



HAL
open science

Neural network-based software sensor: Data set design and application to a continuous pulp digester

Pascal Dufour, Sharad Bhartiya, Prasad S. Dhurjati, Francis J. Doyle Iii

► To cite this version:

Pascal Dufour, Sharad Bhartiya, Prasad S. Dhurjati, Francis J. Doyle Iii. Neural network-based software sensor: Data set design and application to a continuous pulp digester. *Control Engineering Practice*, 2005, 13 (2), pp.135-143. 10.1016/j.conengprac.2004.02.013 . hal-00351646v2

HAL Id: hal-00351646

<https://hal.science/hal-00351646v2>

Submitted on 21 Jan 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**This document must be cited according to its final version
which is published in a journal as:**

**P. Dufour¹, S. Bhartiya¹, P.S. Dhurjati¹, F.J. Doyle III¹,
"Neural network-based software sensor: Data set design
and application to a continuous pulp digester",
Control Engineering Practice, ISSN: 0967-0661
13(2), pp. 135-143, 2005.**

<http://dx.doi.org/10.1016/j.conengprac.2004.02.013>

**All open archive documents of Pascal Dufour are available at:
<http://hal.archives-ouvertes.fr/DUFOUR-PASCAL-C-3926-2008>**

**The professional web page (Fr/En) of Pascal Dufour is:
<http://www.lagep.univ-lyon1.fr/signatures/dufour.pascal>**

1

University of Delaware,
Department of Chemical Engineering,
Newark, DE 19716, USA
<http://www.udel.edu>

Neural Network-based Software Sensor: Training Set Design and Application to a Continuous Pulp Digester

Pascal Dufour^a, Sharad Bhartiya^a, Prasad S. Dhurjati^a
and Francis J. Doyle III^{a,1}

*^aDepartment of Chemical Engineering
University of Delaware
Newark, DE 19716, USA*

Abstract

A neural network based strategy for detection of feedstock variations in a continuous pulp digester is presented. A feedforward two-layer perceptron network is trained to detect and isolate unmeasured variations in the feedstock. Training and validation data sets are generated using a rigorous first principles model. The most important issue discussed here is the design of the data set required to train the artificial neural network. Efficiency and limitation of such an approach are demonstrated using simulations.

Key words: Artificial neural network, fault detection, soft-sensing, pulp digester, fundamental model.

1 Introduction

The pulp and paper industry forms a large and important sector of the chemical process industry. During the period 1996-1998 alone, planned capital investments including expansion of existing facilities exceeded US\$10 billion. This highlights the capital-intensive nature of this industry and the corresponding demand for higher safety and reliability (Leung, Romagnoli, & Bigaran, 2001). In the meantime, practical implementation of advanced industrial control systems, like model predictive control (MPC), requires a watchdog to diagnose the cause of performance degradation. Therefore, early detection of such problems enables the engineer to undertake remedial action including retuning the controller if necessary.

Over the past two decades, a number of approaches for fault detection and diagnosis have been proposed. These myriad techniques (Dash, & Venkatasubramanian, 2000) have been classified and summarized in several reviews and publications. Notable reviews include: observer based methods (Garcia, & Frank, 1997), parameter estimation and knowledge based systems (Patton, Chen, & Nielsen, 1995); knowledge based systems (Sharif, & Grosvenor, 1998); and fuzzy logic methods (Isermann, 1998). More recently, Venkatasubramanian (Venkatasubramanian, 2001, Venkatasubramanian, Rengaswamy, & Kavuri, 2003, Venkatasubramanian, Rengaswamy, Kewen, & Kavuri, 2003a, 2003b) presented a global perspective on these disparate methods. Among these methods, neural networks have very useful properties including the ability to handle nonlinear behavior and tolerance to noise.

¹ Corresponding author: Dept. of Chemical Engineering, UCSB, Santa Barbara, CA 93106, PH (805)893-4731, FX (805) 893-8133, doyle@engineering.ucsb.edu

In this work, preliminary results for a neural network-based software sensor to infer unmeasured variations (referred to as faults) in the feedstock of an industrial pulp digester are presented. This software sensor is developed to predict time-varying model parameters (Qi, Zhou, Liu, & Yuan, 1999, Molga, & Cherbaski, 1999, Porru, Aragonese, Baratti, & Servida, 2000) that affect model-based control performance. The data set required for the neural network creation is usually obtained from plant data experiments. However, in this work, the use of a fundamental model to generate data for unmeasured variables is advocated. A related study is described in (Adal, Bakal, Sönmez, Fakory, & Tsaoui, 1997) where a fundamental model was used to create training data for a neural network. Additional motivation for the use of this approach is that it represents a non-parametric approach.

Various methods exist for the design of the data set considered in this paper:

- Random selection where no clear criterion is applied. This is the most common method in the literature.
- Kohonen self-organizing map (Oja, & Kohonen, 1999) is a clustering method mapping high-dimensional data sets into a space of lower dimension, while preserving the topology (i.e. the spatial relationships) of the data set as best as possible.
- Kennard and Stone design (Kennard, & Stone, 1969) is based on uniform spaced objects over the space. Such an algorithm starts with two most distant objects as members of the data set. Next, all other objects' distances to the first two objects are calculated. The object, which is selected as the next one to join the data set, is the one that has the largest minimal distance to all previously selected objects. The algorithm repeats the selection by comparing all distances of candidates to the already selected objects until

a certain predefined number of objects have been found.

- D-optimal design (Carlson, 1992) aims to minimize the generalized variance of the parameter estimates.
- Full or fractional factorial design (Montgomery, 1997) where combinations between discrete levels are accounted for. Full factorial design contains all the possible combinations of the selected settings of the experimental factors, so that a 2-level full factorial design requires 2^n experiments (n being the number of factors). One way to accomplish a smaller number of experiments, when treating the same number of factors, is the use of fractional factorial design. This operation leads to a partial loss of information, since it might be impossible to discriminate between the effects or the interactions of the effects.

In (Wu et al., 1996), a case study provides a detailed comparison of random selection, Kohonen self-organizing map, Kennard and Stone design and D-optimal design. None of these methods has a distinct advantage over the other methods. Therefore, given the large size of the problem and the simplicity of the approaches, a fractional factorial design method (Zhang, & Subbarayan, 2002, Briceno, El-Mounayri, & Mukhopadhyay, 2002) combined with random selection is applied here for the data set generation.

The main focus of this paper deals with a method to construct, from the digester model, the data set required for the training phase. This paper is organized as follows: first, the problem is described and the fundamental model is briefly introduced. Next, three designs of the training subset of the neural network are discussed. Finally, efficiency, limitations and robustness for the use of each neural network, resulting from the different training subsets, are described.

2 Fault set definition

A continuous digester is a large vertical tubular reactor in which wood chips are converted to wood pulp for papermaking (see schematic in Figure 1). White liquor (aqueous solution of effective alkali and hydrosulfide) strips the presteamed porous wood chips of lignin, freeing the wood fibers. The delignification reaction takes place in the digester section referred to as the cook zone. In this section, both the chips and liquor flow cocurrently. At the end of the cook zone, spent liquor is extracted for chemical recovery. The chips, however, continue their downward flow in the wash zone, composed by modified continuous cooking (MCC) zone and extended modified continuous cooking (EMCC) zone, where they encounter weak liquor flowing countercurrently. The wood pulp is finally extracted at the outlet of the EMCC. Control of the pulp quality, characterized by Kappa number, is the key control objective. Faults addressed in this work were identified through discussions with a mill partner. They mainly deal with variations of the naturally occurring feedstock properties and are summarized below:

- variation in moisture content of the wood chips fed to the chip bin,
- unmeasured density of each of the 5 wood chip components: a) high reactivity lignin; b) low reactivity lignin; c) cellulose; d) galactoglucomannan; and e) araboxyylan,
- unmeasured density of each of the 2 white liquor species: a) effective alkali (EA); and b) hydrosulfide (HS).

Moisture content is measured infrequently (in some cases, once per day). No known measurements exist for the 5 wood components. However, these are

accounted for in the fundamental model. Wood, being a naturally occurring raw material, exhibits wide stochastic variability in moisture content as well as composition. Upsets in the upstream liquor recovery section are often manifested as changes in white liquor densities. Each of the eight previously mentioned faults adversely affect control of the endpoint Kappa number. Indeed, it has been previously determined (Wisniewski, & Doyle III, 1998) that stochastic variations of the feedstock has a significant impact on the Kappa number at the outlet of the digester, even under advanced process control. Therefore, there is a need to estimate these stochastic variations in order to improve closed-loop performance. From a control theoretic point of view, these faults can also be interpreted as unmeasured. Within this framework, identification of these disturbances will enable use of a feedforward scheme for disturbance rejection based on "soft sensor" predictions. However, in the current work, the changes of the eight feedstock quality descriptors are considered as faults and the aim is to estimate their magnitude.

3 Continuous pulp digester model

The first step in the construction of a neural network model is the development of a suitable data set, which is randomly split into a training subset and a validation subset. However, no measurements exist for seven of the eight quality descriptors (see fault summary list above). To circumvent this problem, a rigorous first principles model is employed to create the necessary data set. For modeling purposes, the digester is approximated as a series of 150 continuous stirred tank reactors. The first principles model consists of mass balances for the non-porous solid and the free liquor components and energy

transfer between the 2 phases within each reactor. Sufficient degrees of freedom exist to tune the model to match mill behavior. For modeling assumptions and a description of the conservation laws, the reader is referred to (Wisnewski, Doyle III, & Kayihan, 1997). The fundamental process model is simulated using Matlab (Doyle III, & Kayihan, 1999) and is used as the "plant".

4 Artificial neural network design

The feedforward multilayer backpropagation artificial neural network can be viewed as a nonlinear autoregressive model with external inputs (Su, Mc Avoy, & Werbos, 1992). An overview of training methodologies is presented by Hush and Horne, 1993. In this work, a backpropagation network is trained such that the neural network outputs reconstruct the behavior of a selection of faults among the eight faults discussed previously. A sensitivity analysis reveals that, among all the sensors actually available at the plant, two are important here: measures of EA and HS content in the liquor at the upper extract of the digester (Fig. 1) are the most useful for detecting the cause of changes in the Kappa number. Hence, both of these will be selected to infer the magnitudes of the eight quality descriptors that affect the Kappa number.

Network inputs consisted of EA and HS measurements over a fixed, past receding horizon, thereby, enabling process dynamics to be incorporated. The original data set generated from the fundamental model is split into training and validation sets. The hidden layer consists of a fixed number of sigmoidal nodes (selected by trial and error). Henrique, Lima, and Seborg, 2000 discuss node pruning techniques based on orthogonal least squares, which yield parsimonious models. During the training, the network weights and biases are

adjusted such that the sum of the squared errors (SSE) during prediction is minimized by the backpropagation algorithm. Under these conditions, it is necessary to stop training once an overfit is indicated. A commonly used method to identify overfit is cross-validation: the weights and biases based on the training subset are applied to the validation subset to evaluate the performance in terms of the SSE. Typically, the SSE for the training set decreases with the number of iterations and levels off to a constant value when a local minimum is obtained. In an overfit situation, the SSE for the validation first decreases but finally increases, whereas the SSE for the training is decreasing. When the SSE of the validation subset increases, the algorithm is overfitting the training data. In this work, the training procedure is stopped when the SSE of the validation subset starts to increase while the SSE of the training subset is still decreasing.

A key issue addressed in this paper is the initial design of the original data set: which set of behaviors of the eight faults have to be simulated? Here, use of large data sets for training requires large computational time along with the resulting increase in model complexity². On the other hand, using a small and non-representative data set for training leads to poor validation results. The following procedure is used to build the neural network:

- choose the variation set that describes the combination of the n selected faults that appear during the plant simulation using factorial design and random selection,
- given this variation set, build the related data set using the set of plant simulations,

² With a 800MHz PC, 3000 simulations of 12 hours can be accomplished in 24 hours using IETek model v1.0 and Matlab.

- split the data set randomly into training subset and validation subset,
- train the neural network: the network inputs consist of upper extract EA and HS content measurements over a past horizon. The horizon is chosen to be 18 hours (with sample time of 1 hour), which is long enough to capture the dynamic behavior of the digester. Thus, the neural network inputs are:

$$EA(t) \dots EA(t - 18) \quad HS(t) \dots HS(t - 18)$$

The network outputs consist of the n faults which are shifted in time by the evident delay (from the digester inlet to the upper extract) of 4 hours. Thus, the network outputs are:

$$fault_1(t - 22) \dots fault_n(t - 22)$$

The number of nodes is adjusted by trial and error for each data set.

In the next three sections, three different strategies are discussed for a systematic bottom-up approach for the variation set design.

5 Simple data set design

As a first attempt, faults were introduced in the plant simulator as a combination of initial steps variation for each of the $n_f = 8$ feedstock quality descriptors previously described. In order to avoid computation of an excessive number of simulations, the variation set is designed via fractional factorial design as follows:

- Each of the fault magnitude can take n_v values different from the nominal value (referred to as 100%).

- When more than one fault occurs, all of their magnitudes are the same.

Although these assumptions are restrictive, a large number of $n_v \times 2^{n_f} = 2048$ simulations need to be computed. The magnitudes of the faults range between 92% and 108% with a step of 2%. It is noted that such a full-factorial design method may not be practical if the number of faults is large or for a finer resolution of changes in their magnitudes. In such cases, one may use optimal factorial designs as discussed in (Wu et al., 1996). Each case is simulated for a 12h duration (therefore, a large set is involved in this simple data set design). This duration is sufficient for faults to be completely manifested in the residual EA and HS density measurements at the upper extract. A network with 8 nodes yielded the best results during the validation phase. Detection of the normalized unmeasured density variations are shown, in the validation subset, for the moisture content (Fig. 2a), the cellulose density (Fig. 2b), the araboxylan density (Fig. 3a) and the HS density (Fig. 3b). The evident delay in detection after the step changes is due to the time needed for transport of the feed from the top of the digester to the extraction screen. For these 4 faults, after small transient errors, the static error is small and the neural network behaves efficiently. Large prediction errors were observed for detection of changes in high reactivity lignin, low reactivity lignin, effective alkali and galactoglucomann. These results demonstrate that some variations in feedstock (moisture content, cellulose density and also, to a certain extent, the araboxylan density and the HS density) may be inferred using a neural network with the selected inputs (EA and HS at extraction screen). These results will be used to reduce the size of the data set in the next two data set designs. Testing the neural network in regions outside the data set (with uncorrelated variations for the magnitude of each of the eight faults) shows poor detection

ability for all 8 faults including the 4 faults that were previously reconstructed correctly. Detection of variations in moisture content and cellulose density by the artificial neural network outside the training region are shown in Fig. 4a and Fig. 4b, respectively. These results are very poor and the neural network is not able to estimate correctly the fault magnitudes as done previously. This underlines the drawback of poor extrapolation capabilities of the neural network. The poor results can be attributed to the current training subset that covers the different operating conditions very sparsely.

6 Reduced data set design

Despite selecting a coarse resolution of variations in the fault variables (2% changes) in the previous section, the large dimension of the input space (dynamic description of 8 faults) caused the data set to be large. Moreover, it did not capture all possible system behaviors since, among all possible variations, only step variations were assumed in the fault scenario. In this section, a new variation set is constructed. In the new design, a finer resolution of 1% step changes is employed. Based on results of the previous section, our attention is restricted to three of the four faults that could be detected by the EA and HS measurements at the upper extract. The three faults selected are moisture content, cellulose and HS densities (i.e., the ones that gave the most promising results in the previous simulations). This reduction of the number of faults aims to reduce the data set size involved to yield reasonable computation times. In order to avoid computational burden during the neural network training, the variation set is designed as follows:

- each fault magnitudes can take different values about their nominal value.

- when more than one of the faults occur, any combination of their magnitudes are allowed.

Thus, each magnitude can take any value between 94% and 106% with a step of 1% (i.e., $n_v = 13$). With a full factorial design, this leads to the computation of $n_v^{n_f} = 2197$ simulations to build the data set. 7 nodes were used in the hidden layer of this neural network. The use of this neural network shows enhanced performance in the trained region. The performance outside data set, i.e., in more realistic cases, is now discussed. In the first run, the variation set is defined with the same bounds as in the training, namely [94%, 106%], but the discrete step size is now changed to 0.02% instead of the 1% chosen in the training set. After relatively small transient errors, it is seen (e.g. Fig. 5 for the HS density) that the network is able to track such step changes. The transient errors are obviously still present due to the abrupt changes at the top of the digester but the static errors are quite small. As discussed previously, the delay is unavoidable since it takes finite time for the feed to reach the extraction screen. In the second run, the step variations in the interval [94%, 106%] with a discrete step of 1% are extended to all 8 initial faults described previously. The results in the previous section showed that 4 of these 8 faults were not detectable. However, all of these 8 faults still have an impact on the other neural network responses (e.g. Fig. 6 for the HS density) resulting in static errors during prediction by the neural network. It is also of interest to study the sensitivity of the network outputs to the typical manipulated variables (MVs) since the neural network will be used in the control strategy where the MVs affect the EA and HS measurement used in the neural network. Three such candidate MVs are considered (see Fig. 1 for their respective location on the plant):

- chip flowrate in the digester,
- liquor flowrate at upper extract,
- cook temperature.

Simulations show that the neural network responses are affected by these MV changes. As an example, Fig. 7 depicts the inferred HS density when a step increase in the upper extract flowrate has been introduced as a step variation from 100% to 103% with a simultaneous step variation in HS concentration. Finally, the design of the variation set for this second neural network shows that a sufficiently discretized variation set can lead to good interpolation results for untrained situations. However, the MVs must also be taken into account in the variation set, hence, producing a larger training subset. This will lead to the next training subset design for a new neural network that will be more robust with respect to MV variations.

7 Robust data set design

The design of a third data set is now described for detection of the moisture content, cellulose, araboxylan and HS density variations. For each of these 4 faults, the same procedure is used and is described here. A variation set is constructed including the relevant fault and the 3 MVs introduced previously. In order to reduce the time needed for the calculation of the data set, the magnitude of each of the $n_f = 4$ parameters is allowed to take $n_v = 9$ different values rather than 13 as in the previous case. Therefore, with a full factorial design, $n_v^{n_f} = 6561$ different simulations, where changes in the variation set are still implemented as step function, have to be performed. The number of simulations to run is therefore 3 times higher than previously, hence increasing

the computational burden. Of these simulations, 2000 runs were chosen randomly to create the training subset. 11 nodes were used in the neural network. As expected, the neural network performs well during generalization. The extension to regions outside the data set, where the MV changes are filtered through a first order transfer function (Fig. 8) rather than a step change as in the variation set, leads to promising results. Similar observations are made for detection of variations in the moisture content (Fig. 9). Transient errors exist but the final tracking of the variation is effective. One can see that the deviation from steady state after 8 days in the cook temperature (Fig. 8) does not affect appreciably the neural network response (Fig. 9). A simulation in another region outside the data set allows verification of robustness of the neural network predictions. A periodic fault for HS is introduced as a pulse train (Fig. 10). In the meantime, each of the seven other faults are sequentially introduced as a pulse: the first fault of this sequence is the change in moisture content (Fig. 11a) and, following the order of the list given in the fault set definition section, it ends with the change in EA density (Fig. 11b). After transient errors, it is seen that the neural network rejects these disturbances (Fig. 10) except for the moisture content (the first fault) and the cellulose density (the fourth fault): in these cases, a static error exists after the additional fault occurs. Since these 2 faults may be inferred using a similar neural network, the output of the two related neural networks can be combined to the actual neural networks output to correct such errors. Otherwise, after transients due to the introduction of each fault, the expected value is found with good accuracy.

8 Conclusions

An approach for the evaluation of the magnitude in the variations of the feedstock properties in a pulp digester has been presented. Based on a first principles model, a neural network approach has been used to build an inference of some of the unmeasured variations from the 2 measurements available at the upper extract in the digester. As relatively few plant measurements are available for the necessary training sequence, the design of the data set based on the plant simulator has been discussed.

The main contribution in this paper is a description of the selection of the data set. Given the large volumes of data necessary for the design of the neural network to detect feedstock variations, a systematic bottom-up approach was used. Thus, a very simple data set was initially designed. Even without the MVs at this stage, it was clear from this design that it was not possible to detect changes in 4 out of the 8 proposed faults. Also, it became clear that 2% changes in inputs is too coarse and that the network does not work well with smaller changes in inputs. In the next design, the magnitudes of changes were reduced to 1%, and the results in figure 5 reveal the improvement. The next level of sophistication involved studying the impact of MVs and this was addressed in the robust data set design. A relatively small data set has been built to avoid numerical problems during the training phase of the neural network. Unfortunately, a small data set did not cover every process behavior based on the fault situations. Among the infrequent moisture measurement and the 7 unmeasured feedstock properties, one may estimate the magnitude of the moisture content, the cellulose density and also to a certain extent, the araboxytan density and the HS density.

Future developments will address improvement of the robustness by feeding the current responses in cascaded neural networks to correct the errors. The current approach should be evaluated with realistic changes in moisture and feedstock properties and in a control structure with data obtained from an actual mill. In order to decrease the size of the training subset, one should explore dimension reduction methods before training phase such as PCA or independent component analysis (ICA). In (Li, & Wang, 2002), after reduction through ICA, signals featuring dynamic behavior were reconstructed with good accuracy. Future work will address the use of this model-based sensor to update the corresponding model parameters used in the model predictive control strategy in order to improve closed-loop performance.

9 Acknowledgments

The authors acknowledge funding from the Department Of Energy (Grant DE-FC07-00ID-13882) and Westvaco, Inc.

References

- [1] Adal, T., Bakal, B., Sönmez, M.K., Fakory, R., & Tsaoi, C.O. (1997). Modeling nuclear reactor core dynamics with recurrent neural networks. *Neurocomputing*, 15(3-4), 363-381.
- [2] Briceno, J.F, El-Mounayri, H., & Mukhopadhyay, S. (2002). Selecting an artificial neural network for efficient modeling and accurate simulation of the milling process. *International Journal of Machine Tools and Manufacture*, 42(6), 663-674.

- [3] Carlson, R. (1992). *Design and optimization on organic synthesis*. Elsevier, Amsterdam.
- [4] Dash, S., & Venkatasubramanian, V. (2000). Challenges in the industrial applications of fault diagnostic systems. *Computers & Chemical Engineering*, 24(7), 785-791.
- [5] Doyle III, F.J., & Kayihan, F. (1999). Reaction profile control of the continuous pulp digester. *Chemical Engineering Science*, 54, 2679-2688.
- [6] Garcia, E.A., & Frank, P.M. (1997). Deterministic nonlinear observer-based approaches to fault diagnosis: a survey. *Control Engineering Practice*, 5(5), 663-670.
- [7] Henrique, H.M., Lima, E.L., & Seborg, D.E (2000). Model Structure determination in neural network models. *Chemical Engineering Science*, 55, 5457-5469.
- [8] Hush, D.R., & Horne, B.G. (1993). Progress in supervised neural networks: What's new since Lippmann. *IEEE Signal Process. Mag.*, 10, 8-39.
- [9] Isermann, R. (1998). On fuzzy logic applications for automatic control, supervision and fault diagnosis. *IEEE Transactions on Systems, Man, and Cybernetics.*, 18(2), 221-235.
- [10] Kennard, R.W., & Stone, L.A. (1969). Computer aided design of experiments. *Technometrics*, 11, 137-148.
- [11] Leung, D., Romagnoli, J.A., & Bigaran, C. (2001). Knowledge-based fault diagnosis with adaptive probability update on recausticizing plant. *4th IFAC Workshop CHEMFAS*, Jejudo Island, Korea., 101-105.
- [12] Li, R.F., & Wang, X.Z. (2002). Dimension reduction of process dynamic trends using independent component analysis. *Computers & Chemical Engineering*, 26(3), 467-473.

- [13] Molga, E., & Cherbaski, R. (1999). Hybrid first-principle neural-network approach to modelling of the liquid-liquid reacting system. *Chemical Engineering Science*, 54(13-14), 2467-2473.
- [14] Montgomery, D.C. (1997). *Design and analysis of experiments*. Wiley, Chichester.
- [15] Oja, E., & Kohonen, T. (1999). *Kohonen maps*. Elsevier, Amsterdam.
- [16] Patton, R.J., Chen, J., & Nielsen, S.B. (1995). Model-based methods for fault diagnosis: some guide-lines. *Transactions of the Institute of Measurement and Control.*, 17(2), 73-83.
- [17] Porru, G., Aragonese, C., Baratti R., & Servida, A. (2000). Monitoring of a CO oxidation reactor through a grey model-based EKF observer. *Chemical Engineering Science*, 55(2),331-338.
- [18] Qi, H., Zhou, X.-G., Liu, L.-H., & Yuan, W.K. (1999). A hybrid neural network-first principles model for fixed-bed reactor. *Chemical Engineering Science*, 54(13-14), 2521-2526.
- [19] Sharif, M.A., & Grosvenor, R.I. (1998). Process plant condition monitoring and fault diagnosis. *Institution of Mechanical Engineers, Part E: Journal of Process Mechanical Eng.*, 212(E1), 13-30.
- [20] Su, H., Mc Avoy, T.J., & Werbos, P. (1992). Long-term predictions of chemical processes using recurrent neural networks: a parallel training approach. *Industrial and Engineering Chemistry Research*, 31, 1338-1352.
- [21] Venkatasubramanian, V. (2001). Process fault detection and diagnosis: past, present and future. *4th IFAC Workshop CHEMFAS.*, Jejudo Island, Korea, 3-15.
- [22] Venkatasubramanian, V., Rengaswamy, R., & Kavuri, S. N. (2003). A review

- of process fault detection and diagnosis: Part II: Qualitative models and search strategies. *Computers & Chemical Engineering*, 27(3), 313-326.
- [23] Venkatasubramanian, V., Rengaswamy, R., Kewen Y., & Kavuri, S. N. (2003a). A review of process fault detection and diagnosis: Part I: Quantitative model-based methods. *Computers & Chemical Engineering*, 27(3), 293-311.
- [24] Venkatasubramanian, V., Rengaswamy, R., Kewen Y., & Kavuri, S. N. (2003b). A review of process fault detection and diagnosis: Part III: Process history based methods. *Computers & Chemical Engineering*, 27(3), 327-346.
- [25] Wisnewski, P.A., Doyle III, F.J., & Kayihan, F. (1997). Fundamental continuous pulp digester model for simulation and control. *AIChE Journal*, 43, 3175-3192.
- [26] Wisnewski, P.A., & Doyle III, F.J. (1998). Control structure and model predictive control of the Weyerhaeuser digester problem. *Journal of Process Control*, 8, 487-495.
- [27] Wu, W., Walczak, B., Massart, D.L., Heuerding, S., Erni, F., Last, I.R., & Prebble, K.A. (1996). Artificial neural networks in classification of NIR spectral data: Design of the training set. *Chemometrics and Intelligent Laboratory Systems*, 33(1), 35-46.
- [28] Zhang, L., & Subbarayan, G. (2002). An evaluation of back-propagation neural networks for the optimal design of structural systems: Part I. Training procedures. *Computer Methods in Applied Mechanics and Engineering*, 191(25-26), 2873-2886.

Figure	Legend
1	Schematic of a chemical pulp digester
2	Moisture content (a) and cellulose density (b): normalized variations introduced (solid) and response of the first neural network (dashed) used in the validation subset.
3	Araboxylyan density (a) and HS density (b): normalized variations introduced (solid) and response of the first neural network (dashed) used in the validation subset.
4	Moisture content (a) and cellulose density (b): normalized variations introduced (solid) and response of the first neural network (dashed) used outside the data set.
5	HS density: normalized variations introduced (solid) and response of the second neural network (dashed) used outside the data set.
6	HS density: normalized variations introduced (solid) and response of the second neural network (dashed) used outside the data set.
7	HS density: normalized variations introduced (solid) and response of the second neural network (dashed) used outside the data set with an initial step increase in upper extract flowrate.
8	Faults introduced for the third neural network used outside the data set: (a) normalized chips flow rate, (b) cook temperature deviation.
9	Moisture: normalized variations introduced (solid) and response of the third neural network (dashed) used outside the data set.
10	HS: normalized variations periodically introduced (solid) and response of the third neural network (dashed) used outside the data set.
11	First and last fault of the 7 fault sequence combined to the HS periodic fault: moisture content (a) and EA density (b) for the third neural network used outside the data set.

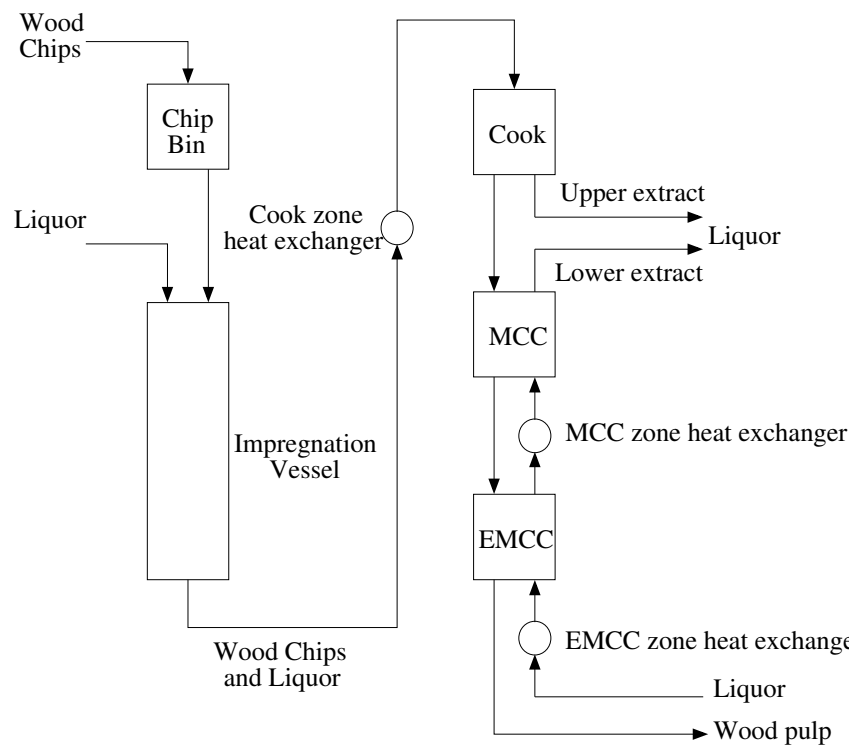


Fig. 1.

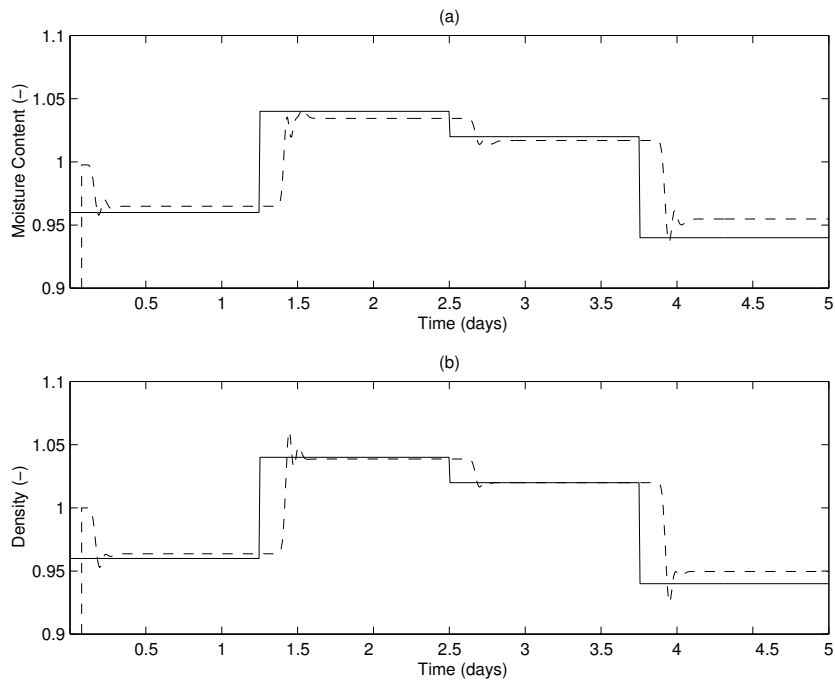


Fig. 2.

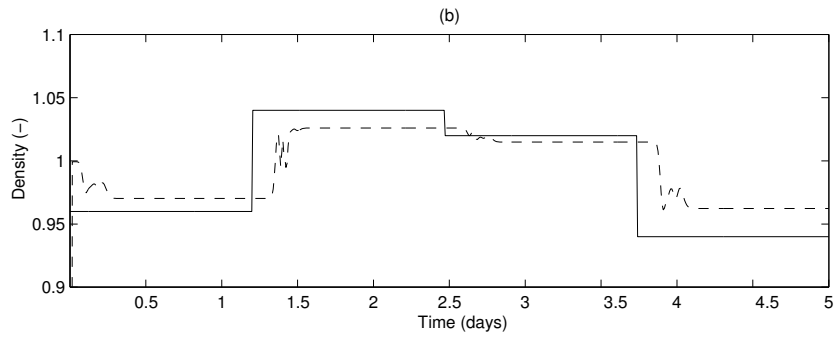
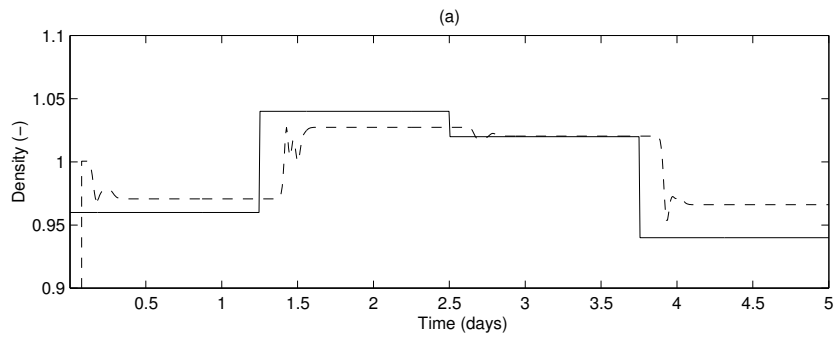


Fig. 3.

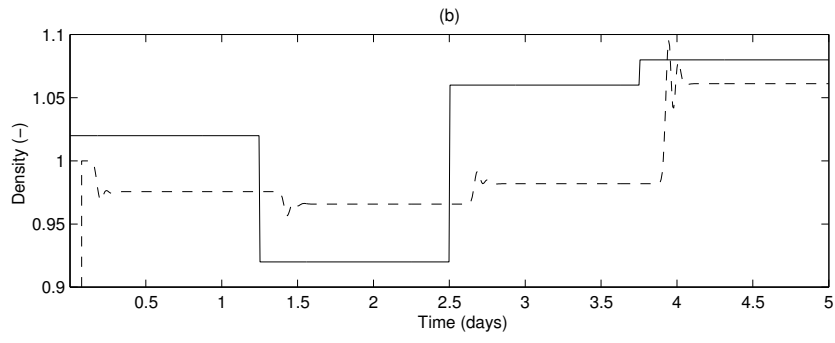
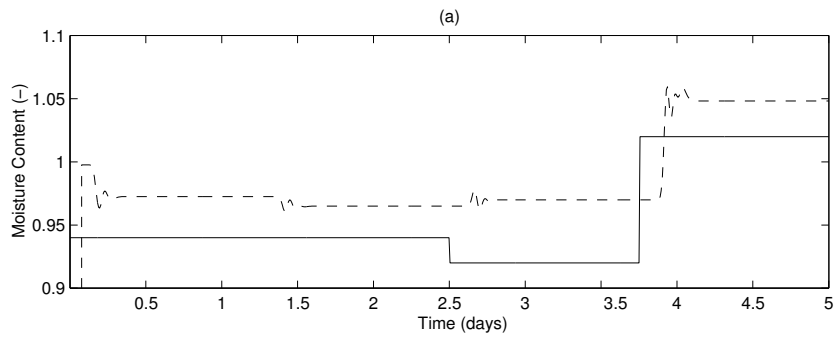


Fig. 4.

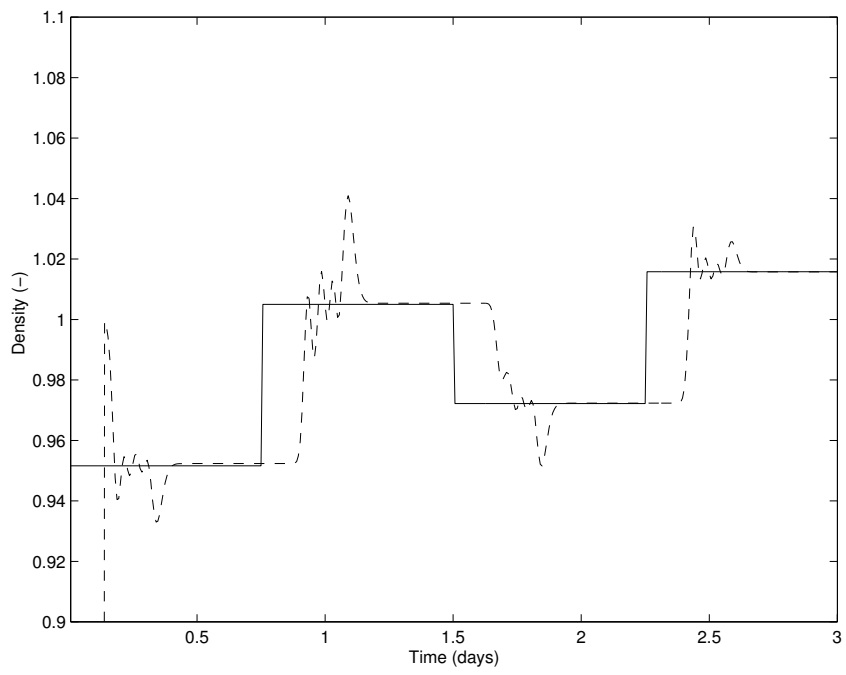


Fig. 5.

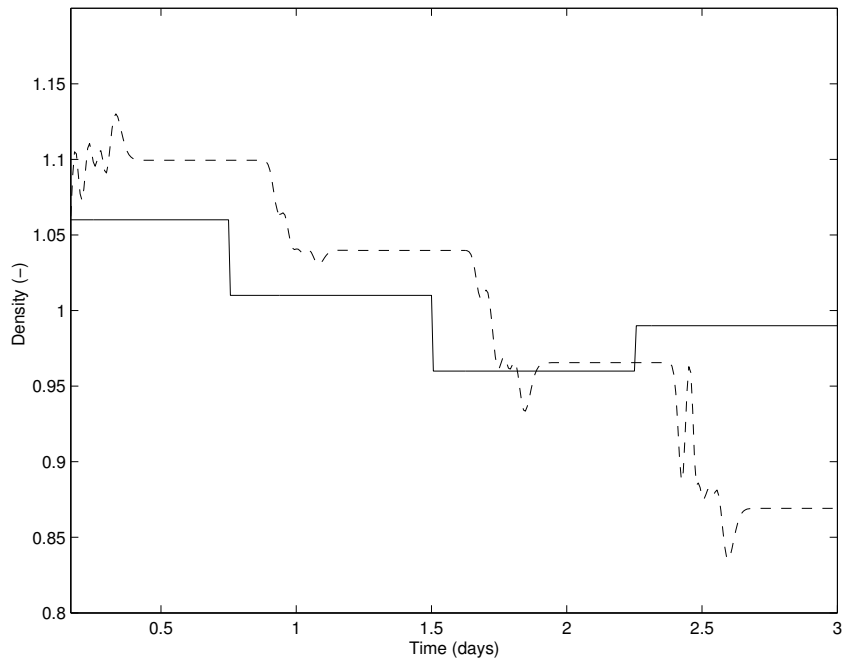


Fig. 6.

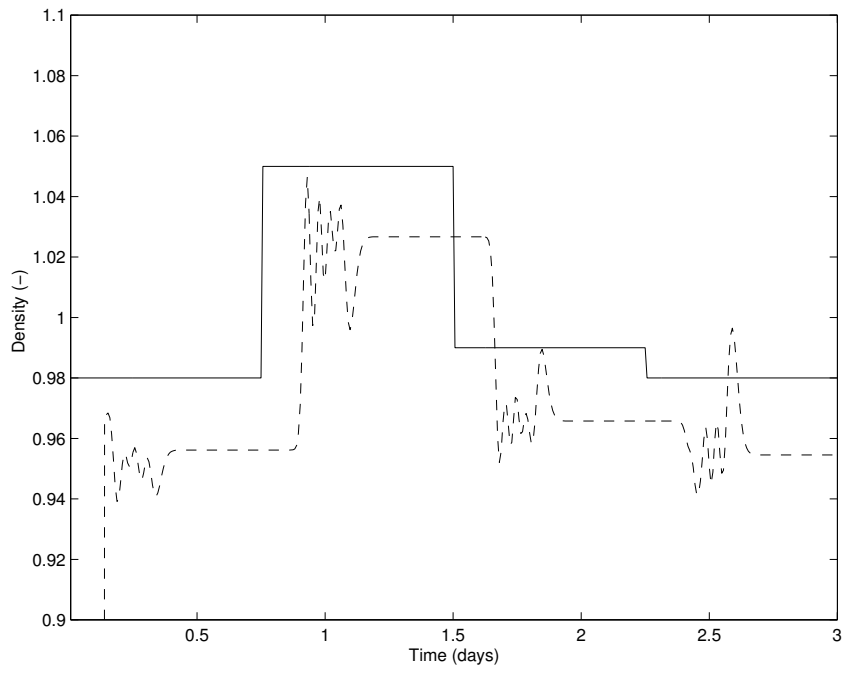


Fig. 7.

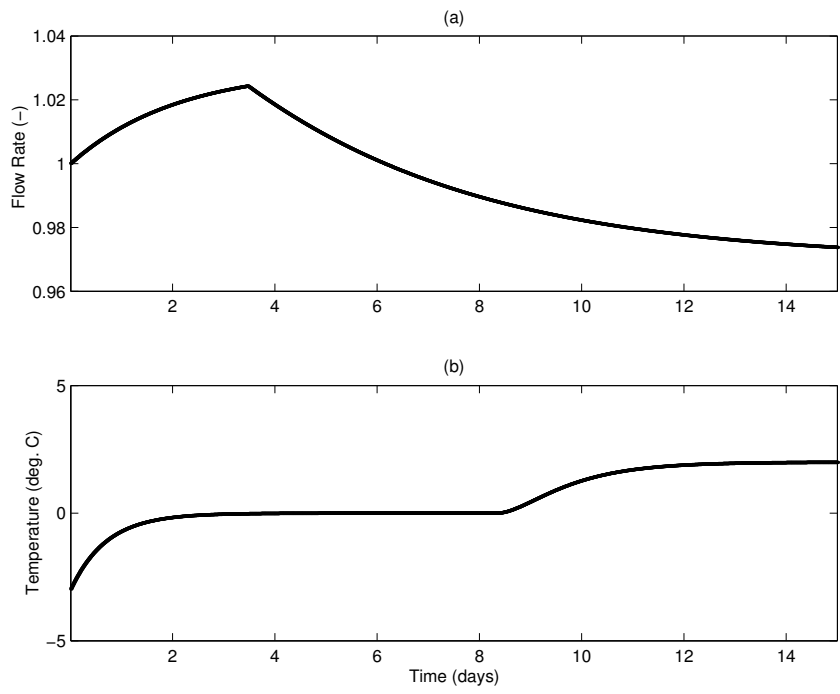


Fig. 8.

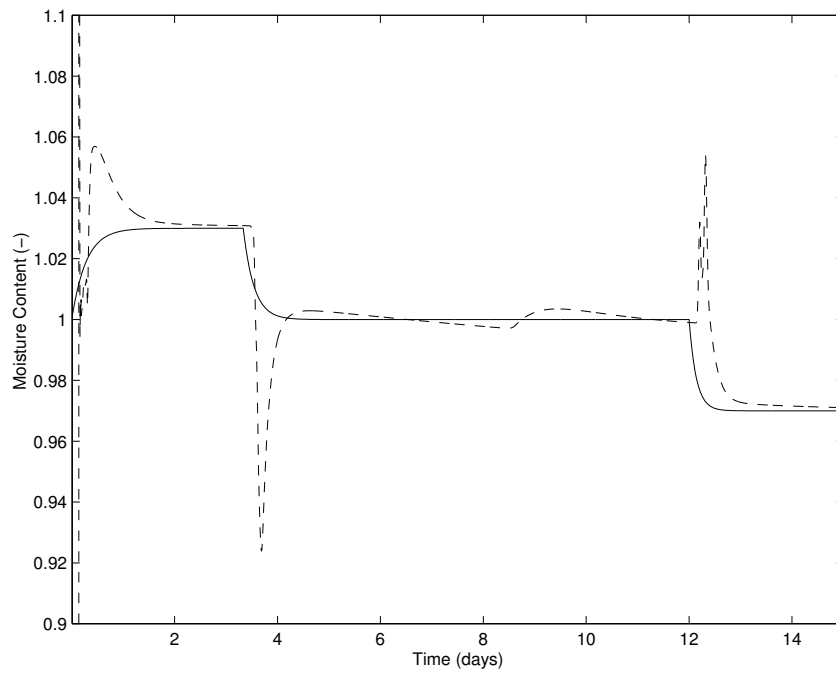


Fig. 9.

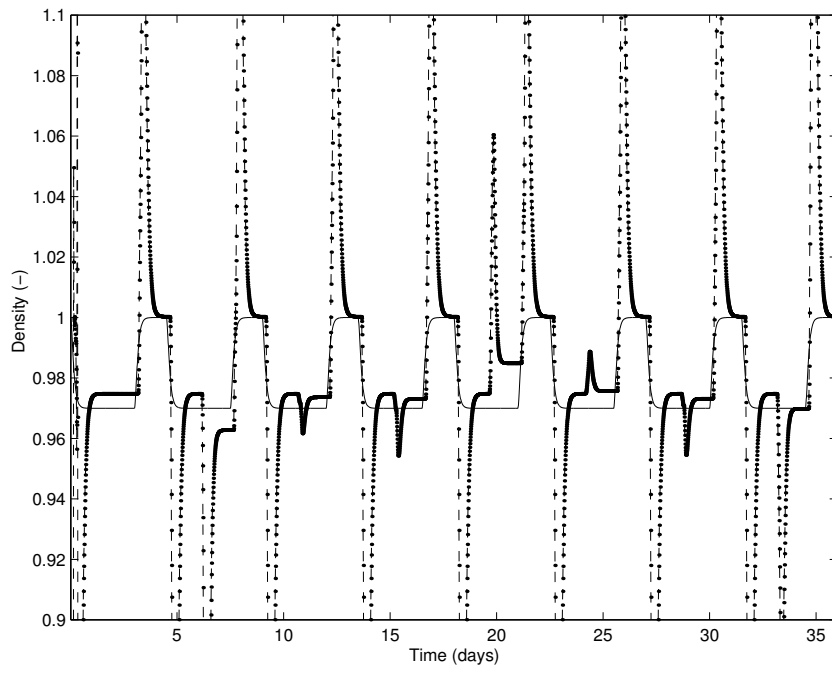


Fig. 10.

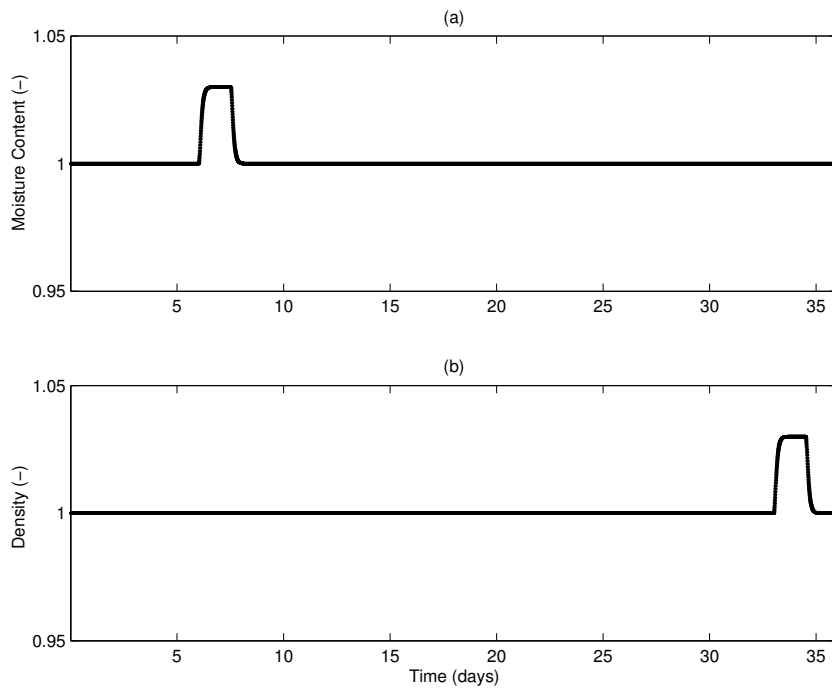


Fig. 11.