Aspects of Artificial Neural Networks as a Modelling Tool for Industrial Processes

Mohamed Khadir
Department of Computer Science, University Badji Mokhtar of Annaba, Algeria

Abstract: *In order to investigate the behavior of industrial processes for design, fault prevention, prediction, control, etc., a model of the process is necessary. Due to inherent nonlinearities proper to industrial processes, and/or nonlinearities due to the characteristics of the valves and pumps forming the entire industrial plant, nonlinear models are desired. Complete mathematical models of such plants proves to be time and efforts consuming, when not totally unrealizable. The fact that Artificial Neural Networks (ANNs) have been proven, by Cybenko, able to represent any nonlinear function, as well as their easy implementation, led to their widespread usage in the modeling community; often not at best and ending in controversial results. This paper proposes a methodology for designing and validating ANN models for modeling industrial plants, taking into consideration typical industrial constraints such as restricted data sets. The approach is applied to an industrial milk pasteurization plant.*

Keywords: *Artificial neural networks, multi-layer perceptron, nonlinear models, pasteurization plant.*

Received July 10, 2004; accepted September 17, 2004

1. Introduction

It has been proven in [7] that a continuous function can be arbitrarily well approximated by a feed-forward network with only one single hidden layer, where each neuron in the hidden layer has a continuous sigmoidal nonlinearity. This type of neurons, when used in a network configuration gives the celebrated Multi-Layer Perceptron (MLP). The MLP is, probably the most-often considered ANN for function approximation or modeling [18]. Narendra and Parthasarathy [16] first used ANN, and particularly MLPs, for modeling and control of dynamical systems in 1990. Since then, ANNs have been widely reported in engineering literature. A computer search revealed 9955 articles between 1990 and 1995 containing the words "neural networks" [17]. However, only 14 articles were concerned with real life applications. The second half of the 90's saw an increasing use of ANN in real life applications and engineering companies started using ANNs e. g., Pavilion Technology Inc [19].

Nowadays, ANNs, especially MLPs rather than other neuron based networks, e. g., Radial Basis Function, are widely used for modeling industrial processes as they do not suffer from the curse of dimensionality. Because of their easy implementation, MLPs often give some sort of results, even when an MLP is trained with data gathered from a given process without prior knowledge and understanding of the latter. It is often the case that the ANN network topology is chosen on a test trial basis, without any consideration of *a-priori* knowledge [2, 3, 13, 6]. This often leads, to a considerable loss of time as well as controversial results.

In an excellent and most recent survey Magali *et al.*, [15], summarized the evolution of ANN usage in modeling industrial plants from the 1980's to date. Although the survey, provides a strong background information for the choice of a given:

- Network's topology and nature.
- Activation function.
- A training method.

For a specific field of application, it fails to link any *a-priori* knowledge, and previously gathered data information for the modeling phase of a specific industrial plant.

The following sections describe typical problems encountered when an ANN approach is used for modeling and validation of a given industrial process, as well as a clear methodology in order to extract the maximum benefits from an *a-priori* knowledge of the plant during modeling.

2. Informative Data for ANN Modeling

The basis of any black-box modeling approach is informative input/output data from the process. This remains true in the case of ANN modeling. A complete theoretical ANN model would be built using a data set covering the entire input/output space. In practice, this is not realizable for feasibility reasons. Instead, models are built to be valid in a specific operative region, the data used for training and validation have to be clustered within this region.

The data are obtained by applying test protocols on the industrial plants. These protocols consist of changing the value of the input (s) e. g., valve position, flow, current, etc. and record the value of the output (s). For models used for control, and in the regulation case, the operative region is around the reference set point. It is often the case when working on industrial plant to encounter the following difficulties:

- Restricted plant availability for productivity reasons.
- The plant is badly and/or not correctly instrumented to permit the gathering of informative data. Adding further instrumentation, may be difficult/impossible for technical/management reasons.
- Specific operative regions may be impossible to investigate for security and/or productivity reasons.

These inconveniences are outside the modeling engineer prerogatives, and can only be dealt with rather than solved. On one hand a lot of diplomacy is often required to negotiate a decent amount of testing time, in order to conduct satisfactory test protocols. On the other hand, one can only accommodate with the existing instrumentation as fitting of further sensors and transmitters may not be allowed/possible.

The test protocol should be lengthy enough to capture the rise time of the process. It is often the case for slow processes that the output appears to be settled when it is, in fact, still converging. Therefore the engineer has to be patient conducting such tests. Once the process rise time is captured, subsequent shorter variations of the input (s) are to be considered in order to capture higher frequency behaviors.

A cross-validation training method may be used in order to maximize the information extracted from a small data set for training and validation. This method will be detailed later, section 4.2.

3. Network Topology

The choice of a good network topology is not a straightforward task. There are no hard rules or theorems to find an optimal topology for a given set of input/output data. However, an appropriate topology can be found by starting with a small network (for example a 1-2-1 topology) that is grown until it reaches a size, which gives a good prediction model. This approach is called network growing.

Alternatively, the network topology can be found by performing network pruning using Optimal Brain Damage (OBD) techniques developed by Le Cun [14]. Starting with a sufficiently big topology, the network is pruned by eliminating the links containing smaller weights using a weight elimination method.

A more practical approach would be to have a clear idea on the needed number of inputs/outputs, and at least a slight idea on how these interact with each other. This comes to have an idea on the system's order, a crude linear model may help to approximate the global order of the system.

As showed by Aeyels [1], a system is observable if 2n+1 measurements of the output are taken, where n is the order of the system. For systems with different outputs for different set of inputs, n measurements are sufficient to ensure observability. Therefore, the order of the system can give valuable information on how many delayed inputs and delayed estimated outputs are to be used in the input layer. This will dictate the use of the number of neurons in the first and second layer, giving an idea on the entire network topology.

4. Training and Validation

After deciding on a given network topology, the MLP has to be trained in order to define its parameters, i. e., the weights and biases of each neuron. The training operation is widely described in the literature [18, 22], and will not be repeated in this paper. Note that the goal of this paper is not to speculate on the efficiency of some training algorithm over another, this is widely covered in [18].

After running the training or learning phase, the prediction capabilities of the MLP have to be tested on a new set of data, this is termed split sample validation.

In the following sub-sections the issues of overtraining and the usage of small data sets for training and validation is addressed and a solution is proposed.

4.1. Overtraining and Early Stopping

The estimation of the parameters, weights and biases of the MLP is performed by minimizing an error criterion [18, 22]. The learning algorithm can then be run until no further improvement is reached, i. e., until a global (or local) minimum is reached. However, it was noted earlier in the ANN literature [20, 21] that if the model is evaluated on a validation set (using the validation error), it first improves with the number of iterations. It then starts to deteriorate with increasing number of iterations despite the fact that the training error will continue to decrease. This phenomenon is termed overtraining. In other words, with the increasing number of iterations, the network tends to learn the training data set, which leads to a poor generalization, and a bad prediction for different set of inputs. In order to overcome this problem, a supervised learning method focussing on the value of the validation error can be used. This method checks the validation error at each iteration (or number of iterations) in order to define when it starts deteriorating (increasing). At that point in training, the value of the weight and biases are saved. In order to make sure that this is a global minimum (or at least a decent local minimum) the training is continued for a sufficient number of iterations to make sure that the value of the validation error will not increase further.

4.2. Cross Validation

Cross validation is an improvement over split sample validation in the sense that all available data is used for validation. This is proven more efficient when the size of the sample space is constrained and limited. The method is also called cross model selection.

Practically, the method entails the dividing of the entire data set D to n subsets. This results in *(n-1)* models to be identified using k data subsets. Each time a different subset is selected for validation and the rest of the data subsets are used for training. Having several validation estimates covering the entire training data set gives a better confidence degree to the estimates.

At the end of the cross validation, *(n-1)* models are obtained, we can then choose the best one or the one that best fits the region of interest, or use all the *(n-1)* models. All *(n-1)* models obtained may be combined in a linear way to give an overall model.

5. Application to a Pasteurization Plant

In order to illustrate the difficulties of industrial plant modeling, and to prove the efficiency of the approach presented in this paper, the modeling of an industrial pasteurization plant is investigated. The plant is located in Drogheda Ireland, and is part of the Glanbia group. It has a strategic importance for the milk production in the province of Leinster.

5.1. Milk Pasteurization Process

The pasteuriser used is a Clip 10-RM Plate Heat Exchanger (PHE) from Alfa Laval. A PHE consists of a pack of stainless steel plates clamped in a frame. The plates are corrugated in a pattern designed to increase the flow turbulence of the medium and the product [5].

The pasteuriser is divided in five sections S1 to S5. Section S4 and S2 are for regeneration, S1 and S3 for heating and S5 for cooling. In the Clip 10-RM the milk treatment is performed as shown in Figure 1. First the raw milk at a concentration of 4.1% enters section S4 of the PHE at a temperature of 2.0°C. It is then preheated to a temperature of 60.5°C by the outgoing pasteurized milk, which as a result is reduced to a temperature of 11.5°C. Passing this section, the milk now at a temperature of 60.5°C, enters section S3 where its temperature increases to 64.5°C using hot water as a medium. The milk, before reaching the next section, is first separated from the fat then standardized and homogenized to a concentration of 3.5%. It then enters section S2, where it is preheated to a temperature of 72°C using the already pasteurized milk as a medium. The milk is finally brought to the pasteurization temperature in section S1 (75.0°C) using hot water at around 77.0°C as a medium. After that the homogenized pasteurized milk is held at the pasteurization temperature for 15s in the holding tube section before being cooled using the incoming cold milk in section S4 and section S2.

Finally, the pasteurized milk enters the cooling section (section S5) at a temperature of 11.5°C. The milk is chilled to a temperature of 1.0°C using propylene glycol as a medium at a temperature of -0.5°C. Note that the water for the heating sections S3 and S4 is brought to the adequate temperature in steam/water heater of type CB76 from Alfa Laval.

As shown in Figure 1, milk pasteurization temperature is a function of three inputs: Steam flow injected in steam/water heater 1, steam flow injected in steam/water heater 2 and the milk input temperature, labeled as F_{v1}, F_{v2}, and T_{im} respectively.

The milk pasteurization temperature is then given by a Multi Input Single Output (MISO) system, having F_{v1}, F_{v2}, and T_{im} as inputs and y, the milk pasteurization temperature, as output.

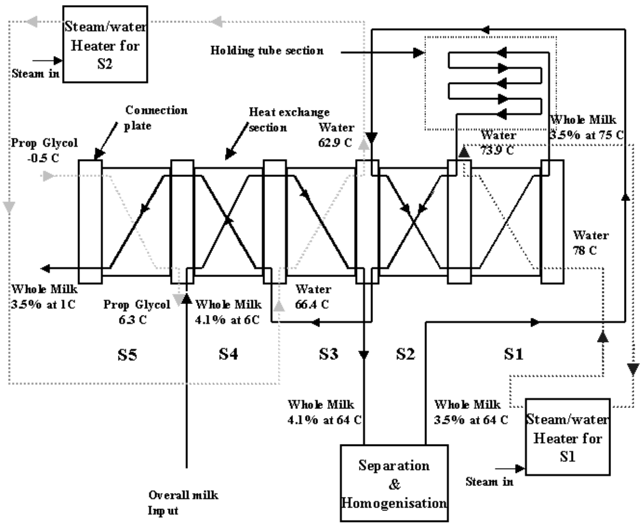


Figure 1. General layout of the pasteuriser plant.

5.2. Test Protocol and Available Data

Five sessions for data collection were generously allowed by the plant management, each of approximatively 4 to 5 hours long, which is the maximum period, for which a second smaller pasteuriser can keep production going. During the testing phase, we were allowed to vary the steam flow valves around their normal operative points in order to provoke changes around the nominal output separation and pasteurization temperatures i. e., 64.0°C and 75.0°C. The plant responses for five test protocols have been concatenated and are given in Figure 2. It can be seen that the changes provoked range between 55.0°C to 85.0°C for pasteurization temperature Top1 and, 50.0°C to 74.0°C for the separation temperature Top3.

Clearly the process rise time is greater than 30 minutes (1800s). Thus a sampling period Ts of 12s was found to be more economical than 1s, and still satisfies the usual Shannon sampling theorems [10] as well as sampling requirements for industrial processes given by equation (1).

$$Ts = \frac{Tr}{N} \tag{1}$$

Where Tr is the process rise time at 63%, and N a constant, $30 < N < 50$.

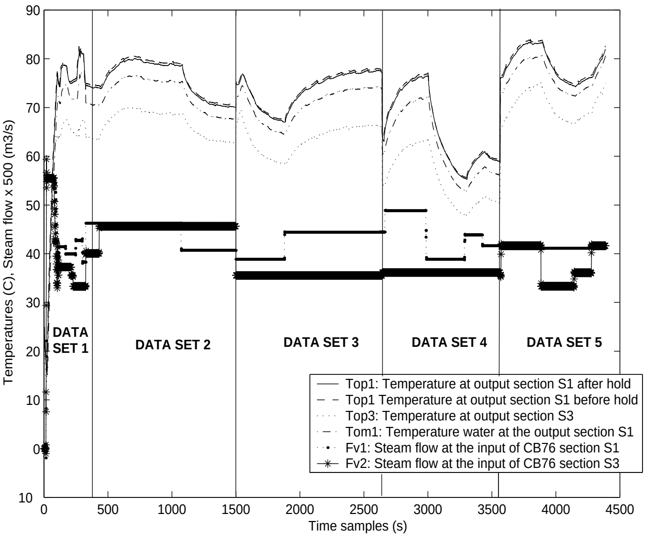


Figure 2. Test protocols and plant response.

5.3. ANN Modeling of the Pasteurization Plant

The MLP topology used is shown in Figure 3. The choice of the inputs has been heavily dictated by the a-priori information gathered from the first principle physical model used in [11]. Where, the output pasteurization temperature can be modeled by an eighth order linear system, this justifies the use of eight delayed signals of $y(k)$ in equation (2). The 12 neurons of the input layer are the results of eight-delayed version of the output milk temperature [1], in addition to the actual inputs F_{v1} and F_{v2} with one delayed version of both. The neurons in the two hidden layers are LOGSIG neurons which are more appropriate for strictly positive data, where the output layer neuron is a PURELIN neuron.

$$y(k) = NNM(y(k-1), y(k-2), \ldots, y(k-8) \\ F_{v1}(k-1), F_{v1}(k-2), F_{v2}(k-1), F_{v2}(k-2)) \tag{2}$$

The overall input milk temperature, is not used in the NNM as the milk is kept at a relatively constant temperature of 2°C, and its use, in the training process, will only introduces a random disturbance to be modeled. In total five subsets of data totaling 4500 samples were used, three for training and a separate subset was used for validation in order to make sure of the validity of the model, the data subsets are clearly shown in Figure 2.

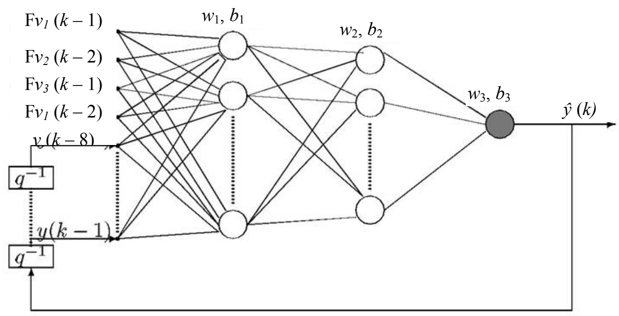


Figure 3. Network topology.

The training algorithm used is the backpropagation algorithm, more precisely a batch version of the backpropagation algorithm [9]. The algorithm is available as a function in the Matlab ANN Toolbox [8].

A cross validation method is used in order to make the best use of all data sets as explained in section 4.2. Three data subsets 2, 3, and 5 shown in Figure 2, were used in turn as a validation set where the rest of the subsets were concatenated and used for training. The choice of these particular subsets is motivated by the fact that they describe best the plant behavior around the pasteurization temperature i. e., 75°C. Validation, to avoid overtraining, a Sum-Squared Error (SSE) on the validation set is calculated and the model parameters are chosen when the SSE is minimum according to the early stopping rule, section 4.1. An example of overtraining can be shown in Figure 4, where it can be seen that the validation SSE in dotted line, starts to rise at the 55th epoch. At the end of the cross validation three distinct ANN models were obtained.

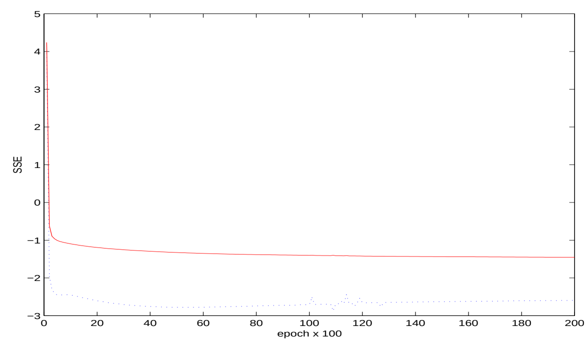


Figure 4. Validation SSE progression versus training epoch number.

The definite NNM consists then, of a linear combination of 3 ANN topologies, given in Figure 5.

The weights of the linear combiner are determined using a simple least square method, training the overall network model NNM using the entire set of data as shown in Figure 2. The weights analytical formulation is given in equation (3).

*Tom*1 is the milk pasteurization temperature values (at the output of section S1, Figure 1) obtained from the test protocols.

$$\begin{bmatrix} W_1 \\ W_2 \\ W_3 \end{bmatrix} = \left(\sigma^T \sigma\right)^{-1} \sigma^T \begin{bmatrix} T_{om1}(1) \\ T_{om1}(2) \\ \vdots \\ T_{om1}(n) \end{bmatrix} \quad (3)$$

where:

$$\sigma = \begin{bmatrix} y_1(1) & y_2(1) & y_3(1) \\ y_1(2) & y_2(2) & y_3(2) \\ \vdots & \vdots & \vdots \\ y_1(n) & y_2(n) & y_3(n) \end{bmatrix}$$

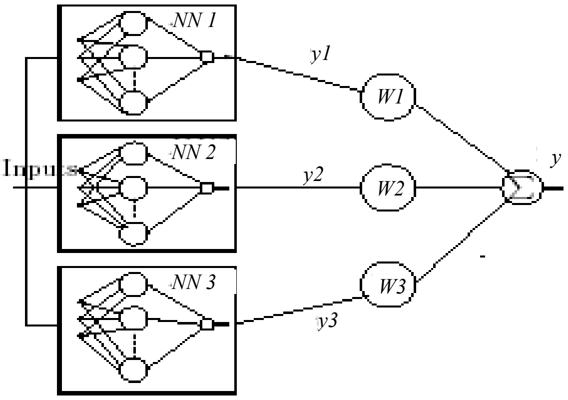


Figure 5. Linear combiner used to obtain the final ANN model.

5.4. Results Assessment

The ANN model response is compared to a previously obtained linear physical model [11], both responses compared in Figure 6. It can be seen, that for the ANN model, the Mean Absolute Error (MAE) between the measured and the predicted output is inferior to 0.8 °C at each sample, which is the uncertainty of the fitted temperature sensor [4]. Therefore, any additional improvement to the model is obsolete unless better sensors are fitted. Note that this is not the case for the linear model at some temperatures further away from the operating setpoint. The physical model, however, offers many laudable attributes as a better comprehension of the plant and thus a better generalization capabilities. This remains a discussion issue between the black box and the physical modeling schools, which is outside of the paper scope. Table 1 gives the overall MAE for the ANN and the benchmark physical model. The ANN model was finally used in a Model Based Predictive Control scheme [12].

6. Conclusions

When modeling industrial plants with ANNs, convergence time and memory allocation are not critical issues, as most of the time the available data set is restricted and the modeling is done off line. Therefore, the objective becomes: How to use, at best, the restricted data set, and choose an adequate topology, rather than determining which gradient method allow fast convergence and/or better memory allocation. The key recommendations for an ANN approach applied to industrial plants, are then summarized in what follows:

1. A careful choice of the data area where the test protocols are to be conducted is paramount in order to gather the maximum amount of informative data.
2. The topology of the MLP used should be heavily dictated by a-priori knowledge of the plant. If possible a crude linear model should be able to give an idea on the order of the system and therefore dictates the number of delayed signals used in the input layer.
3. Early stopping is of paramount importance in order to avoid overtraining.
4. Cross validation proves to be useful when the size of the sample space is constrained and limited. Moreover, having several validation estimates covering the entire training data set gives a better confidence degree to the estimates.

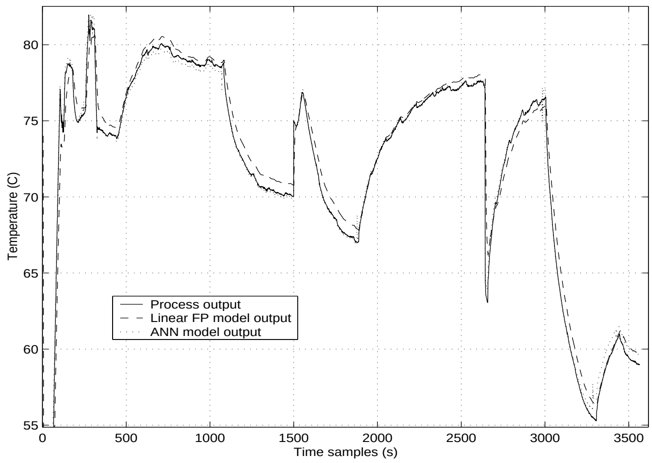


Figure 6. ANN versus physical model response.

Table 1. ANN and physical model MAE.

	ANN Model	Physical Model
MAE (°C)	0.6349	0.2980

References

[1] Aeyels D., "Generic Observability of Differentiable Systems," *SIAM Journal on Control and Optimization*, vol. 19, pp. 595-603, 1981.

[2] Al-Duwaish H., Gouti L., Halawani T., and Mohandes M., "Use of Artificial Neural Network Process Analyser: A Case Study," *in Proceeding of the European Symposium on Artificial Neural Networks*, Bruges, Belgium, pp. 465-470, 2002.

[3] Al-Duwaish H., Othman T., and Amjad S., "Neural Network Identification of a Gaz Plant NGL," *in Proceeding of IASTED International Conference, Intelligent Systems and Control*, Salzburg, Austria, pp. 307-311, 2003.

[4] Alfa Laval, *Technical Manual: Temperature Transmitter for Mounting in Sensor Pt-100*, Tetra Pak, 1998.

[5] Bylund G., *Dairy Processing Handbook*, Lund, Sweden,Tetra Pak Processing, 1995.

[6] Cabassud M., Delgrange-Vincent C., Cabassud M., Bourlier J. M., and Lain J. M., "Predictive Control of an Ultrafiltration Pilot Plant Based on Neural Network," *in Proceeding of the Engineering Applications of Neural Networks*, Gagliari, Italy, pp. 23-30, 2001 .

[7] Cybenko G., "Approximation by Superposition of Sigmoidal Function," *Mathematics of Control, Signals, and Systems*, vol. 2, pp. 303-314, 1989.

[8] Deamuth H. and Beale M., *Neural Networks Toolbox: User's Guide*, version 3.0, The MathWorks Inc., Natick, MA, 1998.

[9] Hagan M. T., Deamuth H. B., and Beale M. H., *Neural Networks Design*, Boston, MA, PWS Publishing, 1996.

[10] Higgins J. R., *Sampling Theory in Fourier and Signal Analysis: Foundations*, New York, Oxford University Press, 1996.

[11] Khadir M. T. and Ringwood J. V., "First Principle Modeling of a Pasteurization Plant for Model Predictive Control," *Journal of Mathematical and Computer Modeling of Dynamical Systems*, vol. 3, pp. 284-301, 2003.

[12] Khadir M. T. and Ringwood J. V., "Linear and Nonlinear Model Predictive Control Design for a Milk Pasteurization Plant," *Control and Intelligent Systems*, vol. 31, no. 1, pp 37-44, 2003.

[13] Lam J. Y. H. and Geddam A., "A Neural Network Approach to Tool Wear Monitoring in End Milling Operations," *in Proceeding of IASTED International Conference, Intelligent Systems and Control*, Salzburg, Austria, pp. 156-161, 2003.

[14] Le Cun Y., Danker J. S., and Solla S. A., "Optimal Brain Damage," *Advanced in Neural Information*, vol. 2, pp. 598-605, 1989.

[15] Magali R., Meireles G., Paulo E., Almeida M., and Simeoes M. G., "A Comprehensive Review for Industrial Applicability of Artificial Neural Networks," *IEEE Transactions on Industrial Electronics,* vol. 50, no. 3, pp. 585-601, 2003.

[16] Narendra K. S. and Parthasarathy K., "Identification and Control of Dynamical Systems Using Neural Networks," *IEEE Transactions Neural Networks*, vol. 1, pp. 4-27, 1990.

[17] Narendra K. S., "Neural Networks for Control: Theory and Practice," *IEEE Transactions Neural Networks*, vol. 84, no. 10, pp 1385-1406, 1996.

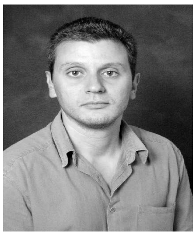
[18] Norgaard M., Ravn O., Poulsen N. K., and Hansen L. K., *Neural Networks for Modeling and Control of Dynamic Systems*, Springer-Verlag, London, 2000.

[19] Qin S. J. and Badgwell T. A., *An Overview of Nonlinear Model Predictive Control Applications in Nonlinear Model Predictive Control*, Birkhauser, Switzerland, 2000.

[20] Sjoberg J. and Ljung L., "Overtraining, Regularisation, and Searching for Minimum in Neural Networks," *in Proceeding of the 10th IFAC Symposium on System Identification,* Copenhagen, vol. 2, pp. 49-72, 1992.

[21] Wahba G., *Three Topics in Ill-Posed Problems*, in Engl H. and Groetsch C. (Eds), Academic Press, 1987.

[22] Zurada J. M., *Artificial Neural Systems*, West Publishing Company, St Paul, 1992.



Mohamed Khadir received his engineering Bachelor degree in electronics, majoring in control from the University of Badji Mokhtar Annaba, Algeria, in 1995. After two years work in the computer industry, he undertook Master of engineering at Dublin City University, Ireland, graduating with first class honors in 1998. He received a PhD degree from National University of Ireland, Maynooth, Ireland in 2002. He then continued with this institution as a post-doctoral researcher until September 2003, when he joined the Department of Computer Science in his original University of Badji Mokhtar Annaba, Algeria, as a senior lecturer.