Publishing House
AKAPIT

COMPUTER METHODS IN MATERIALS SCIENCE
Informatyka w Technologii Materiałów

Vol. 7, 2007, No. 2

FILTERING OF INDUSTRIAL DATA USING THE ARTIFICIAL NEURAL NETWORKS

ANDRZEJ STANISŁAWCZYK[1], JOLANTA TALAR[1], PIOTR JAROSZ[2], JAN KUSIAK[1]

[1] Wydział Inżynierii Metali i Informatyki Przemysłowej, Akademia Górniczo-Hutnicza w Krakowie
[2] Wydział Metali Nieżelaznych, Akademia Górniczo-Hutnicza w Krakowie

Abstract

The form of registered data (superimposed by the measurement noise) and the lack of some data (difficulties in measurement of some process parameters) cause the problems in modelling and control of many real industrial processes. The filtering process of the data with the imposed noise is a complex problem and it is difficult to find appropriate general filtering method, which gives the reliable results. Sometimes, the filtering procedure eliminates important information, and sometimes leaves the unnecessary noise. Therefore, the registered data are often not suitable for the further analysis and modelling of the process.

The main goal of the work is presentation of elaborated filtering procedure, which is able to eliminate these components of the output signals, which can not be predicted on the basis of the registered input signals.

Proposed filtering algorithm gathers the advantages of different techniques: adaptive filters, Fourier Transform Method and techniques based on the Artificial Neural Networks. The paper presents the idea of data filtering system and the results of filtering of the industrial data of the copper flash smelting process. The filtered data can be used to work out a control system based on the Artificial Neural Network.

Key words: filtering of the data, artificial neural networks, adaptive filters, Fourier transforms, modelling of the metallurgical processes

1. INTRODUCTION

The copper flash smelting process (Talar et al., 2005) is very complex and there are many significant parameters (about 30 input and 40 output parameters), which should be considered in its modelling and control, but many of these parameters can not be measured. The existing models of that process are either simplified or based on the FEM models (Donizak et al., 1999; 2002; 2003). These models are useless from the point of view of the control system, because of its low accuracy and a long computation time. The presented idea of the control system of the copper flash smelting process is based on the Artificial Neural Network approach (Talar et al., 2005; Jarosz et al., 2006), and the process variables were limited to 27 input parameters and 16 output process parameters. The input parameters characterise the properties of the charge materials (chemical composition, grain size of the concentrate, water content in the concentrate), oxygen content in the process blow, oil consumption, volume of oxygen per concentrate unit and others. The output parameters consist of the chemical compositions of exhaust gases, blister copper and slag. The registered data, which describe the copper flash smelting process, are very noisy. Therefore, the elaboration of the model must be preceded by a filtering process of registered training and testing data.

In the industrial conditions, the measurements of the parameters of the copper flash smelting process are registered in the different time intervals (from 1 second to 8 hours). Moreover, the data are noisy

 ISSN 1641-8581

because of: the errors of measuring devices, the human mistakes in recording of some data, the disturbances which might have been caused by the factors not registered in the database. Therefore, the obtaining of the complete and correct records of the data for the modelling of the process is crucial. Moreover, some of the output process parameters are dependent on the history of changes (trends) of input parameters. Some parameters can react to the changes of input parameters with delay, but the values of delays are unknown. That situation causes many problems with the gathering of the data, which are useful in modelling of the copper flash smelting process.

The paper presents the filtering results of one of the most important output parameters – the concentration of SO_2 in the output gases. All the 27 input parameters were taken into account in that model.

2. IDENTIFICATION OF TIME-DELAYS

Because of the complex nature of the copper flash smelting process, as well as its dynamics, the measurements of the Input/Output parameters are characterized by the time delays, which are difficult to identify. The time shifting of the measurement of the parameters which are important for the purpose of the process modelling, must be taken into account. Therefore, the cross-correlation analysis (Orfanidis, 1996) was performed to identify the values of the time lags of input parameters of the copper flash smelting process.

The cross-correlation is the measure of similarity of the analysed signals. If the signals are correlated to each other with some delay then the cross-correlation coefficient turns out to its maximum value for that delay. The relationship between the cross-correlation coefficient and time lags for the chosen process parameter is presented in figure 1. In case (a) the time lag between the measurement of the output signal (SO_2 concentration in exhaust gases) and the input parameter IN1 (concentrate flux), is equal about 6 minutes, while the case (b) (correlation between the SO_2 concentration in exhaust gases and the S concentration in copper concentrate) presents the correlation without the time lags. The identification of the time lags is necessary in construction the reliable filters. The time lags obtained as the result of cross-correlation analysis were taken into account in the further filtering using the adaptive filter method. The idea of the developed data filtering system of the copper flash smelting process is based on the following techniques: Adaptive Filters (AF), Artificial Neural Networks (ANN) and Fourier Transforms (FT).

3. ADAPTIVE FILTER METHOD

The main goal of the adaptive filtering is identification of the non-predictable components of the process output signals, which are the consequence of the non-measurable disturbances of the input signals. The adaptive filter is a system, which can adapt to the real conditions of the process. The output of the adaptive filter (AF_{OUT}) is compared to the process output signal (OUT) (as the result of measurable (IN) and non-measurable (IN*) process input signals) (figure 2a). Observed discrepancy of both outputs (process and filter (AF_{ERROR})) results in adjusting of the adaptive filter parameters in such a way, that discrepancy is minimized. In consequence, the

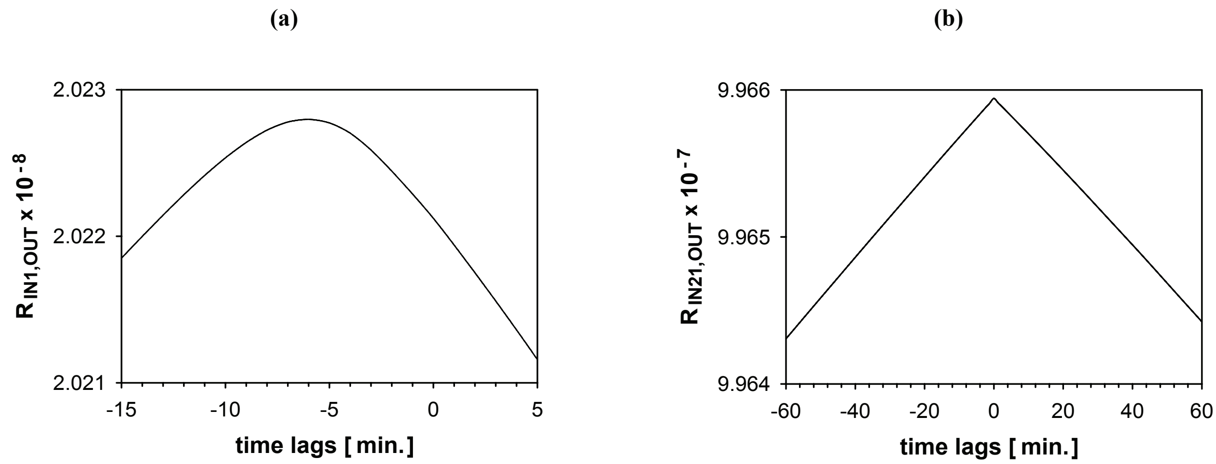

Fig. 1. Plot of the cross-correlation coefficient for sequence time lags. The connection between output parameter (SO_2 concentration in exhaust gases) and chosen input parameters: (a) – IN1 (concentrate flux), (b) – IN21 (S concentration in copper concentrate).

error contains only the part of change output signal dependent of change of unknown inputs signal.

The size of the tapped delay line of the input signals of the adaptive filter was chosen according to the identified time-delay values obtained by the cross-correlation analysis.

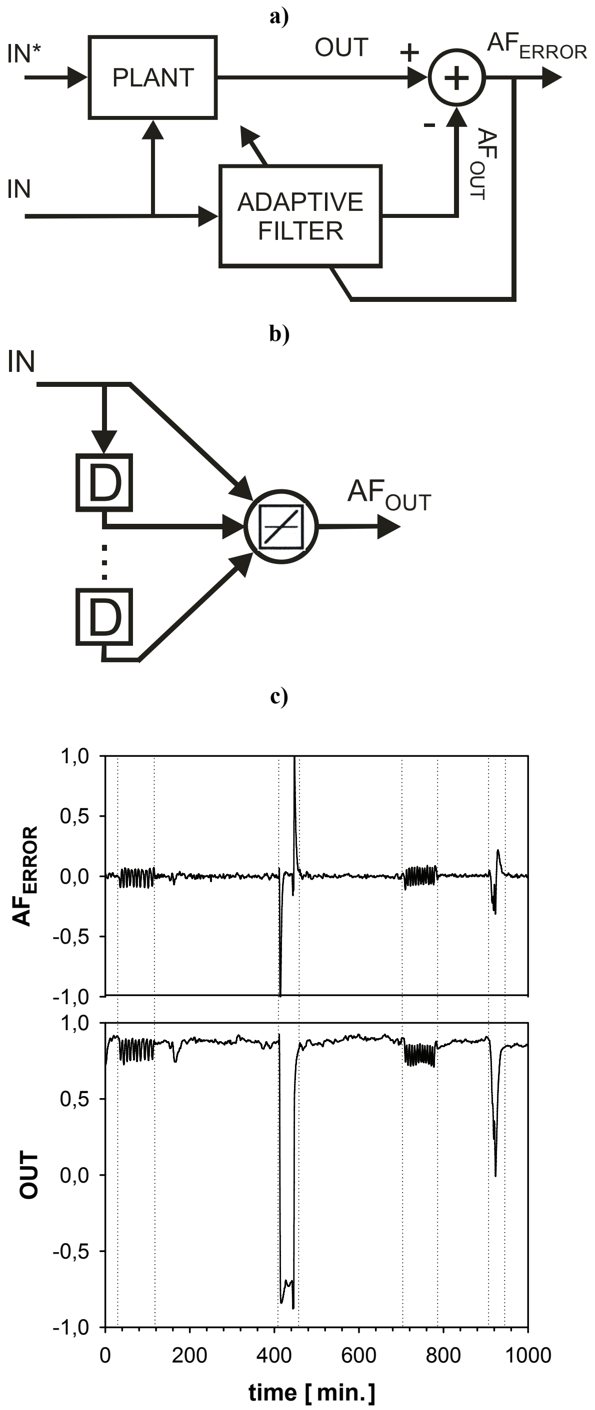

Fig. 2. The flow chart of the filtering system (detector of the unpredictable changes of the output signal) (a), the architecture of the dynamic linear artificial neural network used as the adaptive filter (D – the time-delay of a single time step; the size of the delay line is equal 10) (b); an example of chosen filtering results (c).

The dynamic linear artificial neural network was chosen as the adaptive filter. Linear network is adjusted at each time step based on new input and target vectors. It can find weights and biases that minimize the network's sum-squared error for recent input and target vectors, using the Widrow-Hoff learning algorithm (Widrow & Sterns, 1985). Networks of that kind are often used for noise cutting, signal processing, and control systems (Haykin, 2002).

4. FOURIER TRANSFORM METHOD

The adaptive filter is sensitive to the changes of the output signal which is not related to any changes of the input signals. Therefore, the short-term Fourier transforms were applied as the additional criterion in the filtering of the unpredictable signal components. The short-term Fourier transforms were obtained as the results of calculation of the Fourier transform of the AF_{ERROR} signal in the moving Hamming window. The reposition of the window was performed according to the given time step (20 minutes). The Fourier transform of the signal within the Hamming window is shown in figure 3. The power of the chosen harmonic components were calculated next. The obtained power values correspond to the time matching the centre of the window.

A discrete Fourier transform (DFT) Y of vector y (where y is E·W), computed with a fast Fourier transform (FFT) algorithm is expressed by the following equation:

$$Y(n) = \sum_{k=0}^{N-1} y(k) \cdot e^{-2\cdot\pi\cdot j\frac{k\cdot n}{N}}, \; 0 \le n \le N-1 \quad (1)$$

The frequency corresponding to the $Y(n)$-th is defined as:

$$f(n) = f_s \cdot \frac{n}{N} \quad (2)$$

The power spectrum (the value of the power at various frequencies) is equal:

$$P(f_n) = Y(n) \cdot \overline{Y(n)} \quad (3)$$

The power values between the centres of windows were calculated using the linear interpolation. The obtained Fourier analysis was used as the additional filtering criterion in the final structure of the filtering system. The increase of the power indicates, that there are oscillations within the signal. The power value correspond to the oscillations frequency.

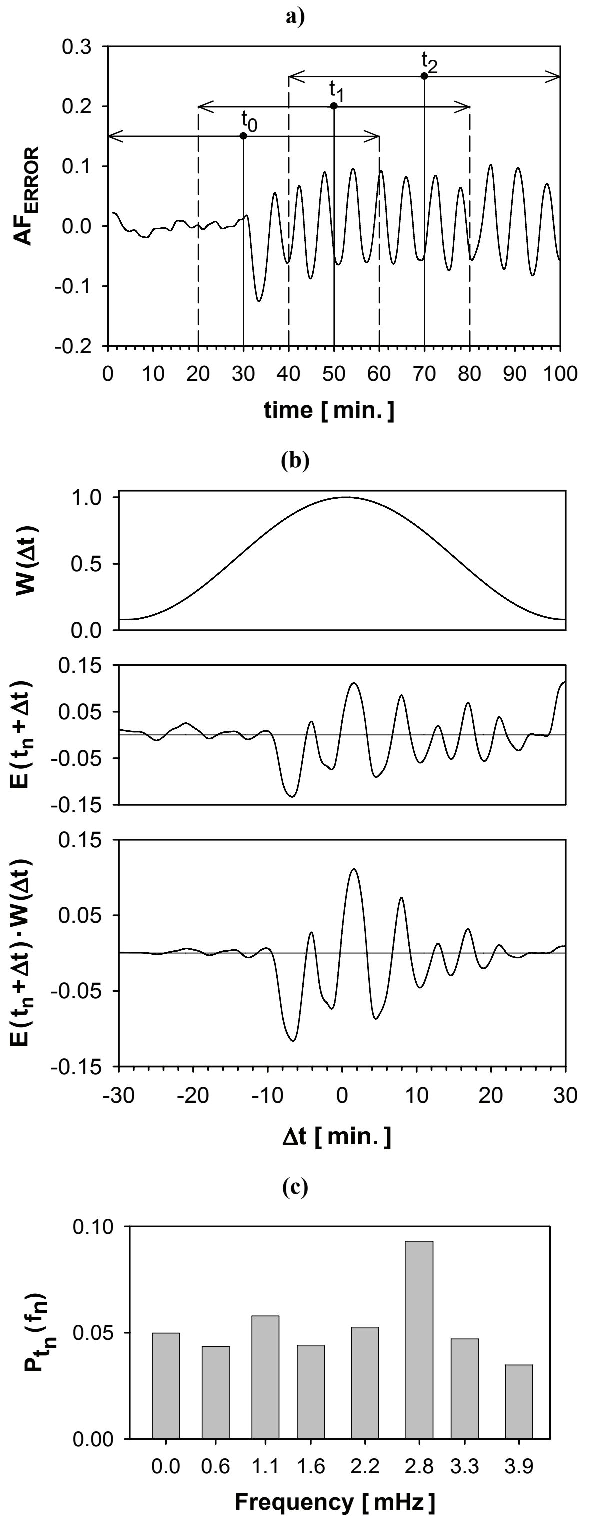

Fig. 3. The Fourier transform of the signal. Moving of the Hamming window along the analysed signal (t_0, t_1, t_2 – centres of the three first windows) (a); preparation of a single window to the FFT analysis (W – Hamming window function, E – AF_{ERROR} signal) (b); result of the Fourier analysis - power of the chosen harmonic product of signals E and W (c).

5. THE IDEA OF THE ANN BASE DATA FILTERING SYSTEM

The idea of the data filtering system is based on the Recurrent Artificial Neural Networks. The outline of the proposed filtering system is presented in figure 4. The inputs of the filtering system are the following three signals:

– the measurement of analysed output process parameter (OUT)
– the result of Adaptive Filter Method (AF_{ERROR})
– the result of Fourier Transform Method ($P(f_n)$ – for eight chosen frequencies).

The applied Recurrent Artificial Neural Network is shown in figure 4. Its architecture is 2-1 with two neurons in the input layer of sigmoidal transfer functions and one output neuron with the bipolar transfer function.

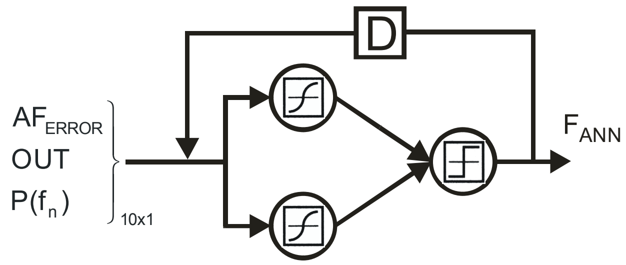

Fig. 4. The general idea of the Recurrent Artificial Neural Network data filtering system (D – time delay).

The ANN output signal F_{ANN} has two values: 1 or -1; where 1 correspond to the non-predictable input record, while -1 means the predictable one. On the base of these signals the artificial neural network selects the correct records from the analyzed data set and allows to eliminate incorrect records.

In the presented case of the validation of the proposed filtering system, the training and testing data set were generated on the basis of the AF_{ERROR} signal. The filtering system can detect the non-predictable and predictable records. Such analysis allows the pre-processing of the measured industrial data, which can be used next in the modelling of considered process.

6. RESULTS OF FILTERING

The elaborated filtering system was tested using the data set collected in the real industrial process of the copper flash smelting. The ANN training data set was a consistent data sequence of 4900 records. 19 clusters have been selected for the filtering. The data marked as non-predictable counted 1523 records (31% of the total number of records), while the predictable data counted 3377 records (69%).

The trained network recognized correctly 98% of all training records and 2% incorrectly. The half of incorrect network answers indicated non-predictable records and the second half pointed erroneously predictable records. It should be mentioned, that all clusters, predictable and non-predictable, were well recognized, and the observed errors are the result of the way the training data set was chosen – the width of the data range was chosen by chance.

The ANN testing data set was a consistent data sequence of 10000 records divided into 36 non-predictable clusters of 3052 records. The network pointed incorrectly only 112 non-predictable records as predictable, while 423 times the predictable data were recognized as non-predictable.

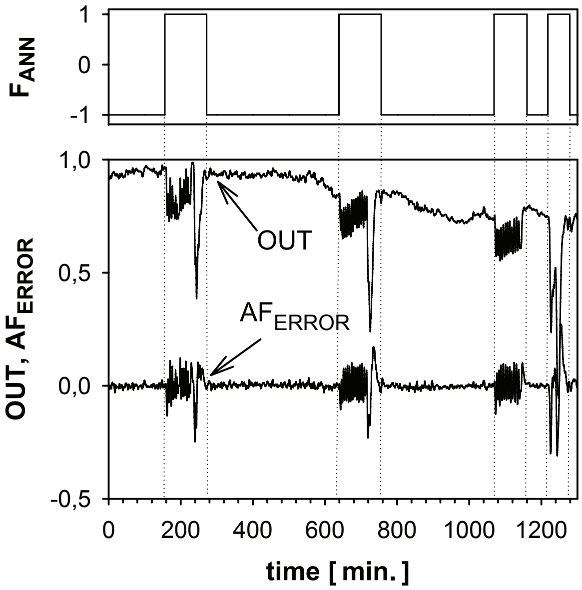

Fig. 5. The results of the ANN filtering of the concentration of the SO_2 in the exhaust gases of the flash smelting process.

The results of the ANN filtering of the testing data of the concentration of SO_2 in the exhaust gases of the flash smelting process are presented in figure 5. The lower part of the figure 5 shows the registered signal of the concentration of SO_2 and the AF_{ERROR} obtained from the adaptive filter. The output signal of the ANN filter is shown in upper part of the figure 5. The values 1 point these records which correspond to non-predictable input records. Identified in that way non-predictable records can be easily eliminated and the resulting filtered data set can be used in the further processing.

7. SUMMARY

Modelling and control of industrial processes is often very difficult because of the measured and registered signals are disturbed. Therefore, the additional pre-processing of these signals can eliminate the erroneous data, and to give the data set which can become in its new form very useful in further processing. The presented idea of the filtering system is based on the Recurrent Artificial Neural Network. The main objective of the filtering system is to eliminate the incorrect records of registered data of industrial process. The filtering system consists of adaptive filter, Fourier transform analysis and the Recurrent Artificial Neural Network. The output signal of the whole filtering system allows the elimination of the erroneous records of the measured data set. Presented results of the filtering of the SO_2 in the exhaust gases of the flash smelting process show, that elaborated filtering system, based on the Artificial Neural Network approach, is reliable in filtering of industrial disturbed data. The obtained filtered data can be used in the modelling of the analyzed process. The presented idea of the filtering technique can be applied in other complex problems connected with the industrial data, disturbed by the non-measurable factors.

Acknowledgements. The financial support of the Polish Science Foundation, project No 3 T08B 034 30 is acknowledged.

REFERENCES

Davenport, W.G., Jones, D.M., King, M.J., Partelpoeg, E.H., 2001. *Flash Smelting: Analysis, Control and Optimization*, TMS, Warrendale.

Donizak, J., Hołda, A., Kolenda, Z., 2002, Zastosowanie deterministyczno-stochastycznego modelu do symulacji zawiesinowego procesu produkcji miedzi hutniczej, *Proc. Conf. XVIII Zjazd Termodynamików*, 325-334 (in Polish).

Donizak, J., Hołda, A., Kolenda, Z., 2003, Modelowanie matematyczne procesów metalurgicznych-jednostadialny proces zawiesinowy produkcji miedzi hutniczej, *Proc. Conf. Teoria I Inżynieria Procesów Metalurgicznych*, Kraków, 23-42 (in Polish).

Donizak, J., Hołda, A., Kolenda, Z., Sukiennik, M., Warmuz, M., Szwancyber, G., Garbacki, J., Gostyński, Z., 1999, The Flash Smelting Process Digital Simulation, *Proc. IVth Int. Conf. on Non-Ferrous Metals and Alloys*, Kraków, 1999, 29.

Haykin, S., 2002, *Adaptive Filter Theory*, Prentice-Hall, New Jersey.

Jarosz, P., Talar, J., Kondek, T., Kusiak, J., Dobrzański, J., Staszak, J., 2006, Modelowanie składu fazy gazowej procesu zawiesinowego wytopu miedzi z wykorzystaniem sztucznych sieci neuronowych (Neural network modelling of exhausts composition of a copper flash smetting process), *Proc. Conf. KomPlasTech2006*, eds., Szeliga, D., Pietrzyk, M., Kusiak, J., Szczawnica, Akapit, 201-208 (in Polish).

Orfanidis, S.J., 1996, *Optimum Signal Processing. An Introduction. 2nd Edition*, Prentice-Hall, Englewood Cliffs NJ.

Talar, J., Kondek, T., Jarosz, P., Kusiak, J., Dobrzański, J., Staszak, J., Byszyński, L., 2005, Artificial intelligence control system of the copper flash smelting process, *Proc. Conf. 11th IFSC: International Flash Smelting Congress*, Bulgaria & Spain, 1–11.

Widrow, B., Sterns S.D., 1985, *Adaptive Signal Processing*, Prentice-Hall, New York.

FILTROWANIE PRZEMYSŁOWYCH DANYCH PRZY WYKORZYSTANIU SZTUCZNYCH SIECI NEURONOWYCH

Streszczenie

W procesie modelowania obiektów przemysłowych często napotyka się na trudności związane ze specyfikacją i sposobem akwizycji danych opisujących rzeczywisty obiekt. Rejestrowane dane są zaszumione oraz często niekompletne. Również nie wszystkie istotne parametry mogą być rejestrowane.

Filtrowanie danych z nakładającymi się szumami oraz zakłóceniami jest bardzo złożonym problemem, a znalezienie uniwersalnej metody filtrowania, która dawałaby wiarygodne rezultaty jest zadaniem skomplikowanym. Czasem procedury filtrowania eliminują istotne informacje, innym razem pozostawiają zbędny szum. Dlatego też, rejestrowane dane są często nieużyteczne dla dalszej analizy i modelowania procesu.

Celem pracy jest prezentacja opracowanej procedury filtrowania, która pozwala na eliminowanie tych komponentów sygnałów wyjściowych, które nie mogą być przewidywane na podstawie rejestrowanych sygnałów wejściowych.

Zaproponowany algorytm filtrowania danych wykorzystuje zalety różnych technik: filtrów adaptacyjnych, metody transformaty Fouriera oraz technik opartych o sztuczne sieci neuronowe. Opisano metodę filtrowania danych przemysłowych oraz zaprezentowano zastosowanie jej do filtrowania danych pochodzących z pieca do zawiesinowego wytopu miedzi.

Submitted: September 26, 2006
Submitted in a revised form: October 17, 2006
Accepted: December 4, 2006