Principal Component analysis and discriminant analysis as a supervised pattern recognition tool for classification of AISI 420 steel samples subjected to a different heat treatment using Magnetic Barkhausen Noise signals

ALVAREZ-ROSARIO, A., PADOVESE, L.R.
Laboratory of Dynamics and Instrumentation - LADIN, Escola Politecnica - Universidade de Sao Paulo; São Paulo, Brasil
Phone:+55 (11) 30919652, e-mail:aarosario@usp.br, lrpadovese@usp.br

MEDINA-AGUAS, R., SERNA-GIRALDO, C.P.
Universidad de Antioquia; Medellín, Colombia
Phone:+57 (4) 2195540 - 2195543, e-mail:rufinomedina@gmail.com, claserna@udea.edu.co

Abstract
Measurements of Magnetic Barkhausen Noise (MBN) were carried out on AISI 420 steel samples of different hardness levels that were produced by means of controlled different heat treatments. The MBN measurements were treated by a supervised recognition method based on principal components and linear discriminant analysis in order to identify the levels of hardness through the MBN signal. Results showed that MBN can be used as a reliable source of information for statistical classification of steels in terms of hardness degree.

Keywords: Magnetic Non Destructive Testing, Magnetic Barkhausen Noise, Multivariate Analysis, Principal Component Analysis, Discriminant Analysis, Supervised pattern recognition

1. Introduction
In the last decades, the MBN phenomenon has received considerable attention as a useful tool for Non Destructive Testing methods (NDT) for magnetic materials. This is due to the fact that, as demonstrated, the MBN can provide accurate information on a number of material characteristics such as grain size [9, 17], carbon content [19, 3], stress state [7, 4], [12] and plastic deformation [2, 1, 21].

Although it is always a complicated issue to choose an appropriate method to analyze MBN signal parameters with the aim of explaining, inferring, and/or correlating changes in material microstructure or mechanical properties, several approaches have been proposed. Among these methods, one can cite spectral analysis [18], 16, chaos theory-based methods [22], wavelet analysis [10, 5, 16], statistical analysis [20, 14, 15] and, in less proportion, multivariate statistic [6, 13].

Multivariate statistical techniques like Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are commonly used in pattern recognition problems. PCA could be used for data dimensionality reduction (by indicating the variables with higher discriminatory power), while LDA is an exploratory tool to find out groups of datasets with similar characteristics [8]. However, these techniques have been little explored as tools for MBN analysis regarding quality control or else NDT purposes.

In this context, the aim of the present study is to show that PCA with LDA can be a valuable option for processing MBN features with a view to identify and classify steel samples having different levels of hardness.

2. MBN Signal Features
In order to correlate the information presented in the MBN signal to magnetic material properties, the following parameters are generally used:
2.1. Signal Energy

The signal energy is calculated by integrating the time dependence of the voltage squared signal on a measured MBN and is known as $MBN_{\text{energy}}$ (Eq. 1).

$$MBN_{\text{energy}} = \sum_{i=1}^{n} V_i^2 \Delta t$$  \hspace{1cm} (1)

Where: $V_i =$ value of the voltage measured at a certain moment; $\Delta t =$ time interval between the points of the signal. In this case the inverse of the sampling frequency.

2.2. Root Mean Square

The root mean square (RMS) is a statistical measure of the magnitude of a varying quantity, and is calculated using (Eq. 2).

$$MBN_{\text{rms}} = \sqrt{\frac{\sum_{i=1}^{n} (V_i - V_m)^2}{n - 1}}$$ \hspace{1cm} (2)

Where:
- $V_i =$ voltage value measured for the $i$-th observation;
- $V_m =$ average signal value;
- $n =$ number of signal observations

2.3. Number of events

An event is considered to have occurred when the measured voltage signal crosses the threshold with a positive slope, and to have ended when it crosses the threshold with a negative slope [11], as shown in figure 1.

![Figure 1: MBN Signal event example](image)

2.4. Envelope features

The signal envelope features are obtained from the analytical signal of the MBN signal. The analytical signal is the $S_+(t)$ signal containing only the positives frequencies of $S(t)$. This way, the associated $S_+(t)$ signal to the $S(t)$ signal, whose Fourier Transform $FT$, expressed by $S(v)$ is defined as follow:

$$S_+(t) = |FTI(Z_s(v))|$$  \hspace{1cm} (3)

where:

$$Z_s(v) = \begin{cases} 2S(v) & \text{for } v \geq 0 \\ 0 & \text{for } v < 0 \end{cases}$$  \hspace{1cm} (4)
From the signal envelope two parameters are commonly calculated: Peak amplitude \( (MBN_{apeak}) \), relative to the maximum value of envelope amplitude and Peak position \( (MBN_{ppeak}) \), relative to the position where the Peak amplitude \( (MBN_{apeak}) \) is located (figure 2).

3. Experimental

An AISI 420 steel bar was selected for the investigation. Its chemical composition is showed in table 1.

<table>
<thead>
<tr>
<th></th>
<th>Fe%</th>
<th>C%</th>
<th>Si%</th>
<th>Ni%</th>
<th>Cr%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>85.76</td>
<td>0.214</td>
<td>0.489</td>
<td>0.504</td>
<td>12.235</td>
</tr>
</tbody>
</table>

Samples used for the experimental measurements were obtained from the same piece of AISI 420 steel. Their geometry can be seen in figure 3.

Samples were subjected to quenching and tempering at different temperatures. Table 2 presents the different heat treatment (HT) sequences for the samples.

In Table 2, \( T00 \) refers to the initial state of the material (as from provider, without any heat treatment) while \( T01 \) indicates the initial annealing treatment (used for obtaining samples with homogeneous hardness). For sequences \( T11, T12, T13 \) and \( T14 \), samples were subjected to a tempering process in two stages as shown in Table 2. These \( T \) temperatures were chosen in order to modify grain size, hardness and material micro-structure in a different way for each sample.
Table 2: Experimental Heat Treatment Sequences

<table>
<thead>
<tr>
<th>HT Code</th>
<th>Number of Samples</th>
<th>HT #1</th>
<th>HT #2</th>
<th>T (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T00</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>T01</td>
<td>2</td>
<td>Annealing</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>T10</td>
<td>2</td>
<td>Quenching</td>
<td>Tempering</td>
<td>-</td>
</tr>
<tr>
<td>T11</td>
<td>2</td>
<td>Quenching</td>
<td>Tempering</td>
<td>200</td>
</tr>
<tr>
<td>T12</td>
<td>2</td>
<td>Quenching</td>
<td>Tempering</td>
<td>300</td>
</tr>
<tr>
<td>T13</td>
<td>2</td>
<td>Quenching</td>
<td>Tempering</td>
<td>500</td>
</tr>
<tr>
<td>T14</td>
<td>2</td>
<td>Quenching</td>
<td>Tempering</td>
<td>600</td>
</tr>
</tbody>
</table>

A personal computer with a data acquisition device (with A/D, D/A and D/D channels) supplies a sinusoidal wave of 10 Hz that feeds the magnetic circuit in order to magnetize the sample with a magnetic field of 71.4 - 104 A/m, which produces magnetization in the samples.

The MBN sensor output is amplified and band pass filtered (1 - 100 kHz). The MBN signals were visualized using a digital oscilloscope and a data acquisition device performed the digital acquisition with a sampling frequency of 200 kHz. The entire process is controlled using the software MAGVIEW.

![Experimental setup for MBN signal measurement](image)

The following statistical parameters were calculated to analyze the signals: $MBN_{rms}$ (2), $MBN_{energy}$ (1), peak position $MBN_{ppeak}$ and peak amplitude $MBN_{apeak}$ (Figure 2) and finally the number of Barkhausen events $MBN_{events}$ (Figure 1).

The quality of measurement dataset was evaluated to identify which records are to be excluded (abnormal feature values or outliers). The number of observations in the final dataset is presented in table 3.

Table 3: Number of observations in final dataset

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Measurements</th>
</tr>
</thead>
<tbody>
<tr>
<td>T00</td>
<td>60</td>
</tr>
<tr>
<td>T01</td>
<td>60</td>
</tr>
<tr>
<td>T10</td>
<td>40</td>
</tr>
<tr>
<td>T11</td>
<td>48</td>
</tr>
<tr>
<td>T12</td>
<td>44</td>
</tr>
<tr>
<td>T13</td>
<td>60</td>
</tr>
<tr>
<td>T14</td>
<td>60</td>
</tr>
</tbody>
</table>

4. Results and discussions

Figure 5 and 6 show the different MBN features Vs. Hardness. As presented in these figures, all MBN features have similar behavior with hardness variation. Then, it can be affirmed that, one could recognize and classify each sample according to its level of hardness through any of these MBN features. In fact, the most commonly studied and used MBN features to correlate hardness ($MBN_{rms}$ and $MBN_{energy}$) show similar behavior for treatments $T10$, $T11$, $T12$ and $T13$, making the task of identification and classification, by these two parameters, a non reliable approach.
Figure 5: Measured MBN features ($MBN_{peak}$, $MBN_{ppeak}$) Vs. Hardness

Figure 6: Measured MBN features ($MBN_{rms}$, $MBN_{energy}$, $MBN_{events}$) Vs. Hardness

Figure 7 shows the shapes of the MBN signal for all the treatments for samples randomly selected. The same difficulty in identifying the hardness for treatments $T_{10}$, $T_{11}$, $T_{12}$ and $T_{13}$ by the hardness curve appears. Shapes of MBN signal are so similar (especially $T_{01}$ and $T_{11}$) that it is not feasible to correlate hardness among these samples just by the MBN shape.
Thus, a better approach is to analyze the features of the MBN signal as a whole and not as isolated data (multivariate approach). Initially the problem is to identify and select the most appropriate features to discriminate the samples in groups of different levels of hardness. In order to find which MBN feature has a better discriminant power, the PCA method was employed. To avoid the influence of the quantity of different features, all features were standardized by the mean and standard deviation using (5).

\[
 z = \frac{(x - \mu)}{\sigma} \quad (5)
\]

Where:
- \(x\) = feature vector observations
- \(\mu\) = mean value of vector \(x\)
- \(\sigma\) = standard deviation of \(x\)

After applying the PCA to the 5-dimensional dataset, one principal component (PC) was extracted. The percentage of the explained variance by this PC is 92.54% (Table 5). The percentage of the first PC is much larger than those of the others PCs. Thus, according to the loadings of variables in the first PC, as shown in Table 5, the most contributing variables were \(MBN_{rms}\), \(MBN_{energy}\), and \(MBN_{apeak}\).
Table 4: Importance of the principal components

<table>
<thead>
<tr>
<th>Variable</th>
<th>PC1</th>
<th>PC2</th>
<th>PC3</th>
<th>PC4</th>
<th>PC5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard deviation</td>
<td>2.151</td>
<td>0.487</td>
<td>0.366</td>
<td>0.038</td>
<td>0.024</td>
</tr>
<tr>
<td>Proportion of variance</td>
<td>0.925</td>
<td>0.047</td>
<td>0.027</td>
<td>0.0003</td>
<td>0.0001</td>
</tr>
<tr>
<td>Cumulative proportion</td>
<td>0.925</td>
<td>0.973</td>
<td>0.9996</td>
<td>0.9999</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

Table 5: Factor loadings for principal components

<table>
<thead>
<tr>
<th>Variable</th>
<th>PC1</th>
<th>PC2</th>
<th>PC3</th>
<th>PC4</th>
<th>PC5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$MBN_{rms}$</td>
<td>-0.461</td>
<td>0.259</td>
<td>0.804</td>
<td>0.257</td>
<td></td>
</tr>
<tr>
<td>$MBN_{energy}$</td>
<td>-0.457</td>
<td>0.363</td>
<td>0.116</td>
<td>-0.552</td>
<td>0.584</td>
</tr>
<tr>
<td>$MBN_{events}$</td>
<td>-0.433</td>
<td>-0.510</td>
<td>-0.726</td>
<td>-0.154</td>
<td></td>
</tr>
<tr>
<td>$MBN_{apex}$</td>
<td>-0.455</td>
<td>0.425</td>
<td>-0.150</td>
<td>-0.768</td>
<td></td>
</tr>
<tr>
<td>$MBN_{ppeak}$</td>
<td>0.430</td>
<td>0.600</td>
<td>-0.671</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

PCA could be used to make an unsupervised classification. However this technique produces just five well-defined groups, as seen in Figure 8, and not seven as expected.

The selected variables in PCA were used as input parameters for LDA. The accuracy of the results was tested by using a Plugin-validation and a Cross-validation test (Table 6). Both tests showed that, the use of MBN features selected from PCA produces successful classification with posterior error probabilities (PEP) close to 0.00%. Additionally, an LDA was carried out for each MBN feature individually, $MBN_{rms}$ and $MBN_{energy}$ showed a good agreement with the PCA-LDA based model but PEP considerably increased (Tables 7 and 8), especially in treatments $T_{10}$, $T_{11}$ and $T_{12}$ reaching values above 70% ($MBN_{energy}$) showing that the analysis based on PCA and LDA techniques provides better classification results than that based on the choice of only one MBN feature.
5. Conclusions

This research showed that MBN signal features can be used for hardness level based classification of AISI 420 steel samples. The LDA, used with previously PCA, was effective in discriminating the seven used groups of AISI 420 steel samples with high degree of accuracy (global error close to 0.0%). In doing so, the MBN could be used as a tool for NDT methods or for quality control of ferromagnetic materials with different levels of hardness.

Acknowledgements

This research was supported by the Brazilian Council of Technological and Scientific Development (Cnpq).

References


