Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 68(2) 2014 237-244

Outlier Detection through Independent Components for Non Gaussian Data

Asis Kumar Chattopadhyay and Saptarshi Mondal
Calcutta University, Kolkata

Received 19 September 2013; Revised 16 January 2014; Accepted 19 April 2014

SUMMARY

Observations lying "far away" from the main part of a data set and probably not following the assumed model may be termed as outliers.It is clear from the definition of outlier that significant presence of outlying observations may lead to erroneous results which in turn affects the statistical analysis of the data. So a very natural consequence of the above phenomenon will lead to the identification of outliers and eliminate them from the data set. Standard outlier detection techniques often fail to detect true outliers for massive, high dimensional and non Gaussian data sets. Several authors proposed different methods for this purpose. Filzmoser *et al*. (2008) proposed an algorithm using the properties of Principal Components to identify outliers in the transformed space. In the present work this method is modified by using Independent Components which is necessary for dealing with non Gaussian data. Primarily the dimension has been reduced through Independent Component Analysis and the proposed method has been applied in the reduced space in order to identify the outliers. The utility of the proposed method has been verified through massive, non Gaussian simulated data as well as real astronomical data related to Globular clusters of the Galaxy NGC 5128.

Keywords: Outlier identification, Independent components, Simulation, Globular clusters, Galaxy.

1. INTRODUCTION

The detection of outliers in multivariate data is considered to be an important and difficult problem. Most standard multivariate analysis techniques rely on the assumption of normality and require the use of estimates for both the location and scale parameters of the distribution. The presence of outliers may distort arbitrarily the values of these estimators and render meaningless results.This problem becomes much more critical for high dimensional massive data. Various concepts for outlier detection in Multivariate analysis exist in the literature. There are works by Atkinson and Mulira (1993), Bacon-Shone and Fung (1987), Bhandary (1992) etc. which are of heuristic nature and by Hara (1988), Caroni and Prescott (1992), Hawkins (1980) etc. of consecutive testing type. A different procedure was proposed by Rousseeuw (1985) based on the computation of the ellipsoid with the smallest volume or with the smallest covariance determinant that would encompass at least half of the data points. This procedure has been analyzed and extended in a large number of articles; see, for example, Rousseeuw and Leroy (1987), Davies (1987), Rousseeuw and van Zomeren (1990), Tyler (1994), Maronna and Yohai (1995), Hawkins and Olive (1999), Becker and Gather (1999), and Rousseeuw and van Driessen (1999).

There are several applications where high-dimensional outlier identification and/or robust

Corresponding author: Asis Kumar Chattopadhyay
E-mail address: akcstat@caluniv.ac.in

estimation are important. The field of Genetics, for instance, has recently received a lot of attention from statisticians. Advances in computing power have enabled biologists to record and store huge databases of information. Such information tends to contain a fair amount of gross errors, however, so robust methods are needed to prevent these errors from influencing the statistical model. Clearly, algorithms that take a long time to compute are not ideal or even practical for such large data sets. In addition, there is a further complication encountered in genetic data. The number of dimensions is typically several orders of magnitude larger than the number of observations, leading to a singular covariance matrix, so the majority of statistical procedures cannot be applied in the usual way. Similarly, Astronomy is another field in which outlier identification is useful; with the introduction of cheap electronic recording and storage devices it is not uncommon for data sets to be measured in terms of terabytes. It can thus be seen that there are a number of important applications in which current robust statistical models are impractical.

For high dimensional data reduction of dimension is also necessary for many data sets. One of the most recent powerful statistical techniques for analyzing such data sets is Independent Component Analysis (ICA). This technique is particularly applicable to non Gaussian data. The common problem is to find a suitable representation of the multivariate data. For the sake of computational and conceptual simplicity such representation is sought as a linear transformation of the original data. Principal Component Analysis, Factor Analysis, Canonical Correlation are some popular methods for linear transformation. But ICA is different from other methods, because it looks for the components in the representation that are both statistically independent and non Gaussian. In essence, ICA separates statistically independent component data, which is the original source data, from an observed set of data mixtures. All information in the multivariate datasets are not equally important. We need to extract the most useful information. Independent Component Analysis extracts and reveals useful hidden factors from the whole data sets. ICA defines a generative model for the observed multivariate data, which is typically given as a large database of samples. ICA can be applied in various fields like speech processing, brain imaging, stock predictions, signal separation, telecommunications, econometrics etc.

In the present work the problem is to identify outlying observations present in a non Gaussian large size data set with moderately large dimension. Attempts have been made to initially reduce the dimension through ICA and then to apply a method for detecting outliers in the reduced space by following the method proposed by Filzmoser *et al*. (2008) where the authors used projection pursuit to identify outliers on the basis of Principal Components. This investigation is organized as follows. We describe the method of Independent Component Analysis in Section 2, projection pursuit method in Section 3. In Sections 4, 5 and 6 we have discussed a simulation study, the Astronomical data set and data analysis respectively. Finally in Section 6, the concluding remarks are listed.

2. INDEPENDENT COMPONENT ANALYSIS

Independent Component Analysis (ICA) was most clearly stated by Comon (1994). Formally, the classical ICA model is of the form

$$X = AS, \tag{2.1}$$

where $X = [X_1, \ldots, X_m]'$ is a random vector of observations, $S = [S_1, \ldots, S_m]'$ is a random vector of hidden sources whose components are mutually independent and A is nonsingular mixing matrix. So A^{-1} is the unmixing matrix. Let we have n independently and identically distributed (i.i.d.) samples of X, say $\{X(j): 1 \le j \le n\}$. The main goal of ICA is to estimate the unmixing matrix A^{-1} and thus to recover hidden source using $S_k = A_k^{-1}X$, where A_k^{-1} is the k^{th} row of A^{-1}.

In the model, it is assumed that the data variables are linear or non-linear mixtures of some latent variables and the mixing system is also unknown. The latent variables are assumed non-Gaussian and mutually independent and they are called the independent components of the observed data. Suppose n random variables $X_1, \ldots, X_n$ are expressed as linear combinations of n random variables $S_1, \ldots, S_n$. Equation (2.1) can be written as:

$$X_i = a_{i1}S_1 + a_{i2}S_2 + \ldots + a_{in}S_n,\ i = 1, 2, \ldots, n \tag{2.2}$$

The S_i's are statistically mutually independent, where a_{ij}'s are the entries of the nonsingular matrix A. All we observe are the random variables X_i, and we have to estimate both the mixing coefficients a_{ij} and the independent components S_i, using the X_i.

There are many computer algorithms for performing ICA. A first step in those algorithms is to whiten (sphere) the data. This means that any correlations in the data are removed, *i.e.* the data are forced to be uncorrelated. Mathematically speaking, we need a linear transformation V such that $Z = VX$, where $E(ZZ') = I$. This can be easily accomplished by choosing $V = C^{-1/2}$, where $C = E(XX')$.

After sphering, the separated data can be found by an orthogonal transformation on the whitened data Z.

In ICA estimation, non-Gaussianity is very important. Without non-Gaussianity the estimation is not possible. Non-Gaussianity is motivated by the central limit theorem. Under certain conditions, the statistical distribution of a sum of independent random variables tends toward a Gaussian distribution. A sum of two independent random variables usually has a distribution that is closer to Gaussian than any of the two original random variables. Here,

$$X = AS,\ VX = VAS$$
$$\Rightarrow Z = (VA)S, \tag{2.3}$$

which implies that Z_i is closer to Gaussian than S_i. S_i is estimated by Z_i through maximization of non-Gaussianity. From equation (2.3) we can write

$$S = WZ, \tag{2.4}$$

where $W = (VA)^{-1}$.

We can measure non-Gaussianity by Negentropy (Hyvärinen *et al*. 2001). The entropy of a discrete variable is defined as the sum of the products of probability of each observation and the log of those probabilities. On the other hand, for a continuous function the entropy is called differential entropy which is given by the integral of the function times the log of the function. Negentropy is the difference between the differential entropy of a source S from the differential entropy of a Gaussian source with the same covariance of S. It is denoted by $J(S)$ and defined as follows:

$$J(S) = H(S_{Gauss}) - H(S), \tag{2.5}$$

where

$$H(S) = -\int p_s(\eta) log p_s(\eta) d\eta, \tag{2.6}$$

$p_s(\eta)$ is the density function of S. Negentropy is always non-negative, and it is zero if and only if S has a Gaussian distribution. Negentropy has an interesting property that it is invariant for invertible linear transformation. It is also a robust measure of non-Gaussianity. Here we estimate S by maximizing the distance of its entropy from Gaussian entropy as the noises are assumed to be Gaussian and if the signals are non-Gaussian only then they can be separated from the noise. If the signals are Gaussian, then ICA will not work.

2.1 Approximation of Negentropy

One drawback of negentropy is that it is very difficult to compute. That's why it needs to be approximated (Hyvärinen *et al*. 2001). The approximation is given by:

$$J(S) \propto (E[G(S)] - E[G(S_{Gauss})])^2, \tag{2.7}$$

where S_{Gauss} is a Gaussian random variable, G is a non-quadratic function. In particular, G should be so chosen that it does not grow too fast. Two popular choices of G are:

$$G_1(S) = \frac{1}{a} log\ cosh(aS) \tag{2.8}$$
$$G_2(S) = -e^{-S^2/2},$$

where $1 \le a \le 2$ is some suitable constant, which is often taken equal to 1.

2.2 The Fast ICA Algorithm

In this method the independent components are estimated one by one. This algorithm converges very fast and is very reliable. This algorithm is also very easy to use. We follow the following steps to perform the algorithm:

1. We center the data such that its mean becomes zero.
2. We whiten the data and denote it by Z.
3. We choose the number of Independent Components to be estimated and set $k = 1$.
4. We take an initial value of unit norm for W_k randomly, *i.e*., we initialize W_k, where W_k is the k^{th} row of W.
5. We set $W_k = E\{Zg(W'_kZ)\} - E\{g'(W'_kZ)\}W_k$, where g is derivative of G and G is defined as in (2.8).
6. We orthogonalize as:

$$W_k = W_k - \sum_{j=1}^{k-1} (W_k' W_j) W_j.$$

7. We set $W_k = W_k/\|W_k\|$.
8. If W_k does not converge, then we go back to step 5.
9. We set $k = k + 1$. If k does not exceed the number of independent components to be estimated, we go back to step 4.

Thus we find the estimated independent components.

3. PROPOSED METHOD

Outliers can be detected by using any one of the two basic approaches. First one is the distance based method where the aim is to detect outliers on the basis of values of similarity/dissimilarity measures to check how far a particular point is from the centre of the data. One can use Mahalanobis distance measure for this purpose. Under projection pursuit (Huber 1985) the target is to find suitable projections of the data in which the outliers are readily apparent and can be eliminated to obtain a robust estimator. This will help to find the outliers. Projection pursuit procedures are not affected by non Gaussianity and can be used for any type of data. Since Independent Component Analysis is developed for non-Gaussian data, it is expected that the outliers in non-Gaussian data will be more visible in independent component space than principal component space or original data space.

The algorithm for outlier detection in high dimension consists of two basic parts: a step for detecting location outliers, and a step for detecting scatter outliers. Scatter outliers possess a different scatter matrix, while a different location parameter describes location outliers (Filzmoser *et al*. 2008).

Let us consider a data set with n individuals and with p variables. To start, it is useful to perform the Independence Component Analysis (ICA) for reducing the dimensions, if the data set is non-Gaussian in nature. To find the optimum number of Independent Components (ICs), Principal Component Analysis (PCA) is performed. Only those eigenvectors/values are retained that explain a certain percent (which is 97% for the present Astronomical data set) of the total variance; suppose the new dimension of the data set is p^*. The remaining components are generally useless noise. For the case $p>>n$, there may be singularity problem, but this dimension reduction solves the singularity problem since usually $p^* < n$. Here the matrix of independent components is obtained as

$$S = WZ \tag{3.1}$$

These independent components are rescaled by the median and the Median Absolute deviation (MAD) as

$$s_{ij}^* = \frac{s_{ij} - med(s_{1j},...,s_{nj})}{MAD(s_{1j},...,s_{nj})},\ j = 1,...,p^*, \tag{3.2}$$

where $MAD(s_{1j}, ..., s_{nj}) = 1.4826 \times med_j|s_{ij} - med_i s_{ij}|$. This S^* is stored for the second phase of the algorithm. After that, the location outlier detection step is initiated by calculating the absolute value of a robust kurtosis measure for each component according to:

$$g_j = \left| \frac{1}{n}\sum_{i=1}^{n} \frac{(s_{ij}^* - med(s_{1j}^*,...,s_{nj}^*))^4}{MAD(s_{1j}^*,...,s_{nj}^*)^4} - 3 \right|,\ j = 1,...,p^*. \tag{3.3}$$

Both small and large values of the kurtosis coefficient can indicate outliers. This enables to assign weights to each component according to how likely we think it is to reveal the outliers.

At the end of the first phase of the algorithm, it is required to determine how large the robust Mahalanobis distance should be to obtain a better classification between outliers and non-outliers. The robust Mahalanobis distance can be calculated as:

$$RD_i = \sqrt{(s_i - T)' C^{-1} (s_i - T)}, \tag{3.4}$$

where T is a robust measure of location computed on the basis of independent components and C is a robust estimate of the covariance matrix of independent components.

In our study we have chosen the estimators of location and scale from the class of *S*-estimators (Marrona *et al.* 2006), defined as the vector T and positive definite symmetric matrix C that satisfy $min|C|$ subject to

$$\frac{1}{n}\sum_{i=1}^{n} \rho(d_i / c) = b_0, \tag{3.5}$$

$$d_i = \sqrt{(s_i - T)' C^{-1} (s_i - T)}, \tag{3.6}$$

where $\rho(\cdot)$ is a non decreasing function on $[0, \infty)$, and c and b_0 are tuning constants to be chosen properly.

According to Marona and Zamar (2002) we see that transforming the robust distances (RD_i) given by

$$d_i = RD_i \times \frac{\sqrt{\chi^2_{p^*,0.5}}}{med(RD_1,...,RD_n)}, i = 1,...,n \quad (3.7)$$

helped the empirical distances $\{d_i\}$ to have the same median as the theoretical distances and thus bring the former closer to $\chi^2_{p^*}$ where $\chi^2_{p^*,0.5}$ is the $\chi^2_{p^*}$ 50th quantile. The translated biweight function is utilized to assign weights to each observation and these weights are used as a measure of outlyingness. The translated biweight function is calculated as:

$$g_{1i} = \begin{cases} 0 & if\ d_i \geq c \\ (1-(\frac{d_i - M}{c - M})^2)^2 & if\ M < d_i < c \\ 1 & if\ d_i \leq M \end{cases} \quad (3.8)$$

where i = 1, ..., n, c = med(d_1, ..., d_n) + 2.5MAD(d_1, ..., d_n) and M is the $33\frac{1}{3}^{rd}$ quantile of the distances $\{d_1$, ..., $d_n\}$. These weights g_{1i} are stored and will be used at the end of the algorithm.

The second phase of our algorithm is similar to the first except that here outliers are searched in the space defined by S^* and here the translated biweight function is calculated by setting $c^2 = \chi^2_{p^*}$ 99th quantile and $M^2 = \chi^2_{p^*}$ 25th quantile. The weights calculated in this phase are denoted as g_{2i}, i = 1, ..., n.

Finally the weights from these two steps are combined to calculate final weights g_i, i = 1, ..., n, according to

$$g_i = \frac{(g_{1i} + k)(g_{2i} + k)}{(1 + k)^2}, \quad (3.9)$$

where typically the scaling constant k = 0.25. Outliers are then classified as points that have weight $g_i < 0.25$.

4. SIMULATION STUDY

In this section simulation results are presented in which the dimension has been taken as p = 100. The number of observations is n = 40000. 30000 observations have been generated from a multivariate Cauchy distribution and the rest 10000 observations from another multivariate Cauchy distribution with widely different parameters. Initially we tried to find out the outliers on the basis of Principal Components. The number of outliers was found to be 17057. The percentage of outlying observations is 42.64, where all the second set of 10000 observations generated from multivariate Cauchy are identified as outliers and remaining 7057 outliers are from the 1st set, *i.e*. from 30000 observations generated from multivariate Cauchy distribution. Next we have used the method on the basis of Independent Components. In this case we identified 14009 outliers of which all the second set of 10000 observations are outliers and remaining 4009 outliers are from the 1st set. The percentage of outlying observations is 35.02. It has also been checked that the nature of the outlying observations with respect to the values of the parameters is more closed to the actual set of outliers for the outlying observations obtained through independent components. The study indicates that the performance of the method is more powerful under the IC space than the PC space.

5. DATA SET

Our analysis is based on the sample of Globular Clusters (GCs) of the earlytype central giant elliptical galaxy in the Centaurus group, NGC 5128, whose structural parameters have been derived by fitting King and Wilson models to the surface brightness profiles based on HST/ACS imaging in the F606W bandpass (McLaughlin *et al*. 2008). The distance is that adopted by McLaughlin *et al.* (2008), namely 3.8 Mpc. The sample consists of 130 GCs whose available structural and photometric parameters are tidal radius (R_{tid}, in pc), core radius (R_c, in pc), half light radius (r_h, in pc), central volume density (logρ_0, in $M_\odot pc^{-3}$), $\sigma_{p,0}$ (predicted line of sight velocity dispersion at the cluster centre, in kms^{-1}), two body relaxation time at the model projected half mass radius (t_{rh}, in years), galactocentric radius (R_{gc}, in kpc), the concentration (c ~ log(R_{tid}/R_c)), the dimensionless central potential of the best fitting model (W_0), the extinction-corrected central surface brightness at F606W bandpass (μ_0 in mag arcsec^{-2}), V surface brightness averaged over r_h ($< \mu_v >_h$) in mag arcsec^{-2}, the integrated model mass (logM_{tot}, in $M_\odot$), Washington T_1 magnitude, extinction corrected color $(C - T_1)_0$ and metallicity determined from color $(C - T_1)_0$.

The radial velocities (V_r, in kms^{-1}) are available for 50 GCs (Woodley *et al*. 2007), the position angles (ψ, east of north) are available for all 130 GCs (Woodley *et al*. 2007). About 51 GCs are common with the sample of GCs observed by Beasley *et al.* (2008) with the sample used in the present case. Among these 33 GCs have published Lick Indices (Beasley *et al.*

2008). These data are used to derive the ages and metallicities ([Z/H]) of 33 GCs of our sample. The metallicities derived by us for the entire data set of GCs from Beasley *et al*. (2008) having published Lick indices and error bars (114 in numbers) is compared with that derived in Beasley *et al*. (2008) by means of a regression line which has a very good correlation ($r \sim 0.9$). The photometric metallicities are available for all the GCs of our sample.

The entire data set of 130 GCs with all the parameters (used from literature as well as derived by authors) are listed in Chattopadhyay *et al*. (2009). Although the number of observations is not very large we have used this data set due to its heteroscedastic and non Gaussian nature.

5.1 Test for Normality

Before applying ICA we have tested Gaussianity of the data set. In order to test the normality of the distribution pattern of a variable, the Shapiro-Wilk test is used. This test was published in 1965 by Samuel Shapiro and Martin Wilk. The null hypothesis of this test is that the data are normally distributed. The test statistic is $W = \frac{(\sum_{i=1}^{n} a_i x_{(i)})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$, where n is the number of observations, $x_{(i)}$ are the ordered sample values ($x_{(1)}$ is the smallest) and the a_i are constants generated from the means, variances and covariances of the order statistics of a sample of size n from a normal distribution.

In our case the data set used is multivariate. So, in this paper we have used the multivariate extension of the Shapiro-Wilk test (Alva and Estrada 2009). Here the null hypothesis is that the entire data set follows multivariate normal distribution. Let X_1, ... X_n be independently and identically distributed as p variate normal with mean vector μ and dispersion matrix Σ. Also let $\bar{X}$ and S be the sample mean vector and dispersion matrix. Define

$$Z_j^* = S^{1/2}(X_j - \bar{X}).$$

Then the coordinates of Z_j^* denoted by Z_{1j}, Z_{2j}, ..., Z_{pj} are approximately independently distributed as univariate standard normal. The test statistic proposed by Alva and Estrada is

$$W^* = \frac{1}{p}\sum_{i=1}^{p} W_{z_i},$$

where W_{z_i} is Shapiro-Wilk's statistic evaluated on the i^{th} coordinate of the transformed observations Z_{i1}, Z_{i2}, ... Z_{in}, $i = 1, 2, ... p$. They have also studied the null distribution of test statistic.

If the p-value is less than the chosen level of significance, the null hypothesis is rejected or in other words, the data are not multivariate normally distributed. We found that the p-value of the test was 4.231×10^{-14}, which is too small. Thus the null hypothesis has been rejected at 5% level of significance and we could conclude that the data set does not follow multivariate normal distribution.

6. DATA ANALYSIS

We have applied the proposed method both for Independent and Principal components. On the basis of the physical properties listed in Table 1 and Table 2, it is apparent that the separation of outliers (in fact outliers and non-outliers may be treated as two clusters here) is more apparent under IC space.

Table 1. Mean Values of the parameters of outliers and non-outliers found by ICA

	Outliers	Non-outliers
No.	9	121
R_{tid}	494.99 ± 85.60	216.18 ± 13.21
R_c	1.67 ± 0.22	1.51 ± 0.11
R_h	6.40 ± 0.72	4.22 ± 0.20
$\log\rho_0$	4.28 ± 0.31	3.68 ± 0.11
$\sigma_{p,0}$	13.60 ± 1.36	7.95 ± 0.46
t_{rh}	46321.44×10^5 ± 61022.40×10^4	16365.60×10^5 ± 98770.73×10^3
R_{gc}	8.76 ± 0.72	7.70 ± 0.39
c	2.45 ± 0.14	2.13 ± 0.04
W_0	7.24 ± 0.47	6.53 ± 0.08
μ_0	15.41 ± 0.48	16.89 ± 0.20
$<\mu_v>_h$	18.29 ± 0.34	18.64 ± 0.17
$\log M_{tot}$	5.78 ± 0.09	5.44 ± 0.04
T_1	19.24 ± 0.22	20.06 ± 0.11
$(C - T_1)_0$	1.41 ± 0.04	1.39 ± 0.03
$[Fe/H]$	–1.10 ± 0.10	–1.13 ± 0.08

Table 2. Mean Values of the parameters of outliers and non-outliers found by PCA

	Outliers	Non-outliers
No.	18	112
R_{tid}	560.44 ± 122.08	243.69 ± 16.70
R_c	1.78 ± 0.30	1.49 ± 0.10
R_h	7.27 ± 1.01	4.29 ± 0.18
$\log\rho_0$	4.20 ± 0.43	3.83 ± 0.11
$\sigma_{p,\,0}$	13.18 ± 1.86	9.23 ± 0.56
t_{rh}	$54212.37\times10^5 \pm 85351.30\times10^4$	$18967.23\times10^5 \pm 12333.33\times10^4$
R_{gc}	9.60 ± 0.88	7.56 ± 0.38
c	2.50 ± 0.19	2.17 ± 0.05
W_0	7.51 ± 0.64	6.55 ± 0.12
μ_0	15.49 ± 0.67	16.57 ± 0.21
$< \mu_v >_h$	18.63 ± 0.46	18.44 ± 0.16
$\log M_{tot}$	5.70 ± 0.12	5.53 ± 0.04
T_1	19.46 ± 0.30	19.81 ± 0.11
$(C - T_1)_0$	1.39 ± 0.06	1.40 ± 0.03
$[Fe/H]$	–1.19 ± 0.17	–1.09 ± 0.06

7. CONCLUSION

In the present work an attempt has been made to extend the projection pursuit method to the Independent Component space. Since it is very usual that most of the real data sets are non Gaussian in nature, it is always necessary to identify some appropriate method applicable under the non Gaussian situation. Both Independent Component Analysis and Principal Component Analysis are used for analyzing large data sets. Whereas ICA finds a set of source data that are mutually independent, PCA finds a set of data that are mutually uncorrelated. ICA was originally developed for separating mixed audio signals into independent sources. For non-Gaussian variables, the p.d.f.s need all moments to be specified, and higher order correlations must be taken into account to establish independence. It is expected that for non-Gaussian situation ICA will perform better than PCA in terms of the homogeneity of data corresponding to the different groups formed with respect to the important components. For PCA although primarily Gaussianity is not required but for making inference regarding eigen values and eigen vectors, Gaussianity is necessary as otherwise it will be too difficult. Further only under normality assumption principal components become independent. Hence the present approach may treated as considered as a good approach for searching outliers in non Normal data.

REFERENCES

Alva, J.A.V. and Estrada, E.G. (2009). A generalization of Shapiro-Wilk test for multivariate normality. *Comm. Statist.-Theory Methods*, **38,** 1870-1883.

Atkinson, A.C. and Mulira, H.M. (1993). The Stalactite plot for the detection of multivariate outliers. *Statist. Comput.,* **3**, 27-35.

Bacon-Shone, J. and Fung, W.K. (1987). A new graphical method for detecting single and multiple outliers in univariate and multivariate data. *Appl. Statist.*, **36,** 153-162.

Beasley, M.A. *et al*. (2008). A 2dF spectroscopic study of globular clusters in NGC 5128: probing the formation history of the nearest giant elliptical. *Monthly Notices of Royal Astronomical Society*, **386,** 1443-1463.

Becker, C. and Gather, U. (1999). The masking breakdown point of multivariate outliers. *J. Amer. Statist. Assoc.*, **94**, 947-955.

Bhandary, M. (1992). Detection of the number of outliers present in a data set using an information theoretic criterion. *Comm. Statist.-Theory Methods*, **21**, 3263-3274.

Caroni, C. and Prescott, P. (1992). Sequential application of Wilks's multivariate outlier test. *Technometrics*, **41,** 355-364.

Chattopadhyay, A.K., Chattopadhyay, T., Davoust, E., Mondal, S. and Sharina, M., (2009). Study of NGC 5128 globular clusters under multivariate statistical paradigm. *The Astrophysical J.*, **705**, 1533-1586.

Comon, P. (1994). Independent component analysis, A new concept? *Signal Processing*, **36,** 287-314.

Davies, P.L. (1987) Asymptotic behaviour of S estimates of multivariate location parameters and dispersion matrices. *Ann. Statist.*, **15,** 1269-1292.

Filzmoser, P., Marona, R. and Werner, M. (2008). Outlier identification in high dimension. *Comput. Statist. Data Anal.*, **52,** 1694-1711.

Hara, T. (1988). Detection of multivariate outliers with location slippage or scale inflation in left orthogonally invariant or elliptically contoured distributions. *Ann. Instt. Statist. Math.*, **40,** 395-406.

Hawkins, D.M. *Identification of Outliers*. Chapman and Hall, London.

Huber, P. (1985). Projection pursuit. *Ann. Statist.*, **13,** 435-475 (see also discussions by multiple authors following this article).

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

McLaughlin *et al.* (2008). Structural parameters for globular clusters in NGC 5128 - III. ACS surface brightness profiles and model fits. *Monthly Notices of Royal Astronomical Society*, **384,** 563-590.

Maronna, R., Martin, R. and Yohai, V. (2006). *Robust Statistics: Theory and Methods*. Wiley, NewYork.

Maronna, R.A. and Yohai, V.J. (1995). The behavior of the Stahel-Donoho robust multivariate estimator. *J. Amer. Statist. Assoc.*, **90,** 330-341.

Maronna, R. and Zamar, R. (2002). Robust estimates of location and dispersion for high-dimensional data sets. *Technometrics*, **44(4)**, 307-317.

Rousseeuw, P.J. (1985). Multivariate estimation with high breakdown point. In: *Mathematical Statistics and Applications*, (eds. W. Grossmann *et. al.*), 283-297.

Rousseeuw, P.J. and Leroy, A.M. (1987). *Robust Regression and Outlier Detection*. Wiley, New York.

Rousseeuw, P. and van Driessen, K. (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, **41,** 212-223.

Rousseeuw, P.J. and van Zomeren, B.C. (1990). Unmasking multivariate outliers and Leverage points. *J. Amer. Statist. Assoc.*, **85,** 633-639.

Shapiro, S.S. and Wilk, M.B. (1965). An analysis of variance test for normality (complete samples), *Biometrika*, **52,** 591-611.

Tyler, D.E. (1994). Finite sample breakdown points of projection based multivariate location and scatter statistics. *Ann. Statist.*, **22**, 1024-1044.

Woodley, K.A. *et al*. (2007). The kinematics and dynamics of the globular clusters and planetary nebulae of NGC 5128. *Astronomical J.*, **134,** 494-510.