

# VISUALIZATION OF PROCESS DATA WITH BILOTS

Sugnet Gardner,<sup>§</sup> Niel J. Le Roux,<sup>§</sup> and Chris Aldrich\*

<sup>§</sup>Department of Statistics and Actuarial Science, University of Stellenbosch,  
Stellenbosch, Private Bag XI, Matieland, 7602, Stellenbosch, South Africa.

\*Institute of Mineral Processing and Intelligent Process Systems,  
Department of Process Engineering, University of Stellenbosch  
Stellenbosch, Private Bag XI, Matieland, South Africa 7602  
Fax +27(21) 808 2059

**Abstract:** In this paper the methodology of biplots is introduced as a means for monitoring the behaviour of process systems. This sophisticated methodology allows for the projection of high-dimensional data to a low-dimensional subspace that can be visualised by a human operator. The projections are highly graphical in nature, and rich in information regarding variation in process variables, correlations among these variables, as well as class separation, taking into account the multivariate character of the data. Moreover, as is shown by way of two case studies, process disturbances can be visualised and explored quantitatively by superimposing alpha-bags on biplots.

**Keywords:** Biplot, Process Control, Modelling, Hydrometallurgy, Froth Flotation.

## 1. INTRODUCTION

Process data are used extensively in modern industrial environments for monitoring product quality, control and optimization. Large volumes of data are routinely collected and stored on many plants and in order to exploit the data to get a better understanding of the behaviour of the process, it is important to identify the salient features underlying the data. By reducing the dimensionality of the problem, the engineer is able to summarize the information captured in a large number of variables by a smaller number of latent variables. Principal component analysis (PCA) is by far the most important technique used for this purpose, and is widely supported by statistical software.

As is indicated in this paper, PCA can be extended with modern biplot methodology by providing a single graph for displaying the variation in multidimensional observations, together with information on all variables concerned. Canonical variate analysis (CVA) biplots on the other hand, are ideally suited for optimally separating classes of observations. Biplots provide an infrastructure for implementing many novel ideas in monitoring industrial plant processes and allows for non-linear relationships among process variables.

## 2. BILOT METHODOLOGY

A biplot is a graphical display consisting of a vector for each row and a vector for each column of a matrix of rank two (Gabriel, 1971). The elements of the matrix are represented by the inner products of the vectors corresponding to their rows and columns. Since any matrix of rank  $k > 2$  can be approximated

by a matrix of rank two, a biplot can be constructed for all matrices by considering its rank two approximation. Although the traditional Gabriel biplot is widely applied as a graphical aid in practice, biplots can also be viewed as multivariate analogues of scatterplots that are easy to interpret (Gower, 1995; Gower and Hand, 1996). In regarding the biplot as a multivariate extension of an ordinary scatterplot the focus is on representing interpoint distances. Since scatterplots have the merit of being familiar, requiring very little formal training to interpret, this biplot is accessible to non-statistical audiences. The multidimensional observations are represented by points in a two-dimensional display while variables are represented as biplot axes – for each variable a separate axis. Furthermore, these (non-perpendicular) axes are calibrated in the original scales of measurement so as to be used in much the same way as the two perpendicular axes of a scatterplot.

Mathematically, the process of finding the coordinates of the original points in the PCA biplot space amounts to performing a singular value decomposition of the data matrix consisting of  $n$  observations on  $p$  variables. However, these axes are not shown but form the scaffolding for constructing the biplot. The relationships between the scaffolding and the original variables are termed *interpolation* and *prediction* (cf. Gower & Hand, 1996). For a given or a new observation, interpolation is the process of finding its representation in the biplot space. Prediction, on the other hand, is inferring the values of the original variables in terms of the biplot scaffolding. Thus, what is shown in a PCA biplot are the observations and a set of axes, calibrated in the original units, representing the variables. It can be shown that two different sets of axes are needed: one

set for interpolation and another for prediction. Since it is natural to use axes for inferring values of the observations for the different variables, only prediction axes will be fitted to biplots shown here. Interpolation is usually performed by a suitable computer programme.

CVA aims to find the linear combination of the predictor or discriminatory variables that maximises the ratio of the between groups to within groups variance. This process amounts to transforming the original means to a new set of means, known as canonical means. The statistical (Mahalanobis) distances in the observation space become ordinary Euclidean distances in the canonical space. Mathematically, the process of finding the coordinates of the original points or means in the canonical space amounts to solving a two-sided eigenvalue problem leading to a set of axes for constructing a biplot.

Moreover, by adding alpha-bags to a CVA biplot a quantitative measurement of multidimensional overlap or separation of groups can be obtained, thereby providing information not only about the degree of overlap, but also the nature of the overlap. Gardner (2001) proposed the alpha-bag for characterising a two-dimensional cloud of points as well as for quantifying overlap among classes. Although trivial to rank univariate data according to a single criterion *e.g.* length, the process of ranking multivariate data is far from trivial. Tukey (1975) and Rousseeuw, Ruts & Tukey (1999) discussed the concept of halfspace location depth as a means of generalising the univariate rank concept to a bivariate data set. This concept forms the basis of the *alpha bag* that encloses the exact innermost  $\alpha\%$  of bivariate data points as specified by the user. The resulting contours are called alpha-bags and are proposed as summaries of data points in two dimensions.

### 3. HYDROLYSIS OF ZINC CHLORIDE IN AMMONIUM CHLORIDE SOLUTIONS

Since aqueous solutions of high concentrations of ammonium chloride are especially appropriate for the treatment of complex raw materials of both oxide and sulphide types, hydrometallurgical processes using ammonium chloride have generated considerable interest in recent years (Figueiredo *et al.*, 1993; Limpo *et al.*, 1992). Process development cannot take place without a detailed knowledge of the solubility of metal chlorides as a function of temperature and composition of solution and in this case study data generated by Limpo *et al.* (1995) are examined by means of biplots.

The data set consisted of 108 samples obtained from an experiment where zinc chloride is hydrolysed in a watery ammoniacal-ammonium chloride solution. Four variables were measured, namely the temperature of the solution (ranging from 30-50°C), the concentration of chloride anions (Cl<sup>-</sup>), the concentration of zinc cations (Zn<sup>2+</sup>) and the ammonia concentration (NH<sub>3</sub>). The concentrations were all

less than 5 M. Under these conditions, three phases of zinc chloride can occur, *viz.* Zn(NH<sub>3</sub>)<sub>2</sub>Cl<sub>2</sub>, Zn(OH)<sub>2</sub> and Zn(OH)<sub>1.6</sub>Cl<sub>0.4</sub>. Of the 108 samples, 22 were associated with the precipitation of Zn(NH<sub>3</sub>)<sub>2</sub>Cl<sub>2</sub>, 37 with and 43 with Zn(OH)<sub>2</sub> and Zn(OH)<sub>1.6</sub>Cl<sub>0.4</sub>, while six samples consisted of mixed precipitates, *i.e.* three observations associated with Zn(NH<sub>3</sub>)<sub>2</sub>Cl<sub>2</sub>-Zn(OH)<sub>1.6</sub>Cl<sub>0.4</sub> and three observations associated with Zn(OH)<sub>2</sub>-Zn(OH)<sub>1.6</sub>Cl<sub>0.4</sub>.

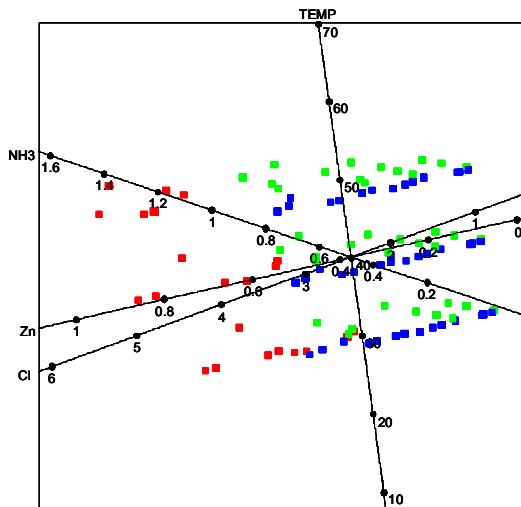
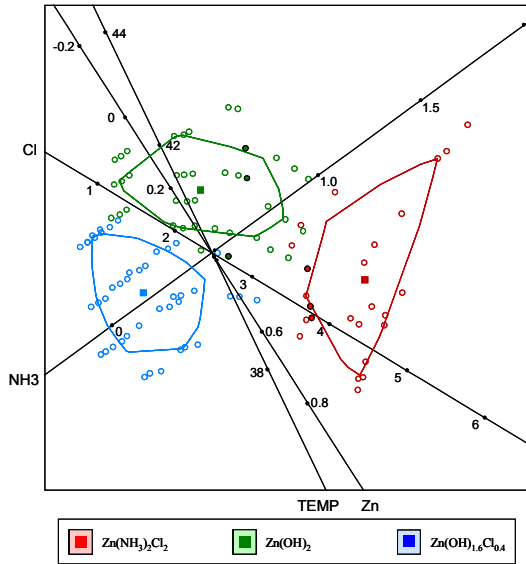


Fig.1. PCA biplot of the hydrolysis of zinc chloride data set.

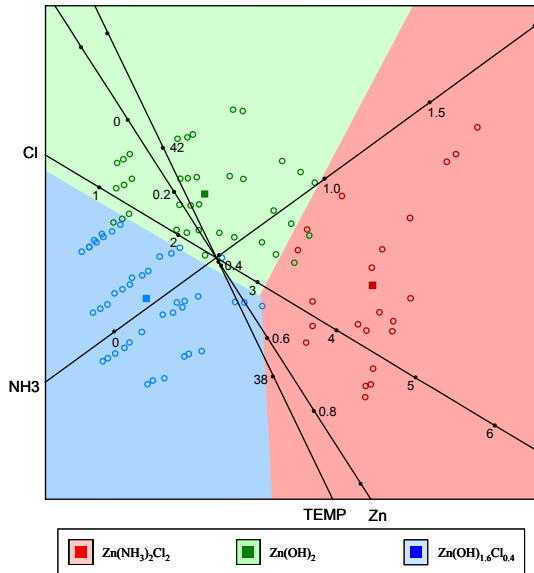
Figure 1 shows a PCA biplot of this data set. The sample points appear in three horizontal bands corresponding to the three temperature settings. The different precipitates, excluding the six mixed observations, are indicated by different colours. The variation in the values of NH<sub>3</sub>, Zn<sup>2+</sup> and Cl<sup>-</sup> for each precipitate can be clearly ascertained from the respective calibrated axes. Although the three precipitates can be clearly discerned, the PCA biplot optimally represents the total variation in the observations. Constructing a CVA biplot of this data optimally separates the three classes.

Limpo *et al.* (1995) have found that for chloride concentrations between 3.5 and 5 M, the solubility of the zinc in the ZnCl<sub>2</sub>-NH<sub>4</sub>Cl-NH<sub>3</sub>-H<sub>2</sub>O system was determined by the solubility of the zinc diammine chloride. For chloride concentrations lower than 3.5 M, there were two zinc compounds, *viz.* zinc hydroxide and zinc hydroxychloride. They have concluded that the solubility of these two zinc compounds depended on the ratio of the total ammonia concentration to the total zinc concentration, ([NH<sub>3</sub>]/[Zn]).

A much richer presentation of these results can be obtained by projecting the experimental data onto the CVA biplot, as indicated in Figure 2. In this figure, the red, green and blue circles indicate the formation of Zn(NH<sub>3</sub>)<sub>2</sub>Cl<sub>2</sub>, Zn(OH)<sub>2</sub> and Zn(OH)<sub>1.6</sub>Cl<sub>0.4</sub> respectively. The alpha-bags delineate the innermost 50% data points within each class. Figure 2(a) shows the CVA biplot with 50%-bags.



(a)



(b)

Fig.2. A CVA biplot and 50%-bags of the hydrolysis of zinc chloride data set.

The three classes are separated very well: Although not shown here, the innermost 95% of the bivariate data points in any one of the classes do not overlap with the 5% most extreme bivariate data points in the remaining two classes. The biplot also indicates fair amounts of variation within classes. Furthermore, it is clear that the formation of  $\text{Zn}(\text{NH}_3)_2\text{Cl}_2$  (red circles) is determined by the chloride concentration  $[\text{Cl}^-]$  in the range 3.5-5 M, while the formation of  $\text{Zn}(\text{OH})_2$  (green circles) and  $\text{Zn}(\text{OH})_{1.6}\text{Cl}_{0.4}$  are determined by the ammonia  $[\text{NH}_3]$  and zinc  $[\text{Zn}^{2+}]$  concentrations when the chloride concentration is below approximately 3.5 M. The centroid of each region is indicated by a solid square.

Figure 2(b) shows the CVA biplot equipped with classification regions, while Table 1 is the confusion matrix resulting from performing a CVA biplot classification of the data. It follows from Table 1 that the biplot classification delineating the classification

regions in Figure 2(b), allows classification of each phase  $\text{Zn}(\text{NH}_3)_2\text{Cl}_2$ ,  $\text{Zn}(\text{OH})_2$  and  $\text{Zn}(\text{OH})_{1.6}\text{Cl}_{0.4}$  with 88.0%, 94.2% and 95.2% correct classifications respectively. The overall correct classification is 93.1%. The mixed precipitates were not included in the construction of the classification map portrayed in Figure 2(b). Table 2 contains the results when these mixed precipitates were classified according to the CVA biplot procedure. In this table the actual phase  $\theta$  represents  $\text{Zn}(\text{NH}_3)_2\text{Cl}_2$ - $\text{Zn}(\text{OH})_{1.6}\text{Cl}_{0.4}$ , while the phase  $\gamma$  represents  $\text{Zn}(\text{OH})_2$ - $\text{Zn}(\text{OH})_{1.6}\text{Cl}_{0.4}$ .

Table 1: Classification of solid precipitates with linear discriminant analysis.

Predicted Phase	Actual Phase		
	$\text{Zn}(\text{NH}_3)_2\text{Cl}_2$	$\text{Zn}(\text{OH})_2$	$\text{Zn}(\text{OH})_{1.6}\text{Cl}_{0.4}$
$\text{Zn}(\text{NH}_3)_2\text{Cl}_2$	22	0	0
$\text{Zn}(\text{OH})_2$	2	33	2
$\text{Zn}(\text{OH})_{1.6}\text{Cl}_{0.4}$	1	2	40
% Correct classifications	88.0%	94.2%	95.2%

Table 2: Classification of the out-of-sample mixed precipitates with the CVA biplot classification procedure.

Temp (°C)	$[\text{Cl}^-]$ (M)	$[\text{Zn}^{2+}]$ (M)	$[\text{NH}_3]$ (M)	Actual Phase	Classification
40	3.97	0.766	0.872	$\theta$	$\text{Zn}(\text{NH}_3)_2\text{Cl}_2$
40	4.07	0.804	0.867	$\theta$	$\text{Zn}(\text{NH}_3)_2\text{Cl}_2$
40	3.48	0.541	0.767	$\theta$	$\text{Zn}(\text{NH}_3)_2\text{Cl}_2$
30	2.28	0.209	0.250	$\gamma$	$\text{Zn}(\text{OH})_2$
30	2.29	0.294	0.890	$\gamma$	$\text{Zn}(\text{OH})_2$
30	2.27	0.233	0.676	$\gamma$	$\text{Zn}(\text{OH})_2$

Using this classification Figure 2(a) also shows the mixed precipitates interpolated into the CVA biplot. These mixed precipitates are shown as three solid red circles ( $\text{Zn}(\text{NH}_3)_2\text{Cl}_2$ - $\text{Zn}(\text{OH})_{1.6}\text{Cl}_{0.4}$ ) just outside the 50% region of the  $\text{Zn}(\text{NH}_3)_2\text{Cl}_2$  precipitates (delineated by a solid red line) and three solid green circles ( $\text{Zn}(\text{OH})_2$ - $\text{Zn}(\text{OH})_{1.6}\text{Cl}_{0.4}$ ) distributed across the region of the  $\text{Zn}(\text{OH})_2$  (delineated by a solid green line).

#### 4. CALCIUM CARBIDE FURNACE

Commercial calcium carbide is a grey to reddish-brown crystalline material consisting of a mixture of  $\text{CaC}_2$  and  $\text{CaO}$  together with impurities introduced by the charge components during manufacture. When brought into contact with water, carbide produces acetylene gas of high purity in accordance with the equation  $\text{CaC}_2 + 2\text{H}_2\text{O} \rightarrow \text{C}_2\text{H}_2 + \text{Ca}(\text{OH})_2$  and it is this property that is largely responsible for its importance as a chemical substance. The volume of acetylene generated is therefore used as a measure of carbide quality in the industry in preference to the percentage  $\text{CaC}_2$ . The volume of gas generated is referred to as the gas yield and is expressed in litres per kilogram at standard conditions of temperature and pressure. The uses of acetylene gas are well known, ranging from a source of illumination when the gas is burned in air to produce a bright white light, to the combustion of acetylene in oxygen to

provide a high temperature heat source for the welding and cutting of metals. Acetylene gas is also the basic material for the synthesis of many organic compounds including solvents, plasticisers, plastics and synthetic rubbers. Calcium carbide is produced by fusing together a mixture of coke (or coal) and lime in an electric furnace according to the reaction  $\text{CaO} + 3\text{C} \rightarrow \text{CaC}_2 + \text{CO}$ .

A data set reflecting process operation over an 8-month period was constructed, containing the following variables (monitored daily): *FURNLOD* = furnace load (ton), *POWERCON* = power consumption (MWh/ton), *RESIST* = average resistance underneath the three electrodes (mΩ), *CHARCOAL* = charcoal consumption (ton), *COKE* = coke consumption (ton), *ANTHRACT* = anthracite consumption (ton), *LIMUNDER* = lime characteristics (% underburnt), *LIMOVER* = lime characteristics (% overburnt), *LIM2O* = ratio of % lime underburnt to % lime overburnt, i.e.  $\text{LIM2O} = \text{LIMUNDER}/\text{LIMOVER}$ , *CARBPROD* = carbide production (ton), and *CARBGRAD* = carbide grade (litre/kg).

In Figure 3 a PCA biplot is constructed for displaying the variation in the data for all the process variables simultaneously in a single graph. The multiplication of *CARBPROD* and *CARBGRAD* was used to quantify the quality of the product. Observations associated with lower quality product are indicated in red and the observations associated with higher quality in green. Although the PCA biplot does not aim to optimally separate these two classes, the difference is already obvious from this graph with higher quality product corresponding to higher values of *FURNLOD*, *CHARCOAL* and *ANTHRACT* and lower values for *RESIST*.

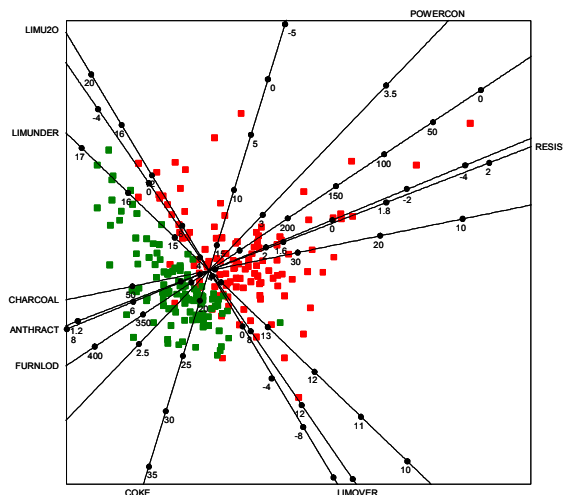


Fig.3. A PCA biplot of the calcium carbide furnace data set.

## 5. CONCLUSIONS

In this paper, biplots resulting from the modern perspective of Gower and Hand (1996) were introduced. In particular, the potential of these PCA

and CVA biplots to monitor multivariate process systems were demonstrated in two case studies, which focussed on the description of multidimensional variation and the separation of different classes of observations.

In the first case study, a CVA biplot could be used for visualising the hydrolysis of zinc chloride in ammonium chloride solutions. Different zinc precipitates could be classified in terms of different process conditions and mixed precipitates could then be assigned to the different classes by mapping them to the biplot.

In the second case study, a PCA biplot was likewise used to identify different operating regimes in a calcium carbide furnace. Once these operating regimes were defined, process conditions differentiating between the desirable and less desirable regimes could be identified with a CVA biplot and this information could then be used to formulate process control strategies for the plant.

Finally, different distance measures can be used with the biplot methodology introduced in this paper. In addition to the PCA and CVA biplots illustrated here, non-linear and generalised biplots can be constructed. These biplots enable users to display observations together with continuous as well as discrete or categorical variables. Moreover, Gardner (2001) has proposed several extensions of biplot methodology in statistical discrimination and classification to provide for non-linear relationships. It follows that biplot methodology is an extremely useful tool for analysing multivariate process behaviour.

## REFERENCES

- Aldrich, C. and Reuter, M.A., Monitoring of metallurgical reactors by the use of topographic mapping of process data. *Minerals Engineering*, 1999, **12**(11), 1301-1312.
- Figueiredo, J.M., Novais, A, Limpo, J.L. and Amer, S., The CENIM-LNETI process for the zinc recovery from complex sulphides. In: Matthew, I.G. (ed.), *Proceedings of the International Symposium on World Zinc '93*, AusIMM, Hobart, 1993, pp. 333-339.
- Gardner, S., *Extensions of biplot methodology to discriminant analysis with applications of non-parametric principal components*. Unpublished Ph.D. thesis, 2001, Department of Statistics and Actuarial Science, University of Stellenbosch, Stellenbosch, South Africa.
- Gabriel, K. R., The biplot graphical display of matrices with application to principal component analysis. *Biometrika*, 1971, **58**, 453-467.
- Gower, J.C., A general theory of biplots. In Krzanowski, W.J. (ed.), *Recent Advances in Descriptive Multivariate Analysis*, 1995, Clarendon Press, Oxford, pp. 283-303.

- Gower, J. C. and Hand, D. J., *Biplots*, 1996, Chapman & Hall, London.
- Limpo, J.L., Figueiredo, J.M., Amer, S. and Luis, A. The CENIM-LNETI process: A new process for the hydrometallurgical treatment of complex sulphides in ammonium chloride solutions. *Hydrometallurgy*, 1992, **28**, 149-161.
- Limpo, J.L., Luis, A. and Cristine, M.C., Hydrolysis of zinc chloride in aqueous ammoniacal ammonium chloride solutions. *Hydrometallurgy*, 1995, **38**, 235-243.
- Rousseeuw, P.J., Ruts, I. and Tukey, J.W. The bagplot: a bivariate boxplot. *The American Statistician*, 1999, **53**, 382-387.
- Tukey, J.W. Mathematics and the picturing of data. *Proceedings of the International Congress of Mathematicians*, 1975, **2**, 523-531.