Comparison of Performance of Different Algorithms in Noisy Signals Filtering of Process in Enzymatic Hydrolysis of Cheese Whey

Gilson Alexandre Pinto, Ruy de Sousa Júnior[*] and Roberto de Campos Giordano
Departamento de Engenharia Química; Universidade Federal de São Carlos - UFSCar; Rodovia Washington Luís (SP-310), Km. 235; C.P. 676; 13565-905; São Carlos - SP - Brasil

ABSTRACT

This work presented the results of the implementation of an off-line smoothing algorithm in the monitoring system, for the partial hydrolysis of cheese whey proteins using enzymes, which used penalized least squares. Different algorithms for on-line signals filtering used by the control were also compared: artificial neural networks, moving average and smoothing algorithm.

Key words: Enzymatic proteolysis, pH control, post-treatment noisy data, on-line noisy signals filtering

INTRODUCTION

The enzymatic hydrolysis of the cheese whey is an interesting alternative to reuse the same as value-added product. Cheese whey, when discharged, presents high environmental hazard (BOD approximately 35000 mgO_2/L, while domestic sewage, as a comparison base, has BOD of 300 mgO_2/L, Viotto, 1993). On the other hand, due to protein characteristics, its controlled enzymatic hydrolysis (proteolysis), with the aid of enzymes immobilized in an inert support can either change or evidence functional properties of the produced peptides, increasing the potential applications of cheese whey. Derivatives of cheese whey have been applied to produce creamers for foaming beverages, edible food films, milk and salt substitutes, besides their use as supplementary food with low text of phenylalanine (used in the diet of phenylketonuria patients).

Sousa Jr. et al. (2004), presented a hybrid algorithm GMC-Fuzzy for pH control during hydrolysis (control of this is basic importance to modulate the MWD of product). Sousa Jr. et al. (2003), presented a supervisory system of this process that used a hybrid "kinetic-neural" model for on-line inference of MWD, in five bands of molecular weights previously determined.

In this scenario, considerable random noise was identified in pH and temperature signals used by pH control and hybrid "kinetic-neural" algorithms cited above. The previous filtering of such signals contributes to a better performance of both algorithms, besides promoting a better understanding of the process. The performance of different digital filters is compared in this work, based on artificial neural networks, moving average (Smith, 1999) and smoother, with discrete penalized least squares (Eilers, 2003). Such smoother is also employed for data post-treatment, implemented in the proper supervised system of
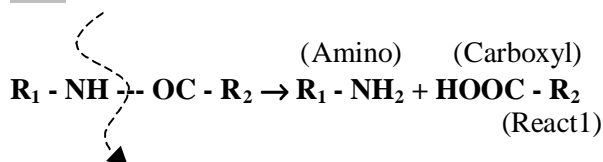
[*] Author for correspondence

the process and preventing, thus, exportation of data record process to other mathematical packages.

Enzymatic Proteolysis of Cheese Whey

Reactions React1, React2 and React3 summarize the proteolysis reactions:

H₂O

(Amino)   (Carboxyl)

$$R_1 - NH - OC - R_2 \rightarrow R_1 - NH_2 + HOOC - R_2 \tag{React1}$$

After hydrolysis, protonation and deprotonation of terminal amino and carboxyl groups occur:

$$R_1 \!-\! NH_2 + H^+ \leftrightarrow R_1 \!-\! NH_3^+ \tag{React2}$$

and

$$R_2 \!-\! COOH \leftrightarrow R_2 \!-\! COO^- + H^+ \tag{React3}$$

The optimum pH of immobilized enzyme here employed, Alcalase® NOVO, was between 9.0 and 10.0 (Sousa Jr., 2003). During the course of the reaction the pH value tends to decrease (React3 completely dislocates to the right, generating hydrogen ions that React2 does not absorb totally). As the result of chemical equilibrium of dissociated reactions, the concentration of H^+ ions in the mean increases affecting the enzymatic activity. The properties of hydrolysate depend on the degree of hydrolysis, which is influenced by the activity of protease, physical and chemical character of protein substrate and conditions of reactor (Adler-Nissen, 1986). Thus, the good performance of pH control algorithm is essential for the product quality. However, such performance is directly related to the quality of pH and temperature sampled signals during the enzymatic process.

Moving Average Filter

The moving average is the most common filter in digital signal processing, mainly because it is the easiest digital filter to understand and use (Smith, 1999). As the name implies, the moving average filter calculates the average of a number of input signals a, in a moving window, to produce an output signal y. For example, in a 3 point moving average filter (M = 3), point 80 in the output signal is given by Equation 1, where the input signals are chosen symmetrically around the output signal (symmetrical approach):

$$y(80) = \frac{a(79) + a(80) + a(81)}{3} \tag{1}$$

It's possible to verify that symmetrical approach implies in a late response for on-line applications.

Filters based on Neural Networks

A possible explanation for the good performance of neural networks (NNs) in the filtering of signals is based on the concept of "microfeature" (Baughman and Liu, 1995). According to this, no single node is directly responsible for associating a certain input with a certain output. In contrast, each node codifies a "microfeature" of input-output pattern. Each node affects the pattern slightly, thus minimizing the effects of noise in any given node. Other computational techniques do not add this concept. In most of the empirical models, for instance, each variable has a significant impact on the output.

Filters based on NNs have been developed for the sampled pH and temperature data. Fig. 1 illustrates the structure of the filters:

The sampled data a (pH or temperature) is the input of the first recurrent NN. A classical moving average filter with no symmetrical approach (Smith, 1999) uses the value inferred from this NN ($\varepsilon 1$). Then, the second recurrent NN infers the smoothed value that is used by the GMC-Fuzzy algorithm ($\varepsilon 2$). The recurrent NNs showed in Fig. 1 were trained off-line with real data (different of data of experimental assays, used to validation, but with the same noisy magnitude) using the Neural Network Toolbox of MATLAB 5.2 (MathWorks Inc., USA). The function TRAINLM, available in its toolbox, was network-training function.
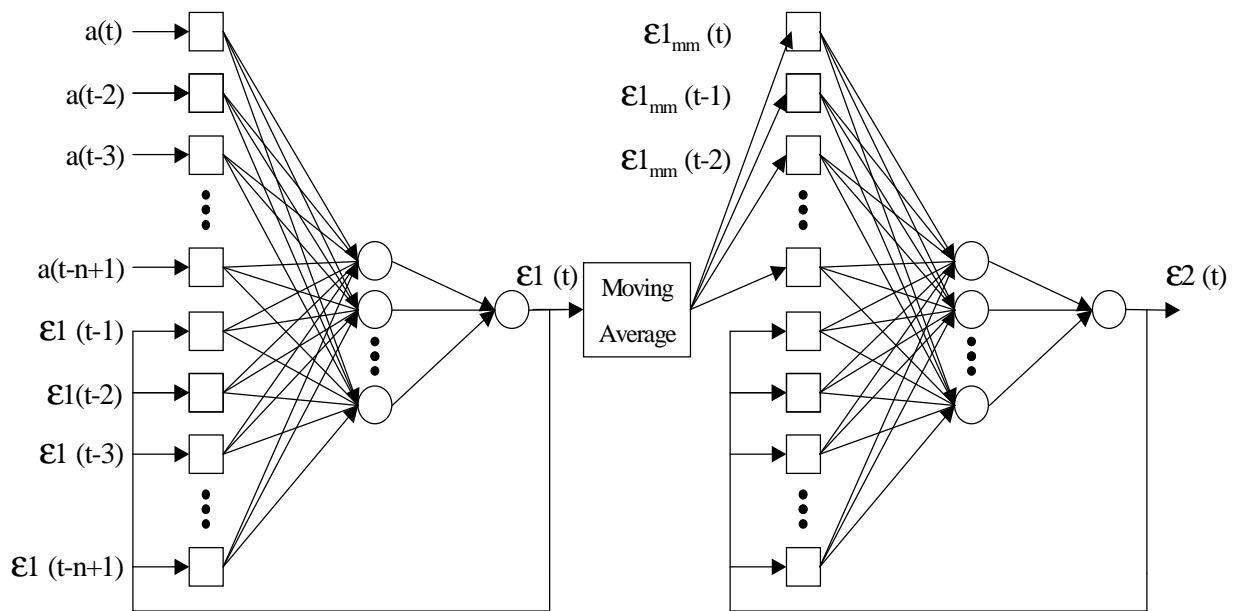
Figure 1 - Structure of filter based on NNs.

Smoother

The smoother here employed, proposed by Eilers (2003), searches to fit a smooth series z to a (sampled with a constant time acquisition). Two conflicting goals appear: (1) fidelity to the original data and (2) roughness. Roughness can be expressed, to a first order smoother, by the differences[•]: $\Delta z(t) = z(t) - z(t-1)$, providing a effective measure of the roughness R:

$$R = \sum_{t} \left[\Delta z(t) \right]^2 \qquad (2)$$

The lack of fit to original data S can be usually measured as the sum of squares of differences (Equation 3):

$$S = \sum_{t} \left[a(t) - z(t) \right]^2 \qquad (3)$$

The sum of Equations 2 and 3 adds the two contradictory objectives mentioned. Thus, we have a standard sum of squares problem whit

---

[•] to a second order smoother:

$$\Delta^2 z = \Delta(\Delta z) = z(t) - 2z(t-1) + z(t-2)$$

penalization, where the goal is to find series z that minimizes Q:

$$Q = S + \lambda R \qquad (4)$$

λ is a parameter chosen by the user which, how much bigger, bigger smoothness will induce to adjusted series z and, however, minor the fidelity to the original data set (there is the penalization concept).

Introducing matrices and vectors (Eilers, 2003), taking $Dz = \Delta z$, the linear system described by Equation 5 arises:

$$(I + \lambda D' D)z = a \qquad (5)$$

I is identity matrix; D' is first (first order smoother) or second derivative (second order smoother) of D.

MATERIALS AND METHODS

Experimental Assays

Cooperativa de Lacticínios São Carlos donated raw cheese whey, which was previously micro-(membrane 0.45 µm, A/G Technology

Corporation®) and ultra-filtered (membrane 10000 Da, Amicon®) up to the concentration of 52 g/L.
Experimental assays of enzymatic hydrolysis yielded five noisy pH and temperature data sets. The reaction was carried out in a 0.5L glass stirred vessel with temperature control (thermostatic bath Brookfield Ex.200). Continuous cheese whey feeding and product withdrawal were done through pumps Masterflex® 7519-10, while titration for pH control was done through ProMinent® Gamma G/4b. Measurements of pH and temperature were carried out using an Orion electrode (diameter 1.5 cm) and thermo-resistance Pt100, respectively. The data acquisition system was from TandS Equip. Eletr., São Carlos (with a constant sampled time of 0.75s).

Statistic Criteria of Performance

To performance analysis of the different digital filters were used the relative mean error $\bar{\mu}_R$ (Equations 6 and 7), and the variance of it σ_R^2 (Equation 8); important parameters to analysis and comparisons between digital filters (Smith, 1999):

$$\mu_R(t) = \left[\frac{\theta(t) - z(t)}{\theta(t)} \right] \qquad (6)$$

$$\bar{\mu}_R = \frac{1}{N} \sum_{i=0}^{N} \mu_R(t-i) \qquad (7)$$

$$\sigma_R^2 = \frac{\sum_{i=0}^{N} \left[\bar{\mu}_R - \mu_R(t-i) \right]^2}{N-1} \qquad (8)$$

N is total number of points in each assay, $z(t)$ is value filtered by one of the analyzed algorithms and $\theta(t)$ is theoretical value (or an "ideal" filter) yield by Fast Fourier Transform (FFT). FFT removes components with frequencies over $(1/(N_F * \Delta t))$, where N_F is the number of points considered in the previous and posterior time around the present data value and Δt the constant interval between two adjacent points.

RESULTS AND DISCUSSION

Implementation and Validation of Smoother

Term $(I + \lambda D'D)$ of Equation 5 yields a tri- or pent-diagonal matrix of order N to a smoother of first or second order, respectively. The algorithms from Press et al. (1996), were used to solve both the problems. Smoother algorithm was implemented using Microsoft Visual Basic 6.0; due to characteristic of the linear system (tri- or pent-diagonal N^{x}N matrix), its resolution and, therefore, the processing of the algorithm was relatively fast (using 1-GHz Pentium PC). Fig. 2 shows the "in-house" software interface of smoother algorithm inserted in the supervisory system of the process to five values of λ previously chosen by user for experimental data post-treatment.

Comparison between the Different Algorithms for On-line Filtering

Due to noisy characteristic of pH and temperature signals, filters were developed for on-line processing (as described, both signals were inputs to the pH control algorithm). Figs. 3 and 4 compared the on-line performance of the filters analyzed with regard to FFT for pH signals.
Figs. 3 and 4 showed the practically equivalent performances of the filter based on NNs and the first order smoother. Temperature data sets showed same behavior, being omitted for concision. To the smoother (first and second order), supply of noisy data set was done in a gradual form, that is, its temporal window grew during the assay. It was observed that the computational processing started to be harmed by N > 5000 points. In this case, a moving window would have to be used to compose the input data.
Fig. 5 presents the smoother response to one of temperature data sets (using a moving window of N = 3000 points). In this case good performance of the cited smoother for on-line filtering was observed.
Fig. 6 compares control action calculated by GMC-Fuzzy algorithm using original noisy data or previously on-line filtered (employing filters based on NNs to the pH and temperature signals). Both assays were realized in bench-scale enzymatic reactor over the same conditions: regulatory problem, pH = 9.5 (set-point) and controlled temperature of 50 ºC.
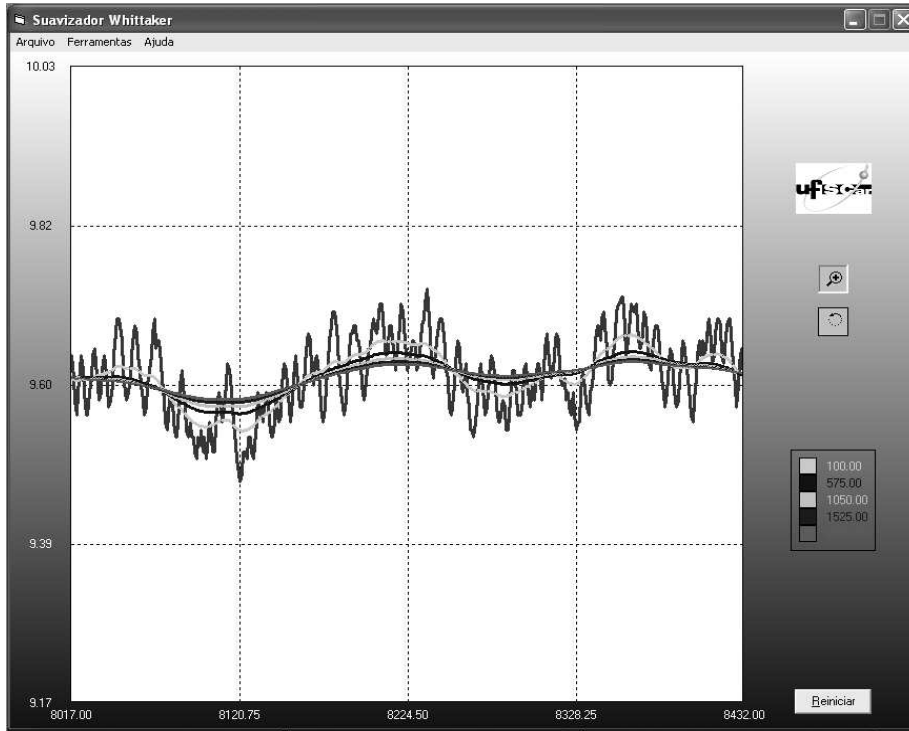
Figure 2 - Interface of smoother implemented in the "in-house" supervisory system.
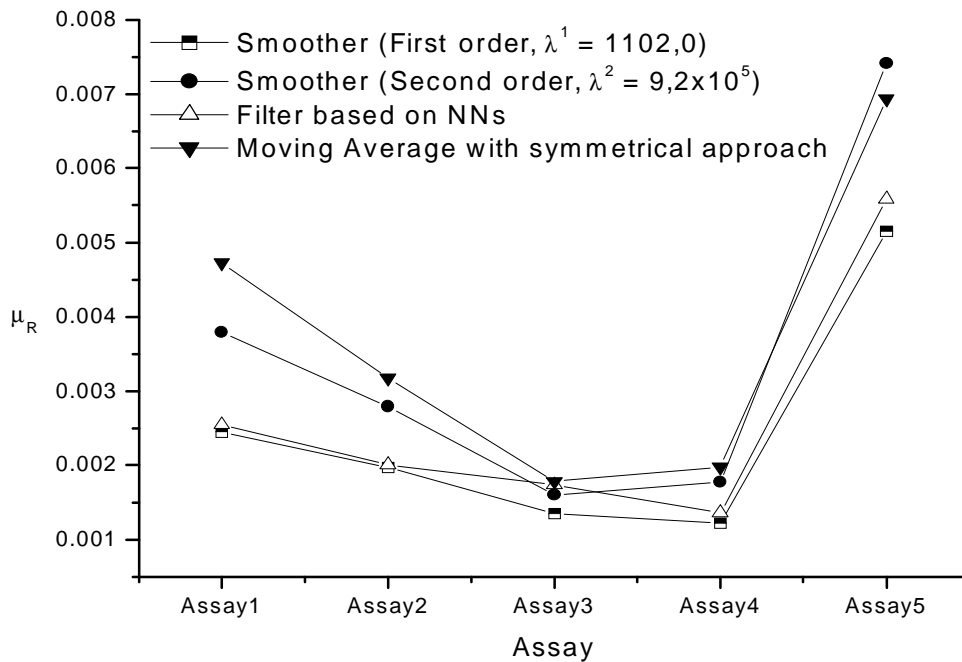

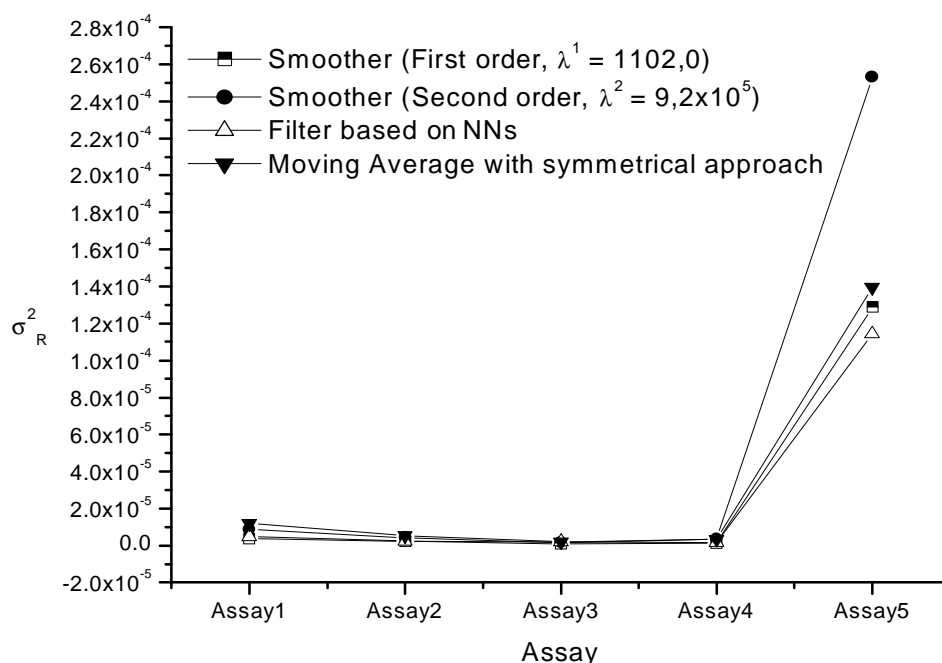**Figure 3 -** Relative mean error $\overline{\mu}_R$ (pH signals).

Figure 4 - Variance of relative mean error σ_R^2 (pH signals).

It was observed that trajectory of titration flow rate calculated by the control algorithm was smoothed when pH and temperature signals were previously filtered, i.e., the control action (titration addition, NaOH 1 N) became smoother.

Fig. 7 compared the values inferred by the kinetic-neural hybrid model when pH and temperature data were supplied to the model in the noisy form or earlier filtered by the algorithm based on NNs. There was no relevant difference between values inferred by the kinetic-neural model when the signals were noisy or previously filtered. This occurred because the hybrid model was trained for four pH values (7, 8, 9 and 10) and to infer reaction rates at intermediate pH values it interpolated (Sousa Jr. et al., 2003). This interpolation procedure for intermediate pH values reduced the influence of noise on the final inference results: if the noise intervened in the prediction of the model, the curves of distribution of peptides along time would be noisy (or at least they would diverge from the expected theoretical values), what was not observed. Nevertheless, the effect of the filters on the control action was noticeable.

The low sensitivity of the NNs state inference with respect to noise observed here was one of the desirable properties of MLP ("Multilayer Perceptron") networks used in the hybrid model (Nelles, 2001). Anyway, it was possible from these simulations to verify that previous filtering of the signals improved the GMC-Fuzzy performance (Fig. 6), without affecting the predictions of the state inference of the hybrid algorithm (Fig. 7).
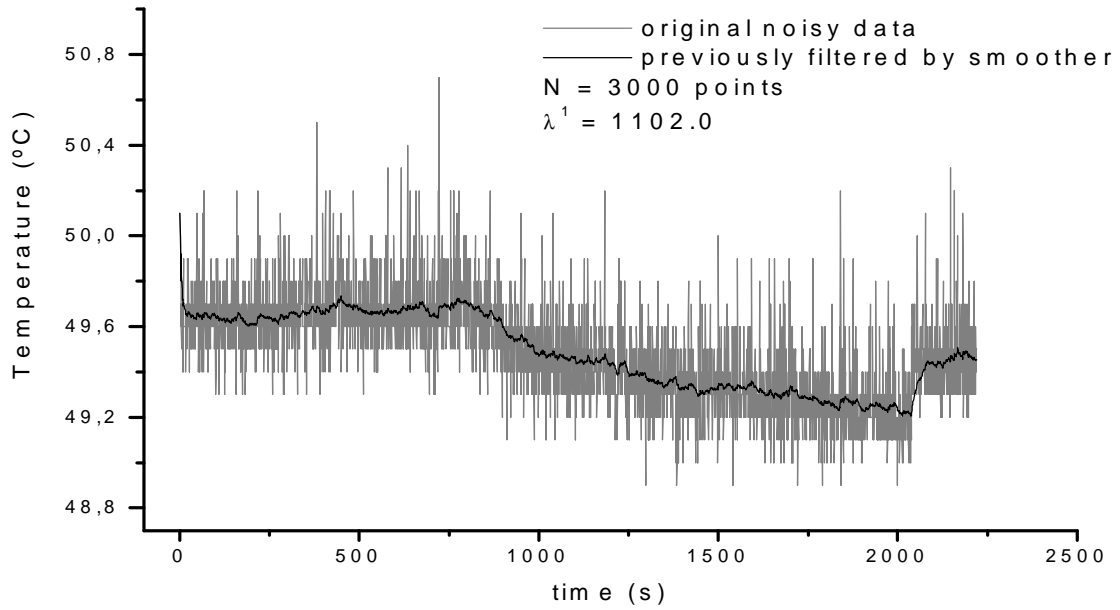
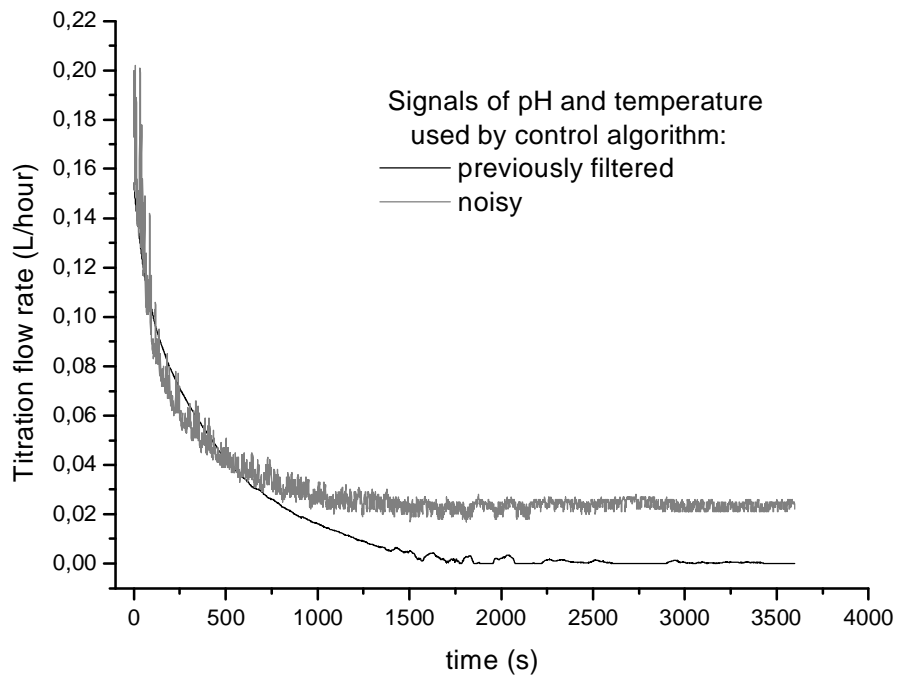**Figure 5 -** Temperature noisy signals and filtered by first order smoother in on-line application (Simulated data).

Figure 6 - Titration flow rates during regulatory pH control using either noisy or previously filtered signals (Experimental data).
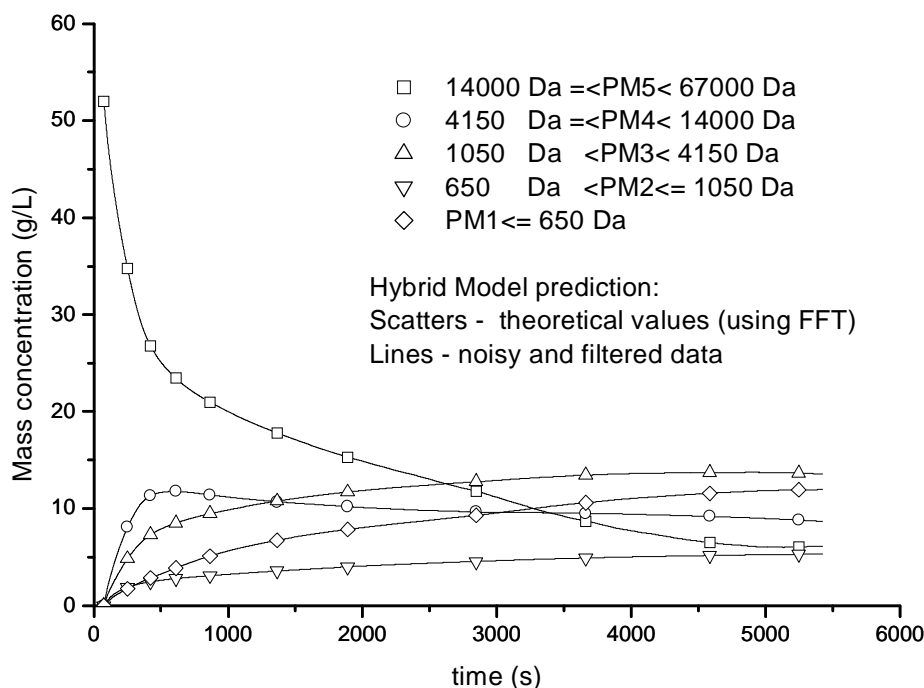
Pinto, G. A. et al.

Figure 7 - MWD along time using either noisy or previously filtered signal (To the
employed scale, the curves of the values inferred with noisy signals are
overlapping the ones that use filtered signals).

CONCLUSIONS

The smoother algorithm here implemented,
validated with the process in study, could be
extended to data post-treatment of other processes
(with constant intervals of sampling). The filters
based on NNs were implemented in the on-line pH
control system, promoting a smoother control
action and did not affecting the MWD inference.
First order smoother also provides good results for
on-line filtering with the considerable advantage
of require just one parameter defined by the user
(λ). Finally, in the case of the process here studied,
characterized by a fast dynamics, both filters fitted
the necessities of the process for on-line
applications: fast processing, composing its
temporal window only with previous and current
data (without delay in the response).

ACKNOWLEDGEMENTS

RESUMO

A hidrólise parcial de proteínas do soro de queijo,
realizada por enzimas imobilizadas em suporte
inerte, pode alterar ou evidenciar propriedades
funcionais dos polipeptídeos produzidos,
aumentando assim suas aplicações. O controle do
pH do reator de proteólise é de fundamental
importância para modular a distribuição de pesos
moleculares dos peptídeos formados. Os sinais de
pH e temperatura utilizados pelo algoritmo de
controle e inferência de estado podem estar
sujeitos a ruído considerável, tornando importante
sua filtragem. Apresentam-se aqui resultados da
implementação, no sistema de monitoramento do

processo, de algoritmo suavizador, que utiliza mínimos quadrados com penalização para o pós-tratamento dos dados. Compara-se ainda o desempenho de diferentes algoritmos na filtragem em tempo real dos sinais utilizados pelo sistema de controle, a saber: redes neurais artificiais, média móvel e o sobredito suavizador.

REFERENCES

Adler-Nissen, J. (1986), *Enzymatic hydrolysis of food protein.* London : Elsevier Applied Science Publishers.

Baughman, D. R. and Liu, Y. A. (1995), *Neural networks in bioprocessing and chemical engineering.* New York : Academic Press. pp. 03-16.

Eilers, P. H. C. (2003), A perfect smoother. *Anal. Chem.*, **75**, 3631-3636.

Nelles, O. (2001), *Nonlinear system identification.* New York : Springer-Verlag Berlin Heidelberg. pp. 246-263.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A. and Vetterling, W. T. (1996), *Numerical recipes in Fortran 90: the art of parallel scientific computing.* Cambridge : Cambridge University Press. pp. 1014-1042.

Smith, S. W. (1999), *The scientist and engineer's guide to digital signal processing.* San Diego : California Technical Publishing. pp. 277-284.

Sousa Jr., R. (2003), *Inteligência computacional aplicada à automação da hidrólise enzimática de soro de queijo em reator contínuo.* PhD Thesis, Universidade Federal de São Carlos, São Carlos-SP, Brazil.

Sousa Jr., R.; Lopes, G. P.; Pinto, G. A.; Almeida, P. I. F. and Giordano, R. C. (2004), GMC-Fuzzy control of pH during enzymatic hydrolysis of cheese whey proteins. *Comp. and Chem. Engineering*, **28** : (9), 1661-1672.

Sousa Jr., R.; Resende, M. M.; Giordano, R. L. C. and Giordano, R. C. (2003), Hibrid model for an enzymatic reactor: hydrolysis of cheese whey proteins by alcalase immobilized in agarose gel particles. *Applied Biochemistry and Biotechnology*, **105**, 413-422.

Viotto, W. H. (1993), *Ultrafiltração de soro doce de queijo minas frescal* - efeito de pré-tratamento do soro no desempenho da membrana de ultrafiltração e na composição e solubilidade do concentrado protéico de soro. PhD Thesis, Universidade Estadual de Campinas, Campinas-SP, Brazil.