, vol. 12, pp. 246-251, // 2013.
- [5] M. D. Prieto, G. Cirrincione, A. G. Espinosa, J. A. Ortega, and H. Henao, "Bearing Fault Detection by a Novel Condition-Monitoring Scheme Based on Statistical-Time Features and Neural Networks," *IEEE Transactions on Industrial Electronics*, vol. 60, pp. 3398-3407, 2013.
- [6] S. Yin, X. Li, H. Gao, and O. Kaynak, "Data-Based Techniques Focused on Modern Industry: An Overview," *IEEE Transactions on Industrial Electronics*, vol. 62, pp. 657-667, 2015.
- [7] P. Leit, x00E, V. Ma, and P. Vrba, "Past, Present, and Future of Industrial Agent Applications," *IEEE Transactions on Industrial Informatics*, vol. 9, pp. 2360-2372, 2013.
- [8] 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.
- [9] M. Delgado, G. Cirrincione, A. G. Espinosa, J. A. Ortega, and H. Henao, "Dedicated hierarchy of neural networks applied to bearings degradation assessment," in *Diagnostics for Electric Machines, Power Electronics and Drives (SDEMPED), 2013 9th IEEE International Symposium on*, 2013, pp. 544-551.

- [10] L. Yi-Sheng, C. Yung-Chi, T. Shiao-Li, and T. Tzung-Cheng, "A novel search scheme for nonintrusive load monitoring systems," in *Industrial Technology (ICIT), 2012 IEEE International Conference on*, 2012, pp. 102-107.
- [11] C. Chaochao, Z. Bin, G. Vachtsevanos, and M. Orchard, "Machine Condition Prediction Based on Adaptive Neuro-Fuzzy and High-Order Particle Filtering," *Ind. Electron. IEEE Trans.*, vol. 58, no. 9, pp. 4353-4364, 2011.
- [12] J. A. Carino, D. Zurita, M. Delgado, J. A. Ortega, and R. J. Romero-Troncoso, "Hierarchical classification scheme based on identification, isolation and analysis of conflictive regions," in *Proceedings of the 2014 IEEE Emerging Technology and Factory Automation (ETFA)*, 2014, pp. 1-8.
- [13] H. Yang, M. Park, M. Cho, M. Song, and S. Kim, "A system architecture for manufacturing process analysis based on big data and process mining techniques," in *Big Data (Big Data), 2014 IEEE International Conference on*, 2014, pp. 1024-1029.
- [14] R. Zhang, A. Xue, and F. Gao, "Temperature Control of Industrial Coke Furnace Using Novel State Space Model Predictive Control," *IEEE Transactions on Industrial Informatics*, vol. 10, pp. 2084-2092, 2014.
- [15] P. Przystalka and W. Moczulski, "Methodology of neural modelling in fault detection with the use of chaos engineering," *Eng. Appl. Artif. Intell.*, vol. 41, pp. 25-40, May 2015.
- [16] C. Brighenti and M. A. Sanz-Bobi, "Auto-regressive processes explained by self-organized maps. Application to the detection of abnormal behavior in industrial processes," *IEEE Trans. Neural Netw.*, vol. 22, no. 12, pp. 2078-90, Dec. 2011.
- [17] T. Kohonen, "The self-organizing map," *Proc. IEEE*, vol. 78, no. 9, pp. 1464-1480, 1990.
- [18] R.-J. Lian, "Adaptive Self-Organizing Fuzzy Sliding-Mode Radial Basis-Function Neural-Network Controller for Robotic Systems," *IEEE Trans. Ind. Electron.*, vol. 61, no. 3, pp. 1493-1503, Mar. 2014.
- [19] J. Vesanto, J. Himberg, E. Alhoniemi, and J. Parhankangas, *SOM toolbox for Matlab 5*: Citeseer, 2000.
- [20] F. S. Wong, "Time series forecasting using backpropagation neural networks," *Neurocomputing*, vol. 2, no. 4, pp. 147-159, 1991.
- [21] F. Rossi, D. Velázquez, I. Monedero, and F. Biscarri, "Artificial neural networks and physical modeling for determination of baseline consumption of CHP plants," *Expert Syst. Appl.*, vol. 41, no. 10, pp. 4658-4669, 2014.
- [22] D. Xuewu and G. Zhiwei, "From Model, Signal to Knowledge: A Data-Driven Perspective of Fault Detection and Diagnosis," *Ind. Informatics, IEEE Trans.*, vol. 9, no. 4, pp. 2226-2238, 2013.
- [23] Z. C. Lipton, "A critical review of recurrent neural networks for sequence learning," arXiv Prepr. arXiv1506.00019, 2015.
- [24] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proc. Natl. Acad. Sci.*, vol. 79, no. 8, pp. 2554-2558, 1982.
- [25] Zhang, G.P.; Kline, D.M., "Quarterly Time-Series Forecasting With Neural Networks," *Neural Networks, IEEE Transactions on*, vol.18, no.6, pp.1800,1814, Nov. 2007.
- [26] Rentz, O. et al., Report on BAT in German Copper Production (Final Draft), University Karlsruhe (DFIU), 1999.
- [27] 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.