# PLANT DATA VISUALIZATION USING NON-NEGATIVE MATRIX FACTORIZATION

**J. Tapson and J. R. Greene**

*Department of Electrical Engineering, University of Cape Town, UCT Private Bag, Rondebosch 7701, South Africa*

Abstract: Non-negative matrix factorization (NMF) is a method for dimensionality reduction and simplification of large data sets. Unlike tools such as principle components analysis (PCA) and factor analysis, NMF produces basis vectors that correspond to perceptible features in the original data. This is particularly useful when working with data where visual interpretation of the simplified representation is required. Typical data of this type is condition monitoring (CM) data, where visual interpretation of vibration spectra is a standard diagnostic tool. The results suggest that NMF processing of CM data simplifies the visual interpretation process, and opens the way for automation of this task. 

Keywords: Failure detection, data reduction, spectrum analysis, computer aided diagnosis.

## 1. INTRODUCTION

Modern process plants generate an enormous amount of data, which is typically stored by automated process control historians. It is common cause that significant improvements in process optimization can be made by appropriate examination of this data – so-called exploratory data analysis (EDA) or data mining. EDA techniques range from those with a rigorous analytical foundation, such as principle components analysis (PCA) (Jolliffe, 1986) or independent component analysis (ICA) (Hyvärinen *et al*., 2001), to those which are notable simply for effectiveness, such as the many variations of neural networks (see Aldrich (2002) for a recent overview of EDA techniques and their applications). Most of these techniques are characterized by producing reduced data sets that are difficult to relate to the original data, except by abstraction (as an *n*th moment in PCA for example). The techniques are useful, but there are many cases where it would be extremely helpful if the reduced data sets could be related in some straightforward and visible way to the original data.

The core requirement in EDA is to devise a linearly transformed version s of the original data **V** such that

$$\mathbf{V} = \mathbf{A}\mathbf{s} \quad (1)$$

(where **A** is a matrix) with the intention that **s** should have more analytical utility than **V**; hence the goal is a factorization of the original data matrix. Hyvärinen, *et al*. (2001) have categorised the available transforms as either second order and "faithful", or higher-order and "meaningful", which is a useful distinction when the purpose is visualisation.

## 2. NON-NEGATIVE MATRIX FACTORIZATION

Non-negative matrix factorization (NMF) is a recently invented technique (Paatero and Tapper, 1994; Lee and Seung, 1999) which falls into the "meaningful" category. The basis of NMF is that a data set **V** (represented as an $n \times m$ matrix) can be approximated by the product of two factors **W** and **H** such that

$$\mathbf{V} \approx \mathbf{W}\mathbf{H} \quad (2)$$

Where **W** is an $n \times r$ matrix and **H** is an $r \times m$ matrix, where $r$ is smaller than $n$ or $m$. There is a critical constraint that all of **V**, **W** and **H** must be non-negative, so the method lends itself to classes of data where this is not a limiting constraint (and where the assumptions of zero mean and Gaussian distribution, on which many techniques depend, do not hold).

The non-negative constraint is key to the utility of the technique, as it means that the components and weights must be positive, so all constructions of components must be additive. Specifically, there can be no component that is subtracted from others (by means of a negative weight for example); hence all components must be intuitively real parts of the whole.

The effect of the NMF approximation is to generate a reduced set of basis vectors **W** which are combined linearly using weights (the rows of **H**). In many types of process data this is a useful simplification, with benefits in terms of subsequent processing, noise reduction, and improved visual intelligibility.

In particular, NMF returns a sparse set of base features and weights, which corresponds better to a parts-based model of the raw data than do either of PCA or ICA, in which there are many basis vectors which are added and subtracted with varying weights to arrive at an approximation of an individual data set.

NMF has so far been used for purposes as diverse as identifying parts of visual images (for example, the parts of faces in photographs of faces (Lee and Seung, 1999)); for analysis of Raman spectroscopy data, hyperspectral images, and human brain chemical shift images (Sajda *et al*., 2003); and for transcription of music (Smaragdis and Brown, 2003). It is this last application that points the way from the plant visualisation point of view, as it was the first to make use of a novel property of the NMF matrices. This is that the weight vector (**H**) gives a time-series representation of the development of intuitively tractable parts of the data. As will be shown in the examples that follow, this is enormously useful in simplifying data analysis.

There are a number of methods for finding the factors **W** and **H**. The methods usually depend on the cost function used to define the quality of the factorization. In this work, Euclidean distance (between the original vector **V** and the product of factors **WH**) is used as the cost function. This cost function provides a simple successive approximation algorithm for generating **W** and **H** (see Lee and Seung (2001) for a description of the algorithm and a proof of its convergence).

The use of NMF in data visualization will depend on the hypothesis to be tested, but might proceed as follows: assume a plant process generates data at a fixed rate, so that in each time interval one set of data is acquired. It is believed that in this data there is an underlying structure that could be visible to an operator or analyst, but this structure is obscured by noise and the large size of the data set. If one forms a matrix **V** where the columns are data sets from successive time intervals, one can approximate the matrix by two factors **W** and **H** of reduced dimension, according to Eqn. 2. The factor **W** will contain base data sets or data features, which should be simpler and more recognizable than the original data, and the factor **H** will give their weights over time, so that one can visualize how the process has evolved in time.

## 3. SIMPLIFICATION OF PLANT CONDITION DATA USING NMF

The use of NMF can be illustrated with examples using vibration data gathered by on-line condition monitoring systems on real industrial plant. It must be stressed that the data sets used in this paper were gathered with no *a priori* knowledge of likely plant behaviour; they encompass plant operational failures that were neither expected nor instigated, although *ex post facto* knowledge was obviously helpful in selecting the data sets.

Many predictive maintenance systems monitor the vibration spectra of machines in order to detect the early onset of mechanical and electrical failures. However, many failure modes, such as mechanical failure of bearings, do not add significantly to the vibration energy and cannot be detected by observation of the average energy, or even by bandwidth-limited energy measurements. The best method of detecting early bearing failure is by observing the shape of the spectrum of vibration, which will show certain signature features depending on the type of bearing and the failure mode. For example, a flaw in the inner race of a roller bearing would produce a regular acoustic impulse each time a roller passes over the flaw, and the impulses would produce a ringing tone whose frequency would depend on the resonant structure of the bearing. A combination of simple mathematics and accumulated experience suggests a set of features which a human analyst could look for in a vibration spectrum to detect early stages of bearing failure.

There is an underlying assumption in predictive maintenance that for each root cause of failure there is a corresponding spectral feature set, and recognition or identification of this feature set will constitute diagnosis of cause of impending failure. Additionally, the magnitude of the feature is assumed to give some sense of the degree of progression of the failure.

Unfortunately, large-scale monitoring of machinery for bearing failure is impossible, if human scrutiny of each vibration spectrum is required. Fortunately, NMF may be used to reduce the complexity of the data. The following examples present three cases showing how this may be achieved.

*3.1 Linear mixture model*

A linear mixture model (after Sajda et al. (2002)) for the vibration spectra is proposed. Using the same notation as in the discussion on NMF, a set of sampled spectra **V** (where each column is a single vibration spectrum) may be modelled as follows:

$$\mathbf{V} = \mathbf{WH} + \mathbf{N} \qquad (3)$$

where the columns in **W** represent the spectral shapes which are characteristic of individual root causes, the rows in **H** represent the preponderance of each root cause, and **N** is additive noise. The sampled vibration spectra are magnitude spectra resulting from a Fast Fourier Transform (FFT) of time series vibration data, so the elements of **V** must be non-negative; similarly, it is reasonable to assume that all of **W**, **H** and **N** would have non-negative elements.

Based on this model, our analysis has three purposes:

- to extract the constituent spectra as columns of **W**, which will enable a diagnosis of root cause of failure;
- to extract the preponderance or weights as rows of **H**, which will give the relative significance of the root causes;
- to examine the columns of **H**, which will indicate the development in time of each possible root cause.

This last point is particular to plant processes, which tend to be time-varying, and indicates our point of departure from prior NMF applications in spectral analysis.

*3.2 Supervised and Unsupervised Methods*

The basic NMF method is an example of unsupervised learning, and its use in vibration spectrum analysis is therefore an example of blind source separation. Sajda *et al.* (2003) have shown that by applying constraints (forcing low spectral amplitudes to zero, and seeding **W** with "endmembers", or expected spectral feature sets), improved results are possible. However, in order to maximise the generality of results, in the present work only the standard (unsupervised) NMF is used.

## 4. EXAMPLES FROM REAL PLANT CONDITION DATA

In the first case, the data consists of twenty vibration spectra taken on successive days from a motor driving a pharmaceutical centrifuge. The motor developed a bearing failure on the eighteenth day and failed before the twenty-first day. The spectra are shown in the traditional waterfall diagram in Figure 1.

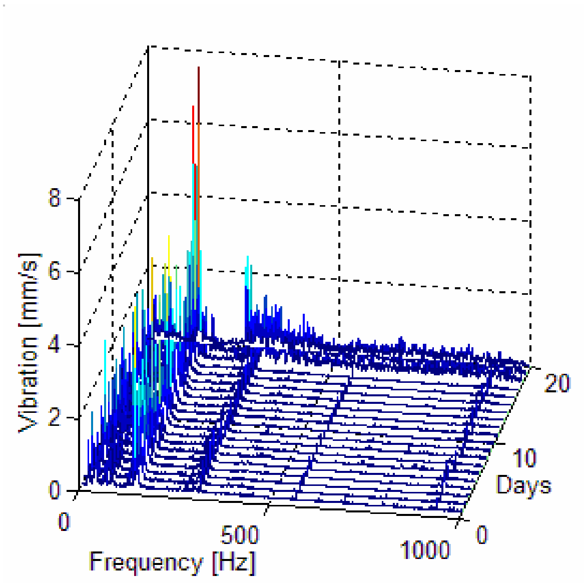


Fig. 1. Waterfall diagram of vibration (velocity magnitude) spectra from an electric motor. The spectra were taken at daily intervals, with the earliest spectrum at the front. The last three spectra show significant change, which would indicate a possible failure developing (the raised noise floor is traditionally considered to indicate bearing deterioration). The motor suffered a catastrophic bearing failure shortly after the last spectrum was sampled.

In order to perform NMF it is necessary to decide on the degree of reduction required. Some guidelines will be developed later in this paper, but at this stage the minimum sensible value of r=2 is used (assuming that there are two states, representing normal and abnormal operation, which it is the goal to distinguish). The figures below show the vectors **W** and **H'**.

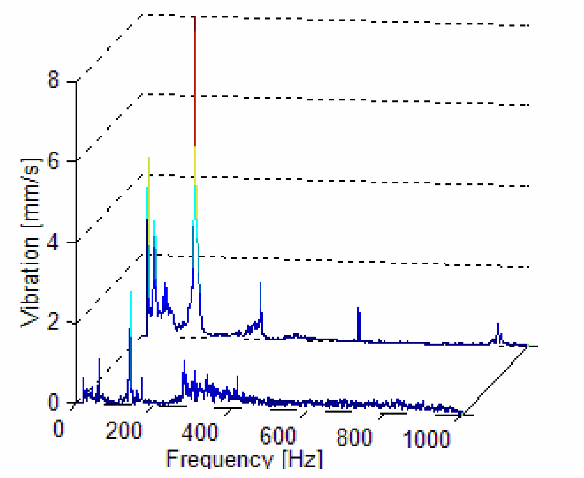


Fig. 2. The two basis vectors produced by the NMF process. These ought to correspond to two distinct feature sets of the vibration spectra. The relative weight of these feature sets and their development in time is shown in Figure 3.

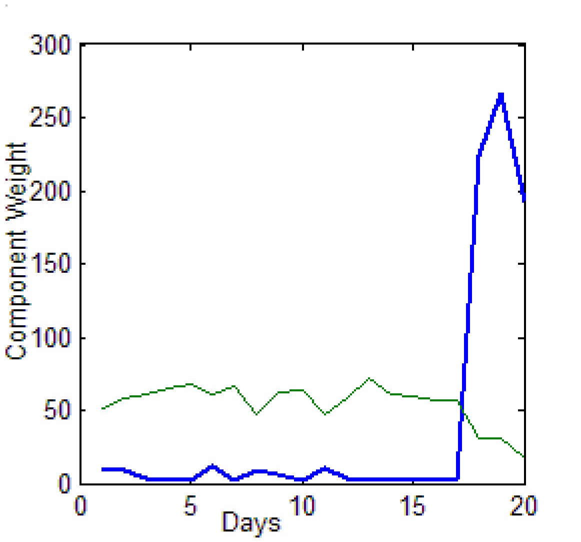


Fig. 3. Graph of **H'** showing the relative weights of the two components and their development in time (the column index of **H** is effectively the day number). The thin line indicates the weight of the rearmost spectrum in Figure 2, and the thick line gives the weight of the foremost spectrum. The point at which the bearing failure starts to affect the vibration (on day 18) is extremely clear.

Figures 2 and 3 illustrate the separation of the data into features and weights. It can be seen that the reduction of the spectra to two distinct feature sets, and the demonstration of their development in time, substantially clarifies the situation for a human analyst.

The interpretation of **H'** is a key to the successful use of NMF in this application. In cases such as Fig. 3 it is straightforward, but it is less so when the data reduction is insufficient. This can be illustrated with the trivial case of performing NMF on a similar data set which has been manipulated to have identical columns. For this case, Figures 4 and 5 are equivalent to figures 2 and 3 above. It can be seen that the NMF process generates two identical feature sets, and that the relative weights oscillate between the two feature sets.

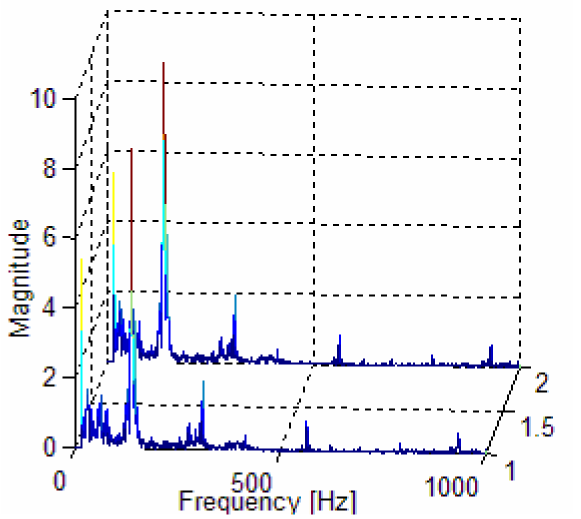


Fig. 4. Waterfall diagram showing the two basis vectors produced by NMF on a set of 20 identical spectra. Not surprisingly, these two basis vectors are identical to the original spectra.

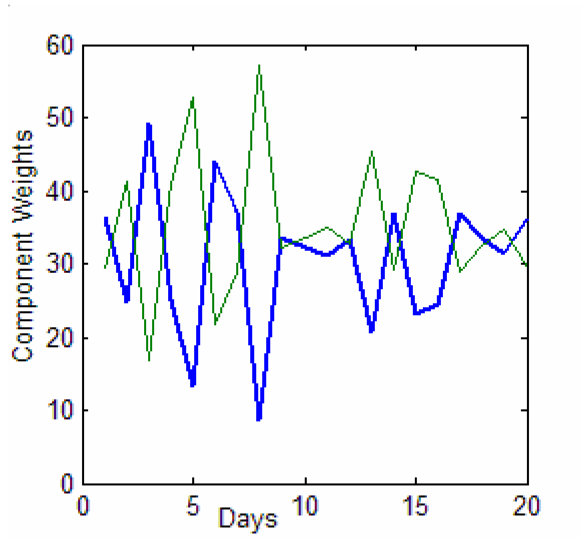


Fig. 5. Graph of **H'** for the NMF of 20 identical spectra. It can be seen that the weights are antisymmetric with respect to an average value. This appearance is characteristically observed when the data is insufficiently separable; although in some cases the weights converge to two constant values.

The effect of insufficiently reducing the data is shown in a trivial case in Figure 5, and in a real case in Figure 6 (for the same data as in Figs 1, 2 and 3). It can be seen that the effect is to produce a pair of weights in **H'** (that is, a pair of rows of **H**) which fluctuate antisymmetrically around a mean value.

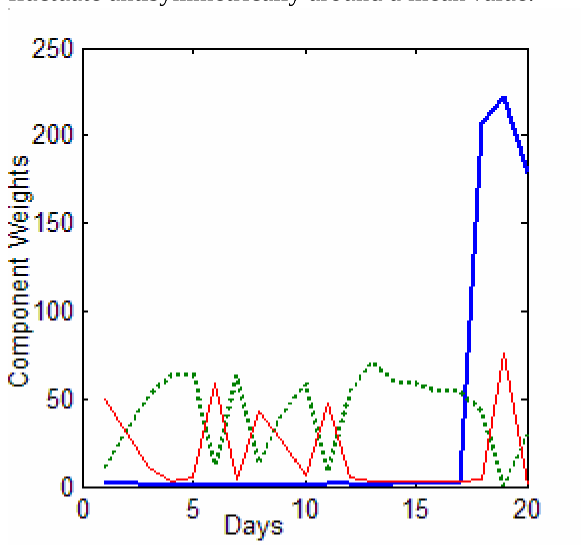


Fig. 6. Graph of **H'** for the original data of Figure 1, reduced using NMF to three feature sets. It can be seen that the weight for the fault condition features is still extremely apparent. However, the attempt to factorize (extract) a third feature set produces weights for the non-fault feature set which fluctuate antisymmetrically, suggesting that there is no significantly different third feature set.

Figures 7 and 8 show the vectors H' and W for a second machine failure, in which a centrifuge bearing deteriorated suddenly. Note that in Fig. 7 the "healthy" component falls to zero as the "failing" component rises; this is counterintuitive as the healthy components should form a baseline for all spectra. The reason for this anomaly is touched on in Section 6.

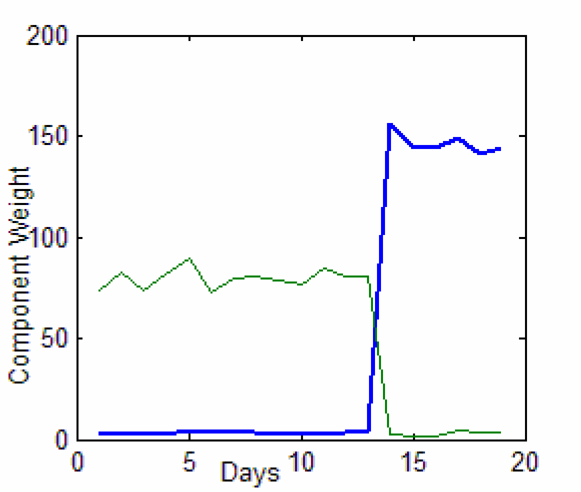


Fig. 7. The weights of the feature sets for the vibration data for a failing centrifuge bearing. It can be seen that the basic features of a healthy bearing (thin line) are replaced on day 14 with the features of a failing bearing.

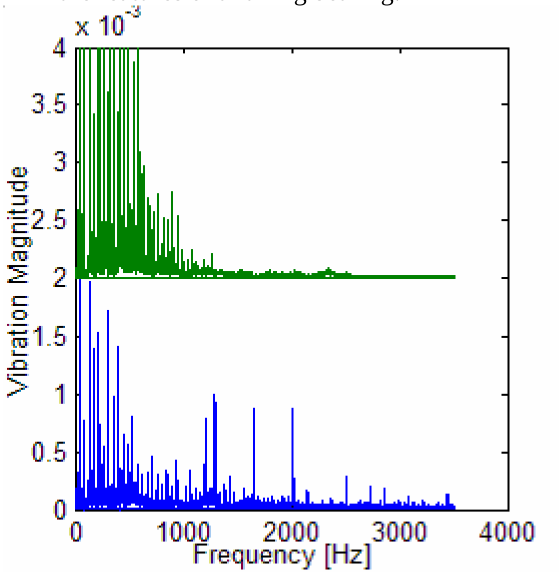


Fig. 8. The basis vectors, or spectral feature sets, for a failing centrifuge bearing. The lower set is the failing spectrum, and the upper (shifted for clarity) is the healthy spectrum. The peaks in the lower spectrum between 1200 and 3000 Hz are caused by regular impact events as the bearing rotates.

## 6. PRACTICAL USAGE OF THE NMF TECHNIQUE

There are a number of variables used in the NMF algorithm which can act to improve or reduce the utility of the result. These include the cost function, the degree of data reduction, the number of iterations of the algorithm, and the number of restarts (from a random initialisation) that may be required to generate a useful convergence.

There is disagreement over the choice of cost function, with Lee and Seung (1999) stating that it is not important, while the results of Sajda *et al.* (2003) indicate otherwise. This is a non-trivial question and requires research beyond the scope of this initial exploration.

In respect to whether a useful factorization – defined as a correct decomposition into parts – can actually be achieved, the reader is directed to Donoho and Stodden (2004), who have derived conditions for that result. In particular, their requirement **R3** ("complete factorial sampling"), that the dataset contains all permutations of the features in all possible combinations, is unlikely to be met in practice. Hence, one must approach the results with the necessary scepticism. Notably, if there is an invariant region in each dataset (which is highly likely in plant condition data), then this region is likely to be found in each base vector (hence the result in Figure 7).

An important observation is that because the output of NMF is inherently tractable to perception, it is immediately clear whether the factors are useful or not. This saves a great deal of speculative number crunching, and allows a fast convergence to useful parameters.

Users of PCA and ICA will be familiar with the use of eigenvalues as a measure of the significance of the components, and the use of this knowledge in deciding the degree of data reduction. In NMF, the degree of data reduction does not appear to be as sensitive to the number of components. Figure 9 below shows the error in NMF factorization relative to the number of components, and it can be seen that the error reduces smoothly as more components are added. Figures 10 and 11 give associated results for the number of iterations and restarts respectively.

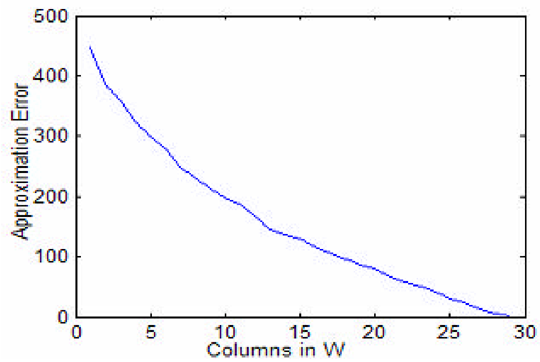


Fig. 9. The error between input data and the approximation result, for raw data similar to figure 1, is shown. The original data set had 29 columns, hence the zero error for that value. It can be seen that there is no step or inflexion in this curve, suggesting that the

choice of degree of reduction is not obvious. The measure of error is Euclidean distance.

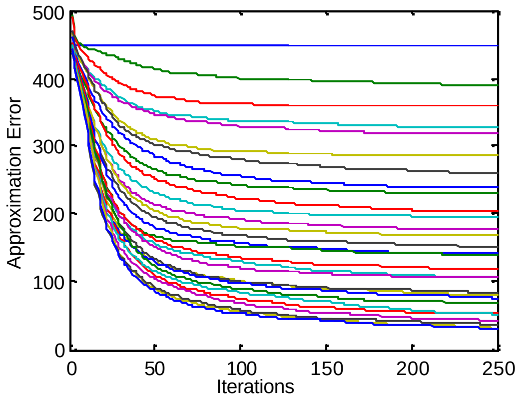


Fig. 10. Convergence to a solution for all possible values of $r$ (reduced dimension), from $r$=1 (top) to $r$=29 (bottom). The number of iterations to convergence does not vary significantly with the value of $r$, as each iteration includes a column-by-column local optimisation; so the processing complexity and time inherently scales with $r$, rather than the number of iterations.

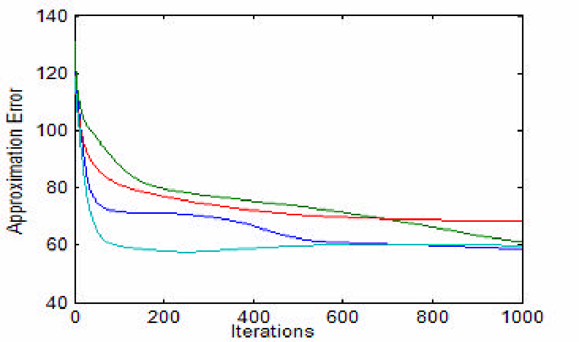


Fig. 11. Convergence to a solution for the same data set and same $r$ ($r$=4) value, with four different random seed data sets. It can be seen that there is some considerable variation in convergence depending on the initial data set.

## 7. CONCLUSIONS

Non-negative matrix factorization is a useful tool in simplifying one type of plant data, namely vibration spectra captured for condition monitoring. The output matrices are directly interpretable as spectral components, and the weights of those components in time, although some care must be taken in this interpretation. Specifically, attempts to over-specify the output must be recognised, and it should be noted that the independence of the base components is not a foregone conclusion, particularly when the input data does not fully characterise the system behaviour.

## ACKNOWLEDGEMENTS


The authors would like to thank Travis Dent of Motornostix (Pty) Ltd for the data sets.


## REFERENCES


Aldrich, C. (2002). *Exploratory Analysis of Metallurgical Process Data with Neural Networks and Related Methods (Process Metallurgy Series 12)*. Elsevier Science B.V., Amsterdam.

Donoho, D. and Stodden, V. (2004). When Does Non-Negative Matrix Factorization Give a Correct Decomposition into Parts? In *Advances in Neural Information Processing Systems 16*, MIT Press, Cambridge MA.

Hyvärinen A., J. Karhunen and E. Oja (2001). *Independent Component Analysis*, Wiley Interscience, New York.

Jolliffe, I.T. (1986). *Principle Component Analysis*, Springer-Verlag, Berlin.

Lee, D.D. and H.S. Seung (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, **401**, 788-791.

Lee, D.D. and H.S. Seung (2001). Algorithms for non-negative matrix factorization. In *Advances in Neural Information Processing Systems 13*, 556-562, MIT Press, Cambridge MA.

Paatero, P. and U. Tapper (1994). Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics*, **5**, 111-126.

Sajda, P., S. Du, and L. Parra (2003). Recovery of constituent spectra using non-negative matrix factorization. *Proc. 4$^{th}$ International Symposium on Independent Component Analysis and Blind Signal Separation*, 71-76.

Smaragdis, P. and J.C. Brown (2003), Non-negative Matrix Factorization for Polyphonic Music Transcription, *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 177-180.