Preprints of the 19th World Congress
The International Federation of Automatic Control
Cape Town, South Africa. August 24-29, 2014

The Application of Classification Methods to the Gross Error Detection Problem

Egardt F Gerber*. Lidia Auret. Chris Aldrich*****

*138 Sherwood Circle NW, CALGARY, AB, T3R 1R7, Canada (e-mail: efgerber@gmail.com)
**Department of Process Engineering, Stellenbosch University, Private Bag X1, MATIELAND, 7602, South Africa (e-mail: lauret@sun.ac.za)
*** Department of Minerals Engineering and Extractive Metallurgy, Western Australian School of Mines, Curtin University, PO Box 4198, WA, 6845, Australia (e-mail: Chris.Aldrich@curtin.edu.au)

Abstract: All process measurements are corrupted by the presence of measurement error to some degree. The attenuation of the measurement error, especially large gross errors, can increase the value of the process measurements. Gross error detection has typically been performed through rigorous statistical hypothesis testing. The assumptions required to derive the necessary statistical properties are restrictive, which lead to investigation of alternative approaches, such as artificial neural networks. This paper reports the results of an investigation into the utility of classification trees and linear and quadratic classification functions for resolving the gross error detection and identification problems.

***Keywords**:* gross error detection, identification, classification trees, classification functions

1. INTRODUCTION

The monitoring, control and optimization of modern industrial processes require the use of measurements of process variables in order to determine process states (Mah, 1990: 385).

All process measurements are corrupted to some degree by the presence of measurement error, so that measurement data do not conform to principles of mass and energy conservation, or other physical constraints pertaining to a particular system. A distinction is made between random noise that can only be described probabilistically and has no assignable cause, and gross or systematic errors (GE's) caused by events such as fouling or poor calibration of instruments, mechanical or electrical failures, etc. (Madron, 1992: 66-74). It is typically accepted that the GE component will have larger magnitude than the random error component, and be non-random to some degree. Depending on the intended use of process measurements, the presence of measurement error may impact adversely on the utility of the measurements, and cause suboptimal or even unsafe process operation. Methods of attenuating the measurement error, especially GE's, are therefore required.

The usual approach to the above problem is to find the adjustments to the process measurements required for the reconciled measurements to verify an imposed process model – typically mass or energy conservation laws. Mathematical programming techniques are used to find the adjustments that are optimal based on the assumed random error model (Narasimhan and Jordache, 2000: 60). GE's are detected based on statistical distributions derived from the random error model for the measurement adjustments or process constraint residuals and the data reconciliation (DR) may have to be performed iteratively until no more GE's are detected. Several difficulties exist with this approach. First, the statistics derived for gross error detection (GED) are strictly valid only for linear process constraints. Second, large systems or dynamic processes may require considerable computational effort to solve.

These difficulties have led to alternative approaches to the DR and GED problems, primarily the application of artificial neural networks (ANN's) (Terry *et al.*, 1993; Aldrich *et al.*, 1994a). While ANN's have been found useful in resolving the DR and GED problems, they are typically difficult to interpret, and only applicable to the systems used to develop them. The GED problem is inherently a classification problem. The success of ANN's in resolving this classification problem poses the question whether other classification methodologies can be applied successfully to the gross error detection and identification problem. This study investigated the utility of two classification methodologies, classification trees and classification functions, in resolving the gross error detection/identification problem.

The layout of this report is as follows: Section 2 introduces classical data reconciliation theory, while section 3 presents theoretical aspects of the classification methods employed. Section 4 provides a description of the case study methodology, and section 5 summarizes the observed results.

2. CLASSICAL DATA RECONCILIATION

Measurement errors are typically assumed to be additive and Gaussian, with zero mean value and known covariance structure, so that the measurement vector is given by (Romagnoli and Sanchez, 2000: 76):

$$\mathbf{y} = \mathbf{x} + \boldsymbol{\varepsilon} \tag{1}$$

Where: $\mathbf{y}$ is the vector of process measurements, $\mathbf{x}$ is the associated vector of state variables and $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{Q}_\varepsilon)$ is a vector of random measurement errors with covariance matrix $\mathbf{Q}_\varepsilon$. For linear systems with all variables measured, the process constraints are given by:

$$\mathbf{A}\cdot\mathbf{x}=\mathbf{0} \tag{2}$$

In this paper it is assumed that $\mathbf{A}$ has m rows and n columns, with $m<n$, and $\mathbf{y},\mathbf{x}$ and $\boldsymbol{\varepsilon}$ are $n\times 1$ vectors. The process constraints are typically not satisfied by the raw measurements due to the presence of measurement error, and the vector of residuals of the constraint equations is given by:

$$\mathbf{r}=\mathbf{A}\cdot\mathbf{y} \tag{3}$$

The Global Test tests for the presence of GE's by constructing $\tau=\mathbf{r}^T(\mathbf{A}\mathbf{Q}_\varepsilon\mathbf{A}^T)^{-1}\mathbf{r}\sim\chi^2_m$ which follows a chi-square distribution with m degrees of freedom under the null hypothesis of no GE's present (Romagnoli and Sanchez, 2000: 112). Under the measurement model of (1) the residuals are distributed as $\mathbf{r}\sim N(\mathbf{0},\mathbf{A}\mathbf{Q}_\varepsilon\mathbf{A}^T)$. The Nodal Test evaluates each individual constraint:

$$(z_r)_j=\frac{|r_j|}{\sqrt{(\mathbf{A}\mathbf{Q}_\varepsilon\mathbf{A}^T)_{jj}}} \tag{4}$$

Under the null hypothesis of no GE's present, $(z_r)_j=N(0,1)$ (Crowe, 1989). These tests can be completed before data reconciliation is performed.

When multiple tests are performed, as in the case of the Nodal Test, it is necessary to modify the univariate level of significance in order to preserve the overall level of significance. The Sidak correction (Sidak, 1967) can be used to calculate the required univariate level of significance from the desired overall level:

$$\alpha_i=1-(1-\alpha_0)^{1/m} \tag{5}$$

Where m univariate tests are conducted in the case of the Nodal Test, α_i is the univariate level of significance, and α_0 is the desired overall level of significance. This correction is conservative, i.e. the actual level of significance should be smaller than α_0.

The optimum estimates of the true process states are obtained by solving the following optimization problem:

$$\begin{aligned} &J=\min_{\mathbf{x}}(\mathbf{y}-\mathbf{x})^T\mathbf{Q}_\varepsilon^{-1}(\mathbf{y}-\mathbf{x})\\ &s.t.\\ &\mathbf{A}\cdot\mathbf{x}=\mathbf{0}\end{aligned} \tag{6}$$

The solution to (6), which is the maximum likelihood solution under the assumption of Gaussian error, is given by:

$$\hat{\mathbf{x}}=[\mathbf{I}-\mathbf{Q}_\varepsilon\mathbf{A}^T(\mathbf{A}\mathbf{Q}_\varepsilon\mathbf{A}^T)^{-1}\mathbf{A}]\cdot\mathbf{y} \tag{7}$$

The vector of measurement adjustments are given by:

$$\mathbf{a}=\mathbf{y}-\hat{\mathbf{x}}=\mathbf{Q}_\varepsilon\mathbf{A}^T(\mathbf{A}\mathbf{Q}_\varepsilon\mathbf{A}^T)^{-1}\mathbf{A}\mathbf{y} \tag{8}$$

Under (1) the vector of adjustments are distributed as $\mathbf{a}\sim N(\mathbf{0},\mathbf{Q}_\mathbf{a})$ where $\mathbf{Q}_\mathbf{a}=\mathbf{Q}_\varepsilon\mathbf{A}^T(\mathbf{A}\mathbf{Q}_\varepsilon\mathbf{A}^T)^{-1}\mathbf{A}\mathbf{Q}_\varepsilon$. The Measurement Test (MT) evaluates the standardized elements of the adjustment vector:

$$(z_a)_i=\frac{|a_i|}{\sqrt{(\mathbf{Q}_\mathbf{a})_{ii}}} \tag{9}$$

Under the null hypothesis of no gross errors present, $z_a\sim N(0,1)$ (Mah and Tamhane, 1982). Since the elements of $\mathbf{a}$ are dependent, it is typical that only $l<n$ unique values are obtained when applying (9).

The measurement test is commonly applied in commercial software packages used for data reconciliation, and was therefore selected as the benchmark for classification methods in this study.

There are four elements to a GED strategy (Narasimhan and Jordache, 2000: 175), namely: detection of the presence of one or more GE's; identification of the type and location of a detected GE; estimation of the magnitude of GE's; and handling multiple GE's. This study focussed on the first two elements, that is on the Detection (DET) and Identification (ID) problems of GED.

3. CLASSIFICATION METHODS

This section will briefly summarize the relevant information regarding the classification techniques investigated in this study.

3.1 Classification Trees

Classification Trees (CT's) are classification rules that are constructed by finding successive disjoint partitions of the input space X that improves the purity of the class membership in each partition. The input space consists of vectors of measurements that can be used to discern class separation. For classification trees, X can contain both ordered and categorical variables.

A tree T therefore consists of a set of nodes $\{t_i\}$ where each node represents a partition of the input space, such that $t_i\subset X$ and $\bigcup_{\forall i}t_i=X$. Construction of the tree T requires a finite number of training vectors $\mathbf{x}_j^k\in X,\ j=1,2,...,n_j^k$ with known class membership $k=1,2,...,n_k$. The purity of a node is a function of the proportion of the different classes contained in the node, i.e.:

$$s(k|t_i)=n_{t_i}^k/n_{t_i} \tag{10}$$

Where $n_{t_i}^k$ is the number of vectors with class membership k in node t_i and n_{t_i} is the total number of vectors in node t_i. Measures of purity on a partition t_i can take various functional forms $\phi(\cdot)$, but needs to meet the following criteria [Breiman *et al.*, 1984: 32]: $\phi(\cdot)$ should be nonnegative; $\phi(\cdot)$ should reach a maximum when $p(k|t_i)=1/n_k,\ \forall k\in t_i$; $p(k|t_i)=1,\forall k\in t_i\Rightarrow\phi(\cdot)=0$, i.e. $\phi(\cdot)$ is zero when any node proportion equals unity; $\phi(\cdot)$ should be symmetric; $\phi''(\cdot)$ exists and $\phi''(\cdot)<0$, i.e. the second derivative exists and the purity function is concave. One example of a purity function that meets these criteria is the Gini index:

$$g(t) = \sum_{k_1 \neq k_2} p(k_1 | t) \cdot p(k_2 | t) \quad (11)$$

Tree construction proceeds by recursively finding the partition of each parent node that maximises the increase in purity of the two child nodes, starting at the root node $t_0 = X$ which contains the complete input space. Partitioning of a node t_i terminates when the node is pure or when some *a priori* criteria for minimum node size are met. The class of a terminal node can be assigned based on the maximum class proportion present in the node, or based on a misclassification cost function.

This mode of construction typically results in large trees that are prone to overfitting of the training data, i.e. the trees fail to generalize the classification relationship, and will perform poorly on similar data not employed for tree construction. It is therefore required to prune the tree so that overfitting is minimized. Pruning is typically achieved by selecting the smallest tree (i.e. least number of nodes) for which the estimated error of misclassification falls within one standard error of the minimum estimated misclassification rate achieved. Details regarding estimation of misclassification rates can be found in Breiman *et al.* (1984).

No assumptions regarding the distribution of the $\mathbf{x}_j^k \in X,\ j = 1,2,...,n_j^k$ is made during construction of the CT, therefore it is a non-parametric classifier.

3.2 Classification Functions

Classification functions assign vectors to a class based on the standardized distance of the vector from the known or estimated class mean. Whereas CT's can handle any data type, classification functions require ordered data.

A training sample of vectors with known class membership $\mathbf{x}_j^k,\ j = 1,2,...,n_j^k$ is required so that estimates of the mean group vectors $\overline{\mathbf{x}}_k;\ k = 1,2,...,n_k$ can be obtained. The sample covariance for each group is calculated by:

$$\mathbf{S}_k = \frac{1}{n_j^k - 1} \sum_{j=1}^{n_j^k} (\mathbf{x}_j^k - \overline{\mathbf{x}}_k)^T (\mathbf{x}_j^k - \overline{\mathbf{x}}_k) \quad (12)$$

The pooled covariance estimate $\mathbf{S}$ is estimated from:

$$\mathbf{S} = \frac{1}{N - n_k} \sum_{k=1}^{n_k} (n_j^k - 1)\mathbf{S}_k \quad (13)$$

Where: $N = \sum_{k=1}^{n_k} n_j^k$ and $\mathbf{S}_k$ is the sample covariance matrix from (12). The pooled covariance $\mathbf{S}$ is an estimate of the population covariance matrix, and its use in the classification function implies that all classes have a common covariance matrix, that is:

$$E(\mathbf{S}) = E(\mathbf{S_1}) = E(\mathbf{S_2}) = ... = E(\mathbf{S_k}) \quad (14)$$

The distance from $\mathbf{x}$ to each $\overline{\mathbf{x}}_k$ is estimated with a distance function:

$$D_k^2(\mathbf{x}) = (\mathbf{x} - \overline{\mathbf{x}}_k)^T \mathbf{S}^{-1} (\mathbf{x} - \overline{\mathbf{x}}_k) \quad (15)$$

Rencher (2002: 304) shows that selecting the minimum of (15) is equivalent to selecting the maximum of a linear classification function of the form:

$$L_k(\mathbf{x}) = \mathbf{c}_k^T \cdot \mathbf{x} + c_{k,0} \quad (16)$$

Where: $\mathbf{c}_k = \overline{\mathbf{x}}_k^T \mathbf{S^{-1}}$ and $c_{k,0} = -0.5 \cdot \overline{\mathbf{x}}_k^T \mathbf{S}^{-1} \overline{\mathbf{x}}_k$. When prior probabilities $p_k;\ k = 1,2,...,n_k$ are associated with each group, the classification rule in (16) becomes:

$$L_k(\mathbf{x}) = \mathbf{c}_k^T \cdot \mathbf{x} + c_{k,0} + \ln(p_k) \quad (17)$$

Since the classification rules derived in (16) and (17) are linear functions of the vector $\mathbf{x}$, they are referred to as linear classification functions (LCF).

The use of a linear classification function depends on the assumption that the groups all have the same covariance matrix. When this is not true, the distance function in (15) can be modified by replacing the pooled estimate of the covariance matrix with the sample estimate of the group covariance matrix:

$$D_k^2(\mathbf{x}) = (\mathbf{x} - \overline{\mathbf{x}}_k)^T \mathbf{S}_k^{-1} (\mathbf{x} - \overline{\mathbf{x}}_k) \quad (18)$$

The classification rule remains to assign a vector $\mathbf{x}$ with unknown class membership to that group for which (18) is a minimum. This classification rule is a quadratic function in $\mathbf{x}$, and hence is called a quadratic classification function (QCF) (Rencher, 2002: 306).

If it is assumed that the $\mathbf{x}_j^k$ follow multivariate normal distributions, and that each group has a known prior probability p_k, the optimum classification rule is to assign $\mathbf{x}$ to the group that maximises:

$$Q_k = \ln(p_k) - 0.5 \cdot \ln|\mathbf{S}_k| - 0.5 \cdot (\mathbf{x} - \overline{\mathbf{x}}_k)^T \mathbf{S}_k^{-1} (\mathbf{x} - \overline{\mathbf{x}}_k) \quad (19)$$

Although classification functions were first derived without any distributional assumptions, the classification rules of (17) and (19) are optimal only for when $\mathbf{x}$ follows a multivariate normal distribution, hence classification functions are parametric classifiers.

4. CASE STUDY: TWO-PRODUCT SPLITTER

The process unit selected for this case study is a two-product splitter:

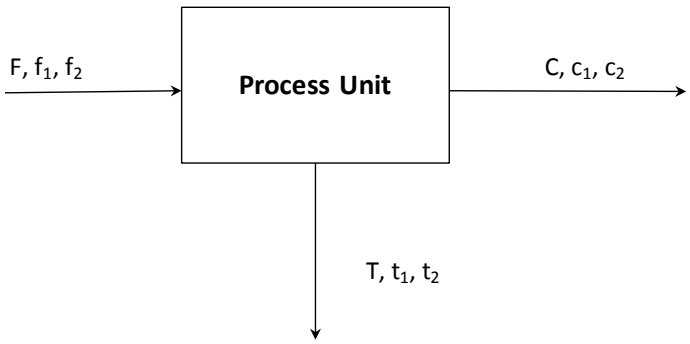

Figure 1 : Two-product Splitter for Case Study

Many process units can be described as two-product splitters, such as hydrocyclones, flotation cells, screens, filters and more. The constraints applied to this unit are given by:

$$F(\mathbf{x}) = \begin{bmatrix} F - C - T \\ F \cdot f_1 - C \cdot c_1 - T \cdot t_1 \\ F \cdot f_1 \cdot f_2 - C \cdot c_1 \cdot c_2 - T \cdot t_1 \cdot t_2 \end{bmatrix} = \mathbf{0} \qquad (20)$$

The constraints in (20) are more complex than the usual bilinear constraints commonly assumed in simulated studies such as these, and can arise when assay-in-size analysis is conducted.

Measurement vectors were generated from a fixed steady-state operating point, defined below in Table 1:

Table 1 : Nominal Steady-State Magnitudes of Variables

	F	C	T	f_1	c_1	t_1	f_2	c_2	t_2
TRUE	100.0	75.0	25.0	10.0%	5.0%	25.0%	10.0%	0.1%	15.9%
COV	2.0%	5.0%	5.0%	2.0%	2.0%	2.0%	1.0%	1.0%	1.0%

Random noise was added to each measurement value generated, with the coefficients of variation indicated in Table 1. Gross errors were simulated to a subset of the simulated measurement vectors, with gross error magnitudes controlled between [-5.0, 5.0] of the random error standard deviation in increments of 0.5. A total of 10 data sets, each containing 11 000 simulated measurement vectors of which 1 100 contained only random errors, was generated.

A naïve objective function for data reconciliation (DR) was assumed as:

$$\begin{aligned} &J = \min_{\mathbf{x}} (\mathbf{y} - \mathbf{x})^T \mathbf{Q}_\mathbf{y}^{-1} (\mathbf{y} - \mathbf{x}) \\ &s.t. \\ &F(\mathbf{x}) = \mathbf{0} \end{aligned} \qquad (21)$$

Where: $\mathbf{Q}_\mathbf{y} = \text{cov}_\mathbf{y} \cdot diag(\mathbf{y})$, where $diag(\mathbf{y})$ is a matrix of appropriate size with the elements of $\mathbf{y}$ on the diagonal and zero's elsewhere, assumes that each measurement has exactly the same relative standard deviation cov_y . This assumption played an important part in the comparison of classifier results with that of the MT, as it allowed manipulation of the type I error rate of the MT to match that of each classifier, so that statistical power of the methods could be compared directly. Data reconciliation was performed using successive linearization of the system in (21).

Constructing the CT and LCF/QCF classification functions requires a set of input vectors to be presented to each method. Two input vectors were developed for this problem, the first containing 14 elements derived from the DR results, and the second 16 elements. The first set of input vectors for each data set was defined as:

$$\mathbf{z}_{1,j}^i = \left[\mathbf{r}_j \quad \sqrt{\mathbf{r}_j^T \mathbf{r}_j} \quad \mathbf{a}_j^* \quad \sqrt{(\mathbf{a}_j^*)^T \mathbf{a}_j^*} \right]^T \qquad (22)$$

Where: $i = 1, 2, ..., 10$ denotes the i^{th} data set, $j = 1, 2, ..., 11000$ denotes the j^{th} set of DR results for data set i, $\mathbf{r}_j$ is the constraint residuals defined in (3), and $\mathbf{a}_j^* = 100 \cdot diag(\mathbf{y}_j)^{-1} \cdot \mathbf{a}_j$ is the normalized measurement adjustments, where $\mathbf{a}_j = \mathbf{y}_j - \hat{\mathbf{x}}_j$. The second set of input vectors was defined as:

$$\mathbf{z}_{2,j}^i = \left[\mathbf{z}_{1,j}^i \quad \sup_l \left| a_{l,j}^* \right| \quad (\mathbf{z}_{1,j}^i - \overline{\mathbf{z}}_{1,j}^i)^T \Sigma_{\mathbf{z}_1}^{-1} (\mathbf{z}_{1,j}^i - \overline{\mathbf{z}}_{1,j}^i) \right]^T \qquad (23)$$

Where: $\mathbf{z}_{1,j}^i$ is the j^{th} element of the i^{th} set of original input vectors defined in (22). That is, the second input vector contained all the elements of the first input vector, as well as two additional elements; the maximum absolute value contained in each $\mathbf{a}_j^*$, as well as the Mahalanobis distance for each of the original input vectors. Since ten data sets of simulated measurement data were created, a classifier of each type (classification tree, linear/quadratic classification function), for each problem (gross error detection/identification), could be developed for each data set, and tested on the remaining nine data sets. A separate classifier was developed for each of the Detection and Identification problems, with the classifier outputs for each problem defined as:

$$\begin{aligned} &\textit{Detection}: \quad Y^{j,i} \in \{0,1\} \\ &\textit{Identification}: \quad Y^{j,i} \in \{0,1,...,9\} \end{aligned} \qquad (24)$$

Where: $Y^{j,i} = 0$ corresponded to training exemplars containing only random errors and the other values of $Y^{j,i}$ correspond to either the presence (DET) or location (ID) of a gross error. In order to compare the results of the classifiers with that of the MT, the McNemar test for correlated proportions (McNemar, 1947) was used. The test is best illustrated by referring to the layout of Figure 2:

		Classifier		
		Success	Failure	Total
MT	Success	a	b	a + b
	Failure	c	d	c + d
	Total	a + c	b + d	n

Figure 2 : The McNemar Test for Correlated Proportions

The McNemar statistics is given by:

$$T_M = \frac{(b-c)^2}{b+c} \qquad (25)$$

Under the null hypothesis that no difference in performance exists between two methods compared, that is $H_0 : \pi_{a+b} = \pi_{a+c} \Rightarrow H_0 : \pi_b = \pi_c$, the McNemar statistic follows a chi-square distribution with one degree of freedom, which allows formal inference on the null hypothesis. The level of significance used for hypothesis testing throughout this study was 0.05. A statistical test may commit two different kinds of error: Type I error, which consists of rejecting the null hypothesis when it is true, and Type II error which consists of failing to reject the null hypothesis when it is false. These concepts are illustrated in Figure 3:

		H_0	
		TRUE	FALSE
Test	Fail to Reject H_0	Correct	Type II Error
	Reject H_0	Type I Error	Correct

Figure 3 : Type I and Type II Error associated with Hypothesis Testing

5. RESULTS AND DISCUSSION

The empirical Type II and Type I error rates for the MT and classifiers are presented below in Figure 4 for both the Detection and Identification problems:

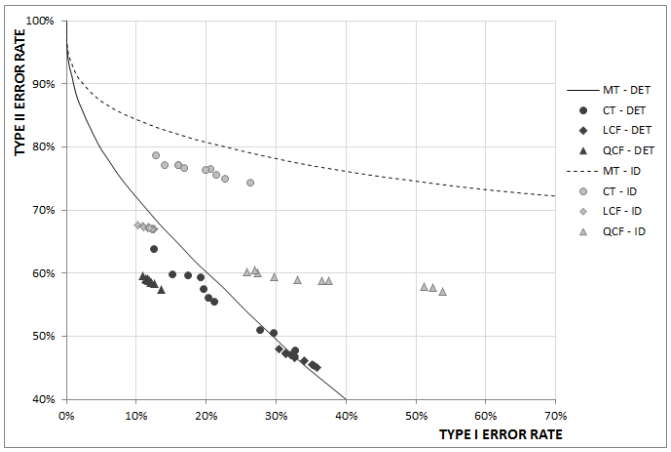

Figure 4 : Comparison of overall Type II vs. Type I Error Rates – First Input Vector

The low power (high Type II error) for both the Detection and Identification problems is apparent and is indicative of the complexity of this trilinear constraints. In general the classification methods performed similar to or better than the MT.

The classifiers' results exhibit a wide range of Type I error rates, both within the same classification methodology, and between different methodologies. The QCF seems to have the optimal trade-off between power and Type I error for the Detection problem, and the LCF for the Identification problem, although in practice what is 'optimal' in terms of this trade-off depends on the particular situation, e.g. a nuclear power plant would need to ensure a very low Type II error rate, while a mineral processing plant may want to have fewer false alarms, i.e. lower Type I error rate.

The McNemar Test as used to determine the significance of differences in power between the MT and each classifier on each of the 90 test sets, where the difference is defined as:

$$\Delta P = P_{Classifier} - P_{MT} = \beta_{MT} - \beta_{Classifier} \qquad (26)$$

Where: P represents statistical power, and β is the corresponding Type II error rate. That is, a negative ΔP indicates that the MT was superior, and vice versa. The number (Counts) and average values (Means) of differences determined to be significant are displayed in Table 2 and Table 3 respectively for different gross error magnitudes and locations:

Table 2 : Mean Values and Counts of differences in Power for different GE magnitudes determined to be significant by McNemar Test (upper half shows MT outperforming classification methods; lower half shows classification methods outperforming MT)

		DETECTION						IDENTIFICATION					
		Means			Counts			Means			Counts		
		CT	LCF	QCF	CT	LCF	QCF	CT	LCF	QCF	CT	LCF	QCF
GE Magnitude - b^*	0.5												
	1.0	-2.8%	-4.6%		1	16							
	1.5	-3.5%	-3.9%	-3.5%	6	31	90	-2.2%	-1.5%	-4.2%	5	35	72
	2.0	-6.0%	-4.6%	-3.7%	3	4	37	-2.8%			9		
	2.5												
	3.0												
	3.5												
	4.0												
	4.5												
	5.0												
GE Magnitude - b^*	0.5												
	1.0												
	1.5	3.7%			1								
	2.0	2.6%	3.4%		3	12		2.2%		3.3%	1		14
	2.5	4.5%	4.7%	8.5%	35	64	90	3.5%	11.0%	16.2%	31	90	90
	3.0	5.8%	5.9%	15.7%	66	77	90	5.3%	20.1%	24.8%	76	90	90
	3.5	7.0%	6.5%	19.1%	77	81	90	6.5%	26.1%	30.7%	86	90	90
	4.0	8.2%	6.9%	21.1%	80	87	90	7.0%	28.7%	34.1%	85	90	90
	4.5	8.7%	6.7%	21.8%	83	85	90	7.2%	28.4%	35.5%	71	90	90
	5.0	10.0%	7.3%	22.5%	80	83	90	6.9%	28.7%	35.0%	51	90	90

Table 3 : Mean Values and Counts of differences in Power for different GE locations determined to be significant by McNemar Test (upper half shows MT outperforming classification methods; lower half shows classification methods outperforming MT)

		DETECTION						IDENTIFICATION					
		Means			Counts			Means			Counts		
		CT	LCF	QCF	CT	LCF	QCF	CT	LCF	QCF	CT	LCF	QCF
GE Location	F	-8.1%	-3.6%		85	20		-13.2%			90		
	C	-8.4%	-7.8%	-8.7%	88	90	90	-8.9%	-4.2%	-8.0%	66	18	76
	T	-4.5%			28								
	f_1												
	c_1												
	t_1	-5.1%			4			-7.7%			90		
	f_2							-1.0%			2		
	c_2		-3.2%			2							
	t_2												
GE Location	F		3.5%	14.7%		1	90		25.2%	18.2%		90	90
	C												
	T	4.5%	3.3%	10.1%	17	17	90	13.6%	20.8%	16.5%	89	90	90
	f_1	12.3%	10.1%	20.2%	86	90	90	21.7%	32.2%	25.2%	90	90	90
	c_1	16.7%	15.8%	26.8%	90	90	90	11.3%	31.5%	35.3%	82	90	90
	t_1	5.4%	4.4%	7.2%	16	30	90		6.8%	9.6%		88	84
	f_2	12.4%	7.6%	16.4%	89	87	90	6.8%	20.8%	30.2%	72	90	90
	c_2									19.8%			90
	t_2	12.2%	6.2%	10.7%	87	72	90	5.1%	2.0%	23.2%	71	54	90

There is a general trend of the difference in power increasing with increasing GE magnitude, with the MT being superior for smaller magnitudes (1.0 – 2.0), and the classifiers superior for larger magnitudes (>2.0), which applies to both the Detection and Identification problems. Although the mean magnitude of outperformance of the MT at smaller GE magnitudes is fairly consistent for the different classifiers, the repeatability of this outperformance (as measured by the counts associated with the means) is markedly different for both the Detection and Identification problems. The repeatability of the QCF is generally highest, followed by the LCF and CT classifiers. This trend of repeatability also holds at larger magnitudes where the classifiers are superior.

The results for different gross error locations also exhibit general patterns; the MT is consistently superior for location

C for all classifiers, and consistently superior to the CT classifier for locations F and t_1 (for the Identification problem). Location c_2 presented unique results with none of the classifiers being able to resolve the Detection problem with higher power than the MT, while the MT achieved greater power than the LCF on this location with low repeatability. No significant results for the CT and LCF classifiers solving the Identification problem occurred, while the QCF achieved significantly higher power than the MT with high repeatability for the location.

In general then it can be concluded that the performance of a classifier relative to the MT is dependent on both the magnitude and location of a GE, as well as the type of classifier. This is true for both the Detection and Identification problems. The significance of type of classifier in determining performance must be due to the different principles employed to achieve classification. Although the classifiers collectively performed superior to the MT for the majority of GE magnitudes and locations, this does not guarantee superiority in general, as the MT may be superior for the particular requirements of an application, i.e. detecting and/or identifying GE's of a particular magnitude or in a particular location.

The low power achieved by classifiers lead to the application of a second input vector, as described in the section on methodology. The overall Type II error rate vs. Type I error rate for the classifiers using the second input vector is displayed in Figure 5, along with the results for the first input vector:

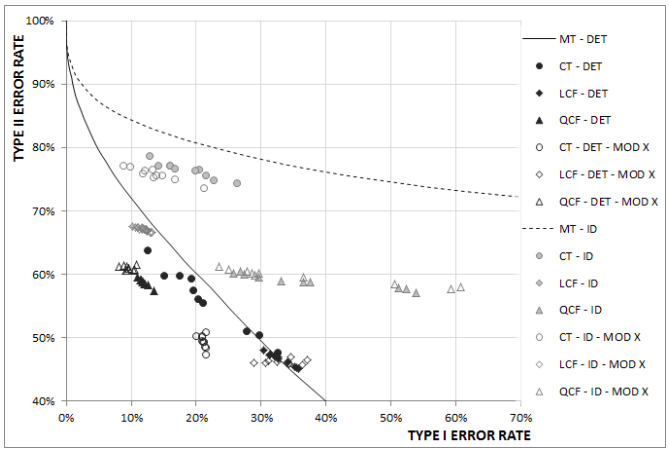

Figure 5 : Comparison of overall Type II vs. Type I Error Rates – Second Input Vector

The overall performance for the LCF and QCF classifiers for both the Detection and Identification problems using the second input vector remained virtually unchanged. The CT classifiers developed on the second input vector appear to have improved overall performance, i.e. lower Type II error rates for similar Type I error rates. This improvement in power is greater for the Detection problem than for the Identification problem. In addition, the variation in Type I error rates for the Detection problem decreased, i.e. the various CT classifiers developed for this problem seem to have more consistent performance. The detailed results of significant differences for the second input vector follows the same general trends as that of the first input vector, and is not presented.

6. CONCLUSIONS AND RECOMMENDATIONS

This investigation has shown that classification methodologies may be applied successfully to the Detection and Identification problems of GED, that is, classification methodologies may achieve lower Type II error rates than the Measurement Test for similar Type I error rates. However, future work on more complex (and simpler!) case studies would be required for a more definitive conclusion.

7. REFERENCES

Aldrich, C and van Deventer, J S J (1994) "The use of connectionist systems to reconcile inconsistent process data", *The Chemical Engineering Journal*, 54, 125-135

Breiman, L. Friedman, J H, Olshen, R A and Stone, C J (1984) *Classification and Regression Trees*, Chapman and Hall, Boca Raton

Crowe, C M (1989) "Test of Maximum Power for Detection of Gross Errors in Process Constraints", *American Institute of Chemical Engineers Journal"*, 35(5), 869-872

Madron, F (1992) *Process Plant Performance,* Ellis Horwood LTD, Chichester

Mah, R S H and Tamhane, A C (1982) "Detection of Gross Errors in Process Data", *American Institute of Chemical Engineers Journal*, 28(5), 828-830

Mah, R S H (1990) *Chemical Process Structures and Information Flow*, Butterworth Publishers, Stoneham

McNemar, Q (1947) "Note on the Sampling Error of the differences between correlated Proportions or Percentages", *Psychometrica*, 12 (2), 153-157

Narasimhan, S and Jordache, C (2000) Data Reconciliation and Gross Error Detection – An Intelligent Use of Process Data, Gulf Publishing Company, Houston

Rencher, A C (2002) *Methods of Multivariate Analysis* (*2nd Edition),* John Wiley & Sons, Inc., New Jersey

Romagnoli, J A and Sánchez, M C (2000) *Data Processing and Reconciliation for Chemical Process Operations*, Academic Press, San Diego

Sidak, Z (1967) "Rectangular Confidence Regions for the Means of Multivariate Normal Distributions", *Journal of the American Statistical Association*, 62 (318), 626-633

Terry, P A and Himmelblau, D M (1993) "Data Rectification and Gross Error Detection in a Steady-State Process via Artificial Neural Networks", *Industrial and Engineering Chemistry Research,* 32, 3020-3028