Optimal experimental design and artificial neural networks applied to the photochemically enhanced Fenton reaction


* Lehrstuhl für Umweltmesstechnik, Engler-Bunte-Institut, Universität Karlsruhe, 76128 Karlsruhe, Germany
** Process Control and Simulation Laboratory, Department of Chemical Engineering, University of São Paulo, CEP 05508-900 São Paulo, Brazil

Abstract Among advanced oxidation processes (AOPs), the photochemically enhanced Fenton reaction may be considered as one of the most efficient for the degradation of contaminants in industrial wastewater. This process involves a series of complex reactions. Therefore, an empirical model based on artificial neural networks has been developed for fitting the experimental data obtained in a laboratory batch reactor for the degradation of 2,4-dimethyl aniline (2,4-xylidine), chosen as a model pollutant. The model describes the evolution of the pollutant concentration during irradiation time as a function of the process conditions. It has been used for simulating the behavior of the reaction system in sensitivity studies aimed at optimizing the amounts of reactants employed in the process, an iron(III) salt and hydrogen peroxide, as well as the temperature. The results show that the process is most sensitive to the concentration of iron(III) salt and temperature, whereas the concentration of hydrogen peroxide has a minor effect.

Keywords Advanced oxidation processes; 2,4-dimethyl aniline; experimental design; Fenton reaction (photochemically enhanced); neural networks

Introduction

Toxic contaminants in wastewaters from chemical, pharmaceutical and dye industries have to be eliminated, but, depending on the nature of the pollutant and on the level of contamination, detoxification might be difficult and/or expensive to achieve by conventional methods, e.g. adsorption on activated carbon. Photochemical degradation processes, also referred to as advanced oxidation processes (AOPs) have become increasingly popular in recent years as an alternative or complementary treatment (Legrini et al., 1993). The Fenton reaction, and especially the photochemically enhanced Fenton reaction, are considered most promising for the remediation of wastewaters containing a variety of toxic compounds (Bishop et al., 1968; Feuerstein et al., 1981; Barbeni et al., 1987; Lipcyynska-Kochany, 1991; Pignatello and Chapa, 1994). The successful treatment of industrial wastewater, contaminated with xylidines (dimethyl anilines), by such a process on a large pilot scale has already been reported (Oliveros et al., 1997).

Advanced oxidation processes, and particularly the photochemically enhanced Fenton reaction, involve complex reaction systems. The mechanisms of oxidation by the Fenton reagent (a mixture of hydrogen peroxide and iron(II) salt) are still under controversial discussion. Although until recently hydroxyl radicals (HO\(^{-}\)) were considered as the primary oxidizing species (reaction 1) (Walling, 1975; Bishop et al., 1968; Feuerstein et al., 1981; Barbeni et al., 1987; Lipcyynska-Kochany, 1991; Pignatello and Chapa, 1994), other metal-based oxidant intermediates, such as the hydrated higher-valent iron species (Fe(IV) ferryl species, reaction 2), are most likely involved (Wink et al., 1994; Bossmann et al., 1998; Pignatello et al., 1999).

\[
\begin{align*}
\text{Fe}^{2+} + \text{H}_2\text{O}_2 & \rightarrow \text{Fe}^{3+} + \text{HO}^- + \text{HO}^- \\
\text{Fe}^{2+} + \text{H}_2\text{O}_2 & \rightarrow \{\text{Fe}^{IV} = \text{O}\} \rightarrow \text{Fe}^{3+} + \text{HO}^- + \text{HO}^- 
\end{align*}
\]
Moreover, enhancement of the Fenton reaction under UV/visible irradiation implies a series of photochemical reactions leading to a more efficient recycling of Fe^{2+} (reduction of Fe^{3+}) than in the thermal process (see, e.g., Bossmann et al., 1998). It should be noted that the photochemically enhanced Fenton reactions may be carried out using iron(II) or iron(III) as initial reactant.

Due to the complexity of these reaction systems, the kinetic parameters of the various steps involved are very difficult to determine or even estimate, leading to uncertainties in the design and scale-up of chemical reactors of industrial interest. Under these conditions, an empirical approach based on artificial neural networks for modeling the process might prove particularly useful (Nascimento et al., 1994; Oliveros et al., 1998; Göb et al., 1999). Indeed, neural networks do not require the mathematical description of the phenomena involved in the process. A representative set of experimental data may be obtained by planning the experimental work according to the experimental design methodology (Box et al., 1978). This methodology provides a means of building a statistically significant model of a phenomenon by performing a minimum set of experiments adequately distributed in the experimental region. This approach was adopted in the investigation of the degradation of 2,4-xylidine by the photochemical enhanced Fenton reaction. An important objective was to obtain a model of the process that could make reliable predictions for the optimization of the amounts of reactants employed in the process, iron(III) sulfate and hydrogen peroxide, as well as the reaction temperature.

**Experimental**

Experiments were carried out using 2,4-xylidine (Xyl) (Sigma Aldrich) purified by distillation, hydrogen peroxide (30% w/w H_{2}O_{2} in water, extra pure medical, Merck) and iron(III) sulfate (Fe_{2}(SO_{4})_{3}•5H_{2}O, > 97%, Sigma Aldrich). The concentration of hydrogen peroxide in the commercial solution was approx. 8 mol/L (Merck). Solutions were prepared with ultra pure water (Elgastat UHQ II).

The photochemical set-up used in this work has been previously described in detail (Bossmann et al., 1998). Briefly, it consisted of a 2 L thermoregulated reservoir and a flow-through annular photochemical reactor equipped with a TQ 718 medium pressure mercury arc (Heraeus Noblelight). The solution (total volume of 2.5 L) was pumped at 10 ± 1 L/min. under continuous jet-injection of compressed air.

The processing conditions were the same in all experiments. The initial concentration of 2,4-xylidine was 500 mg C/L or ppm C (5.2 \times 10^{-3} mol/L). The pH was set to an initial value of 3 by adding sulfuric acid. The iron(III) salt was then introduced at a concentration ranging from 0.8 to 2.6 \times 10^{-3} mol/L. The temperature was set to 25°C, 35°C or 45°C. H_{2}O_{2} was introduced continuously during 120 minutes up to the chosen concentration value (2 to 8 \times 10^{-2} mol/L). Samples of the reaction mixture were taken at regular time intervals during irradiation and were immediately treated by adding a “reduction and precipitation agent” (Bossmann et al., 1998), in order to ensure complete reduction of the residual H_{2}O_{2}. Samples were then filtered. The 2,4-xylidine concentration was monitored by HPLC (HP Series II 1090, LiChrospher-100 RP18), using a mixture of acetonitrile and triethylamine as an isocratic eluent. From the replications, the standard error for [Xyl], was estimated to be ± 20 mg C/L.

**Methods**

**Experimental design**

The experimental design methodology is based on multivariate methods, where the levels (settings or values) of the different variables are simultaneously modified from one experiment to another (Box et al., 1978). A number of classical experimental designs
(experimental matrices) adapted to different types of problems are available. For easier data treatment and analysis, the levels of the variables are standardized or normalized. Usually, a reduced and centered variable $X_i$ is associated to each natural variable $U_i$:

$$X_i = \frac{(U_i - U_{i0})}{\Delta U_i}$$

(3)

where $U_{i0}$ is the value of $U_i$ at the centre of the experimental region ($= (U_{i, \text{max}} + U_{i, \text{min}})/2$) and $U_i$ the step ($= (U_{i, \text{max}} - U_{i, \text{min}})/2$).

Experimental data used in this work for neural network modeling were obtained by performing the series of designed experiments of a three-dimensional Doehlert uniform array. The effects of three key variables of the photochemically enhanced Fenton reaction, the concentrations of Fe$^{3+}$ and H$_2$O$_2$ as well as the temperature, were studied. The 13 experiments to be performed in the case of a Doehlert uniform array for three variables are uniformly distributed and, in coded variables, may be represented by the apexes and the centre of a cube-octahedron. The concentrations of H$_2$O$_2$ and Fe$^{3+}$ used for the different experiments and the temperature values are listed in Table 1. Additional control experiments were carried out for the validation of the system.

### Artificial neural networks

Artificial neural networks have been attracting great interest during the last decade as predictive models as well as for pattern recognition (Rumelhart et al., 1986; Zupan and Gasteiger, 1993). The potential for employing neural networks in data treatment is especially high in the case of systems presenting non-linearities and complex behavior. Indeed, neural networks possess the ability to “learn” from a set of experimental data, (e.g. processing conditions and corresponding experimental responses, without actual knowledge of the physical and chemical laws that govern the system (supervised learning).

The most commonly employed artificial neural networks for modeling steady-state systems are “feed-forward” networks. In general, the network consists of processing units or neurons and information flow channels between the neurons, usually denominated interconnects. Neurons are distributed in three layers (input, hidden and output layers). The neurons in the input layer store the normalized input variables. Each processing neuron in the hidden and output layers first calculates the weighed sum of all interconnected signals from the previous layer ($S_j$, Equation 4) and then generates an output ($O_j = f(S_j)$) through its activation function. This function may assume different forms but is commonly expressed as a sigmoid function.

<table>
<thead>
<tr>
<th>Temperature(°C)</th>
<th>0.2</th>
<th>0.8</th>
<th>1.4</th>
<th>2.0</th>
<th>2.6</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>50</td>
<td>50</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28.4</td>
<td>80</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31.6</td>
<td>20</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>38.4</td>
<td>80</td>
<td>80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>41.6</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>50</td>
<td>50</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
where \( j \) refers to the \( j \)-neuron of the hidden layer, \( x_i \) is the normalized input variable \( i \), \( W_{i,j} \) is the weight between the \( i \)-neuron of the input layer and the \( j \)-neuron of the hidden layer.

The neurons of the output layer perform the same operations as the neurons of the hidden layer, the outputs from the hidden layer \((O_j)\) representing the input signals to the output layer. The network output consists of one value (or more generally of a series of values), \( O_k = f(S_k) \). \( O_k \) represents the response variable calculated by the network.

The network “learns” by making changes in its weights in order to reach a convergence between the response values \( O_k \) calculated by the network and the experimental responses \( y_k \). At present, the most extensively adopted algorithm for the learning phase is the back-propagation algorithm (Rumelhart et al., 1986; Zupan and Gasteiger, 1993) which is a generalization of the steepest descent method. The neural network training algorithm used in this work was developed at the LSCP (Process Control and Simulation Laboratory). A series of normalized input–output pairs of experimental data \((x, y)\) are used for training the network (learning set), and the weights are modified. The back-propagation algorithm uses gradient information to change the weights. However, weights are calculated with respect to only one input–output pair at a time. For the output layer, the weights are changed according to the following expressions:

\[
W_{j,k}^{(m+1)} = W_{j,k}^{(m)} + \Delta W_{j,k}^{(m)}
\]

where

\[
\Delta W_{j,k}^{(m)} = \eta \cdot f'(S_k) \cdot (y_k^{(m)} - O_j^{(m)}) \cdot O_j^{(m)}
\]

and \( \eta \) represents a dumping or accelerating factor.

During the training process, not only network parameters are varied, such as the number of neurons in the hidden layer (NH) or the normalization limits, but the learning parameters, such as the dumping/accelerating factor (\( \eta \)) and the number of iterations (NI), are also dynamically changed. The quality of the fit depends on the combination of the parameters mentioned above. As a rule, the number of neurons in the hidden layer should be as low as possible in order to avoid over-fitting problems. The network is tested and validated by comparing its predicted output values with the experimental values on an independent set of data (test set). Note that the test set was not used for modifying the weights of the network during training.

**Results and discussion**

**Neural network treatment**

Five input variables were chosen: (1) irradiation time \((t)\); (2) initial concentration of iron(III) salt \([\text{Fe}^{3+}]_0\); (3) total concentration of hydrogen peroxide \([\text{H}_2\text{O}_2]\); (4) temperature; and (5) initial concentration of the substrate \([\text{Xyl}]_0\).

In the present case, the concentration of 2,4-xylidine measured at a given time of irradiation \([\text{Xyl}]_t\) is the only experimental response or output variable \((y)\). Therefore, there is only one neuron in the output layer of the neural network. The number of data pairs was 104. The data pairs were alternatively distributed into the learning set (LS) and the test set (TS).

Heuristic rules may be applied to select the optimal neural network parameters in order to prevent over-fitting and over-training. These rules consist of the combination of the minimum NH with the minimum NI which provide the lowest mean square errors \((E)\) in TS and...
The number of neurons in the hidden layer (NH) was varied between 4 and 12 and the number of presentations of the LS data pairs (NI) to the network ranged from 10,000 to 500,000. Based on the trend of the E function and the distribution of the distance function, values of NH and NI equal to 8 and 100,000, respectively, were adopted for the simulations (dumping/accelerating factor $h = 1$).

**Validation of the model**

Figure 1 shows a comparison between calculated and experimental values of the output variable $([\text{Xyl}])_t$ using the neural network model with NH equal to 8. The agreement between the model predictions and the experimental data is good for the learning set. Some test points are not very well predicted, because these points belong to experiments where xylidine degradation was very fast, and the discrepancies might be explained by errors in the sampling times.

Plots of the $[\text{Xyl}]_t$ distance function (difference between calculated and experimental values scaled in increasing order) are represented in Figure 2 for both LS and TS and show a reasonably good distribution.

**Simulations**

Experiments were simulated with the neural network model and compared with the experimental results (Figure 3). The simulations of the experiments included in the neural network, as well as of the independent experiment, show a good correlation between predicted and experimental values.

![Figure 1](https://example.com/figure1.png)  
**Figure 1** Correlation between calculated and experimental results

![Figure 2](https://example.com/figure2.png)  
**Figure 2** Distribution of distance function for both LS and TS

![Figure 3](https://example.com/figure3.png)  
**Figure 3** Experimental results and simulations of degradation kinetics for experiments partially included to train the network (1 and 2) and for an independent test experiment (3). Experimental conditions of 3: $[\text{Fe}^{3+}]_0 = 1.2 \times 10^{-3}$ mol/L; $J = 32.5^\circ$C; $[\text{H}_2\text{O}_2] = 62.4 \times 10^{-3}$ mol/L.
Based on the fitted neural network model, simulations were carried out in order to predict the variation of the efficiency of 2,4-xylidine degradation as a function of the levels of the input variables. Figure 4 shows the computed initial reaction rate as a function of the initial concentrations of H\textsubscript{2}O\textsubscript{2} and Fe\textsuperscript{3+}.

By comparison, Figure 5 shows the variation of the initial reaction rate as a function of the Fe\textsuperscript{3+} concentration and the temperature. In both cases, the value of the third input variable has been set constant, in the first case the temperature was 35°C and in the second one the concentration of hydrogen peroxide was equal to 5 \times 10^{-2} \text{ mol/L}. Both values coincide with those at the center of the Doehlert uniform array.

The response surfaces in Figures 4 and 5 indicate that, in the experimental domain investigated, the initial concentration of iron(III) and the temperature have important and comparable effects on the degradation of xylidine, whereas the influence of hydrogen peroxide is relatively small.

**Conclusions**

In this work, we have shown that a neural network model based on a minimum set of designed experiments could be used to describe the rate of degradation of a pollutant as a function of three input variables in the experimental domain chosen for the photochemically enhanced Fenton process. By carrying out simulations based on this model, the sensitivity of this complex reaction system to the three key variables (initial concentrations of the additives and temperature) could be evaluated and response surfaces drawn. Additional simulations may be performed in order to describe the behavior of the system under...
different conditions. Based on these data, optimal values of experimental parameters may be selected for treating a given volume of wastewater and achieving degradation of the pollutant(s) in the required time period. Therefore, neural network models represent valuable tools for scaling-up and designing industrial scale batch reactors for the efficient degradation of organic contaminants in wastewaters.

Acknowledgements
Financial support of the experimental work done in Karlsruhe was granted by the BMBF through the Forschungszentrum Karlsruhe (Project number PWT-A94194-02WA9625/0). The authors also thank DAAD and CAPES for financial support of the exchange of scientists between Germany and Brazil in the framework of the PROBRAL program.

References