ROBUST AND EFFICIENT JOINT DATA RECONCILIATION – PARAMETER ESTIMATION USING A GENERALIZED OBJECTIVE FUNCTION

Yen Yen Joe[i), iii)] [1], David Wang[ii)], Jose Romagnoli[ii)] and Arthur Tay[i)]

[i)] *Dept. of Electrical & Computer Engineering, The National University of Singapore, 10 Kent Ridge Cres., Singapore 119260*

[ii)] *Dept. of Chemical Engineering, The University of Sydney, NSW 2006, Australia*

[iii)] *Institute of Chemical and Engineering Sciences, Ayer Rajah Cres., Block 28, Unit #02-08, Singapore 139959*

Abstract: This paper proposes the use of a generalized distribution, namely the Generalized T (GT) distribution in the joint estimation of process states and model parameters. The desirable properties of the GT-based estimator are its robustness, simplicity, flexibility and efficiency for the wide range of commonly encountered distributions (including Box-Tiao and t-distributions) which belong to the GT distribution family. To achieve the efficiency, the parameters of the GT distribution are adapted from the data through preliminary estimation. The strategy is applied to the virtual version of a practical chemical engineering plant.

Keywords: Parameter estimation, data processing, robust estimation, robustness, estimators, error correction

1. INTRODUCTION

Data reconciliation (DR) and parameter estimation (PE) are important components in process modelling, control, optimization and other applications that require reliable process data and accurate process model. Due to measurement errors, measurements of variables are usually equilibriated according to some conservation equations prior to their use in parameter estimation. In obtaining the model parameter estimates, however, the measurement data are further adjusted such that the model parameters and the adjusted data satisfy the process model equations. The inefficiency of the two-step DR and PE is hence noted. A remedy will be to couple the two procedures, thereby subjecting the parameter estimates and the measurement data to both the conservation equations and the process model, yielding estimates that are consistent to both sets of equations. Such estimates are expected to be more accurate. This is confirmed by the findings of MacDonald and Howat (1988), who examined both the sequential and coupled DR and PE and concluded that more reliable estimates are obtained using the coupled method.

The coupled DR and PE (DRPE) can also be viewed as the error-in-all-variables-measured (EVM) strategy, although DRPE is a more general formulation due to the inclusion of DR constraints that do not include unknown parameters. Mathematically speaking, the two strategies are similar in terms of issues that might arise in obtaining reliable estimates. Several aspects that have been discussed in the literature are: the general algorithm for the solution strategy (Valko and Vadja, 1987), the optimization strategy (Kim et al, 1990; Tjoa and Biegler, 1991) and the robustness of the estimation (Albuquerque and Biegler, 1996; Arora

[1] Author to whom correspondence should be addressed: g0202653@nus.edu.sg

and Biegler, 2001). Our proposed strategy will mainly deal with the robustness of the estimation.

The most prevalent form of estimator is the weighted least squares (WLS) formulation, where estimates minimizing the squared error in the measurements, normalized by the measurement covariance, are sought. Then again, it is also well known that the optimality of the weighted least squares is conditional upon the normality of the error distribution. Departure from normality, which could stem from the presence of gross errors and the empirical character of the error, often renders the weighted least square estimates unreliable. Attempts to treat these deviations include the a priori assignment of distributions with thicker tails, such as the bivariate Gaussian distribution (Tjoa and Biegler, 1991), or other particular forms of estimators to the same effect, i.e. ones that down-weight measurements with large residuals (e.g. the Fair Function by Albuquerque and Biegler, 1996; the redescending estimator in Arora and Biegler, 2001). An argument against pre-determined distributions is that the efficiency of the estimates may not be optimal, as the actual error distribution may not be close to the one assigned. If the parameters of these pre-determined forms of estimator are estimated posteriori, partially adaptive estimator results, and efficiency may be improved, although it will still be limited by the forms spanned by the possible values of the tuned parameters. A more flexible alternative is the non-parametric approach, where the estimator does not assume any fixed form and can fully adapt to the error distribution. The drawback of this approach is the relatively large number of unknown tuning parameters, and the possibility that it might not perform well for sample sizes frequently encountered in practice (Butler et al, 1990).

Our proposed approach is the robust DRPE strategy using a generalized distribution function, i.e. the Generalized T (GT) distribution. We will estimate the distribution parameters, and as the GT covers a wide range of commonly encountered distributions, it is expected to be more efficient than other partially adaptive estimators mentioned above. It is also simpler than the fully adaptive approach, as there are only a few tuning parameters. The GT-based estimator has been applied to parameter estimation (Butler et al, 1990) and data reconciliation (Wang and Romagnoli, 2003). In this paper, we extend the application to the more general and efficient DRPE strategy. The organization of the paper is as follows. The next section discusses about the robustness of M-estimators. Our proposed GT-based estimator is introduced here and its robustness properties discussed, along with that of two other estimators that we use for comparison in the case study, namely the weighted least square and the bivariate normal estimator. Next, Section 3 discusses the partial adaptiveness and the efficiency of the proposed strategy. In Section 4, the strategy is applied to an application case study and the results presented and discussed. Finally, Section 5 concludes the paper.

2. THE M-ESTIMATOR AND ITS ROBUSTNESS

Our proposed GT-based estimator falls under the category of M-estimator, which is the generalization of the maximum likelihood estimation and can be expressed as:

$$z = \arg\max f(u) = \arg\min -\log(f(u)) = \arg\min \rho(u) \quad \text{s.t. constraints} \tag{1}$$

where u = y-x, y=measurements, x=variables estimate, z=[x, p], p=model parameters to be estimated, f(u) is the probability density function (PDF) of u and $\rho(u) = -\log(f(u))$. It is of paramount importance, then, that the measurement error distribution indeed follows the PDF f(u), otherwise it will be forced to yield estimates that maximize f(u) when actually it does not resemble f(u), leading to inaccurate estimates.

The weighted least squares (WLS) is the most widely used estimator. Within the context of M-estimation, it corresponds to f(u) being the normal distribution, i.e. the error is assumed to be normally distributed.

To overcome the shortcomings of the WLS estimator in the event of gross errors, various robust estimators have been developed. Huber (1981) and Hampel et al (1986) provide a unified theoretical framework for statistical robustness, by means of which the various robust estimators can be analysed and compared. The analysis based on Hampel et al's Influence Function (IF) will be adopted in our analysis. To simplify presentation, the derivation of formulae will be omitted here; the interested reader could refer to Hampel et al (1986) for details. The IF aims to describe the behaviour of an estimator in the neighbourhood of the parametric distribution assumed. If the residual u is drawn from a PDF f(u) and if T[f(u)] is the unbiased estimate corresponding to u, then the IF of a residual u_0 is given by:

$$IF = \psi(u_0) = \lim_{t \to 0} \frac{T[(1-t)f + t\delta(u-u_0)] - T[f]}{t} \tag{2}$$

where $\delta(\mathrm{u}-\mathrm{u}_0)$ is the delta function centred about u_0. Thus, the IF quantifies the amount of influence that a residual u_0 has on the estimation. For M-estimators, the influence function IF is proportional to the derivative of $\rho(u)$, i.e.:

$$\mathrm{IF} = ?(u) \propto \partial\rho / \partial u = \partial \ln(f(u)) / \partial u \, . \tag{3}$$

In order for the estimator to be robust, the IF should be bounded such that a single large residual cannot dominate and distort the estimation. Additional desirable properties for the IF would be for it to be descending to very small values as the residual gets large, and to be continuous such that the estimator is well-behaved.

We will now use IF to analyse the robustness of a few estimators, starting from the weighted least squares (WLS). For WLS, $\rho(u) = u^{\mathrm{T}} \Phi^{-1} u$, where

Φ is the covariance matrix of the measurements and represents the weight. The IF of the WLS is then

$$\psi(u) = \ln \rho(u) = \ln(u^T \Phi^{-1} u) \propto u \quad (4)$$

which is a straight line and clearly unbounded for large values of residuals u (Figure 1). The amount of influence is proportional to the residual magnitude, allowing large residues to distort the estimation.

A robust alternative proposed by Tjoa and Biegler (1991) combines the normal PDF with another normal PDF with much larger variance to account for outliers. This results in the bivariate normal distribution with PDF:

$$f(u)=(1-p)\frac{1}{\sqrt{2\pi}\alpha}\exp(-\frac{u^2}{2\alpha^2})+p\frac{1}{\sqrt{2\pi}\alpha b}\exp(-\frac{u^2}{2\alpha^2 b^2}) \quad (5)$$

The PDF has two parameters: p is the probability of the gross error occurrence, and b is the ratio of the standard deviation of the larger normal PDF to that of the smaller one. The IF of this estimator is shown in Figure 1. The improvement in robustness can be observed: the IF behaves like a WLS for small residuals, but starts to descend after a certain point where the residuals are considered too large.

Our proposed GT distribution has the following mathematical form:

$$f_{GT}(u;\sigma,p,q)=\frac{p}{2\sigma q^{1/p}B(1/p,q)\left(1+\frac{|u|^p}{q\sigma^p}\right)^{q+1/p}};\ -\infty<u<\infty \quad (6)$$

It is symmetric about zero and unimodal. It is characterised by the distribution parameters $\{p,q,\sigma\}$: p and q determine the shape of the distribution, while σ determine the scale of the distribution. The GT defines a very general family of density function and combines two general forms that include as special cases most of the stochastic specifications encountered in practice. Some of the more commonly known special cases, along with the particular values of $\{p,q,\sigma\}$ for each case, are depicted in Figure 2.

The IF of the GT-based estimator can be obtained as:

$$\psi_{GT}(u;\sigma,p,q)=\frac{(pq+1)\,sign(u)|u|^{p-1}}{q\sigma^{1/p}+|u|^p} \quad (7)$$

Figure 3 shows the IF with several different sets of values of $\{p,q,\sigma\}$. It shows that generally the IF is bounded and actually descending when the residuals get large. However, we also observe that as p increases, the IF for large residuals increases and as q increases, the IF becomes less bounded. In fact, we notice that one special case where $q\rightarrow\infty$, with $p=2,\sigma=\alpha\sqrt{2}$ (α is the standard deviation) is none other than the normal distribution. To ensure that the GT-based estimator is insensitive to large residuals, therefore, bounds must be imposed on the values that p and q can take. In our work, $1\leq p\leq 5, 0.5\leq q\leq 50$.

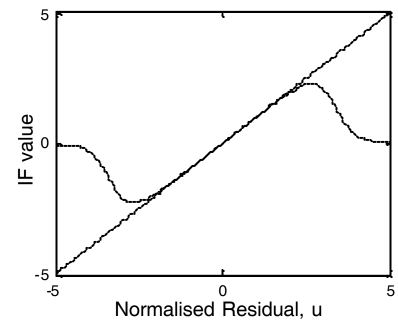

Figure 1. Plots of Influence Function for Weighted Least Square Estimator (dashed line) and the Robust Estimator based on Bivariate Normal Distribution (solid line)

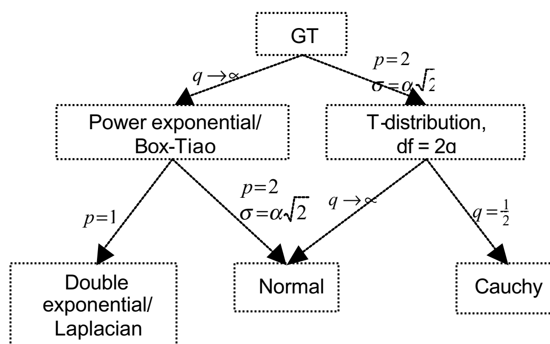

Figure 2. GT Distribution Family Tree, Depicting the Relationships among Some Special Cases of the GT Distribution

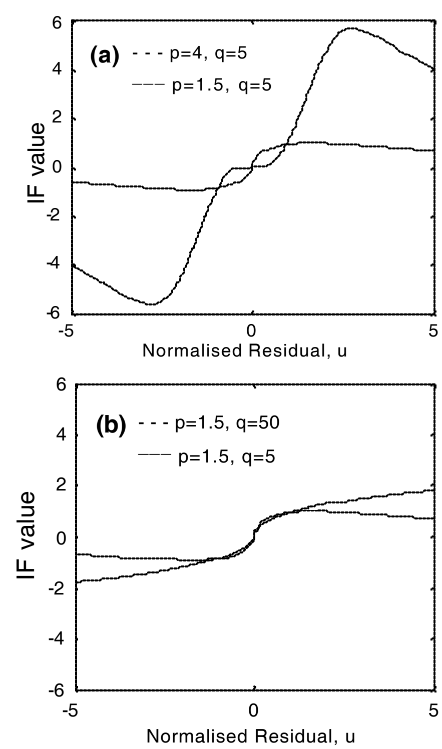

Figure 3. Plots of Influence Function for GT-based Estimator with different parameter settings

We will estimate the distributional parameters $\{p,q,\sigma\}$ using residuals obtained by performing a preliminary estimation on historical data. These residuals should have the same characteristics as the actual error, provided that the preliminary estimation is robust enough. As such, using the estimated $\{p,q,\sigma\}$ will ensure that the estimator characterises the error better, and in turn, yields more efficient estimates. The efficiency of this partially adaptive estimator will be further discussed in the next section.

3. PARTIALLY ADAPTIVE ESTIMATION AND EFFICIENCY

The motivation for partially adaptive estimation is to include the information about the error characteristics into the estimation. This inclusion of prior information corresponds to the formulation of posterior density which is a Bayes estimation. For our GT-based estimator, the distribution parameters $\{p,q,\sigma\}$ are estimated from the residuals of preliminary estimates. The residuals should be descriptive of the underlying error distribution, hence ensuring that the estimates of $\{p,q,\sigma\}$ result in the GT shape that best characterizes the error. It is therefore important for the preliminary estimator to be sufficiently robust so that the preliminary estimates are unbiased. In addition, the amount of historical data must also be sufficient to represent most of the data that are likely to be obtained. However, since there are only three distributional parameters to be tuned, the tuning is comparably simpler than that of non-parametric estimators.

For the GT-based approach, the distributional parameters $\{p,q,\sigma\}$ can be estimated using the maximum likelihood estimator (McDonald and Newey, 1988):

$$\{p,q,\sigma\} = \arg\max \sum \log f_{GT}(u_0;\{p,q,\sigma\}) \qquad (8)$$

where u_0 is the residual from the preliminary estimates. The values of $\{p,q,\sigma\}$ are then obtained as the parameters of a GT member from which the data are most likely sampled. The GT estimator using such estimated values of $\{p,q,\sigma\}$ has been shown to be asymptotically efficient among all estimators, when the error distribution is within the GT family. Rigorous mathematical proof can be found in McDonald and Newey (1988). Nothing can be said about the efficiency in the case of non-GT distributed errors, and some loss of efficiency is possible. However, since the GT family includes a wide range of commonly encountered distributions, the fact that it is asymptotically efficient for these distributions make its application very appealing.

In our approach, we use the maximum likelihood estimator in (8) to estimate $\{p,q,\sigma\}$. To obtain the residuals u_0 in (8), we take the median of the data set as the estimated values. Taking the median as the estimate corresponds to the use of robust L-estimator (Albuquerque and Biegler, 1996; Hampel et al., 1986). It is feasible in our case as we assume steady state measurements, i.e. the actual values are assumed to be constant over the time horizon considered. This method of estimating the distribution parameters is simple, robust and computationally more convenient, as compared to performing a full preliminary DRPE. Since the asymptotic distribution of the estimates depend only on the limit of the distribution parameters, and not on the particular way by which they are estimated (McDonald and Newey, 1988), our approach is justifiable. In cases where the measurements are not constant over the time horizon considered, we can assign fixed values of $\{p,q,\sigma\}$ that result in sufficiently robust GT estimator. In this case, the asymptotic efficiency is not guaranteed, but it is still robust and it is straightforward to apply.

4. APPLICATION CASE STUDY

The proposed GT-based DRPE strategy is applied to a case study of the pilot-scale setting of a general purpose plant containing two CSTRs, a mixer and a number of heat exchangers (Figure 4). Material feed from the feed tank is heated before being fed to the first reactor and the mixer. The effluent from first reactor is then mixed with the material feed in the mixer, and then fed to the second reactor. The effluent from the second reactor is, in turn, fed back to the feed tank and the cycle continues.

Steady-state analysis of the system structure results in three redundant equations involving seven redundant variables. The model parameters estimated are the product of the heat transfer coefficient with the effective heat transfer area of the cooling coil of the first reactor. For parameter estimation, two more model equations with five non-redundant variables are included.

Associated with the pilot-scale plant, a virtual environment that mimics the actual plant behaviour has been developed within the Matlab/Simulink framework and will be used in this paper while the plant is being commissioned. Simulation data are generated with several different distributions: Normal, Laplacian and t distribution. The different distributions are considered as outliers. A data set having Normal distribution and with large random shifts as gross error is also generated.

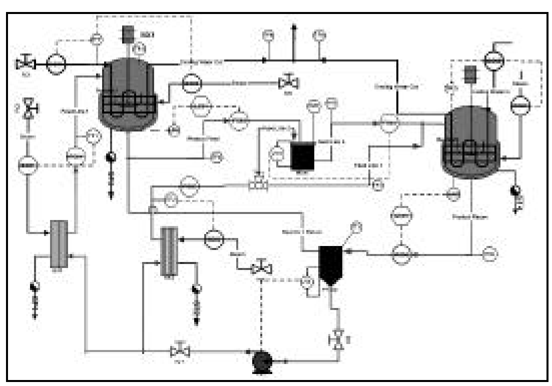

Figure 4. Flow Diagram of General Purpose Plant for Application Case Study

The different noise distributions are considered as outliers. From optimisation point of view, when an outlier is detected by the estimator, the variable is regarded as unmeasured. Consequently, depending on the structure of the constraint equations, there are certain variables that when regarded as unmeasured, become indeterminable. There is also a certain maximum number for outliers in a dataset, which if exceeded, will cause some variables to be indeterminable. Such issues will cause difficulty in the optimisation, and we are careful to avoid them in generating the outliers.

Five DRPE methods are compared, i.e. **(i)** weighted least squares (WLS); **(ii)** bivariate normal with distribution parameters fixed at p = 0.05, b = 10, and σ = standard deviation of the normal noise; this configuration is suggested in Tjoa and Biegler (1991) if the number of outliers are not known apriori, which is the condition we assume here; **(iii)** GT method with distribution parameters fixed at p = 1.5, q = 5, and σ = standard deviation of the normal noise multiplied by square root of two; this configuration is assigned based on our analysis on the influence function with different sets of parameter values, where the analysis shows that the current configuration will be sufficiently robust; **(iv)** partially adaptive bivariate normal with distribution parameters p, b and σ estimated from preliminary residuals; we obtain the preliminary residuals and estimate the distribution parameters in a similar way as we do for the partially adaptive GT method (as described in Section 3); and **(v)** partially adaptive GT method, with distribution parameters p, q and σ estimated from preliminary residuals.

Two measures are used to quantify the efficiency of the DRPE methods, i.e. the mean square error (MSE) and the percentage of model parameter estimation accuracy (%-discrepancy). The MSE is calculated as:

$$\text{MSE} = \frac{1}{\text{mK}} \sum_{j=1}^{\text{K}} \sum_{i=1}^{m} \frac{(\hat{x}_{i,j} - x_{i,j})^2}{\sigma_i^{\ 2}} \tag{9}$$

where m is the number of measured variables and K is the number of data sets used for the DRPE (m=12, K=24 in this study). $\hat{x}_{i,j}$ and $x_{i,j}$ are the estimates of the reconciled data and the actual value of the variable, respectively, while σ_i is the standard deviation of the Gaussian noise on sensor *i*. The %-accuracy is calculated as:

$$\%\,\text{discrepancy} = 100\% \times \frac{|\,\text{parameter estimate - actual value}\,|}{\text{actual value}} \tag{10}$$

Figure 5 and 6 chart the MSE and %-discrepancy, respectively. The partially adaptive GT method with performs most efficiently for all types of outliers considered, i.e. it has the smallest overall estimation errors for both parameter and states. The advantage of partially adaptive scheme is apparent by the lower estimation errors achieved by GT and bivariate methods with distribution parameters estimated from the data, as compared to their counterparts with distribution parameters assigned and fixed at certain values. However, although the partially adaptive bivariate normal method performs very well, it is seen that the partially adaptive GT method is still more efficient for Laplacian and t noise, which are within the GT family of distributions.

Comparing the non-partially adaptive GT and bivariate normal methods, it is seen that in this particular case study, GT generally performs better. However, the choice of distribution parameters greatly influences the estimation results, and since we have only one instance of all possible settings of distribution parameters, the results here cannot be generalized to the conclusion that all non-partially adaptive GT estimators perform better than all non-partially adaptive bivariate methods.

The results of the preliminary estimation for the GT distribution parameters for some variables when the error follows Cauchy and t distributions are listed in Table 2. Referring back to the distribution tree in Figure 2, we can see that the values of the estimated p and q are close to the ideal p and q for the respective distributions. For example, for Laplacian noise, where the ideal p=1 and $q \to \infty$, the estimated p are close to one, while q are large or close to the upper bound, i.e. q=50.

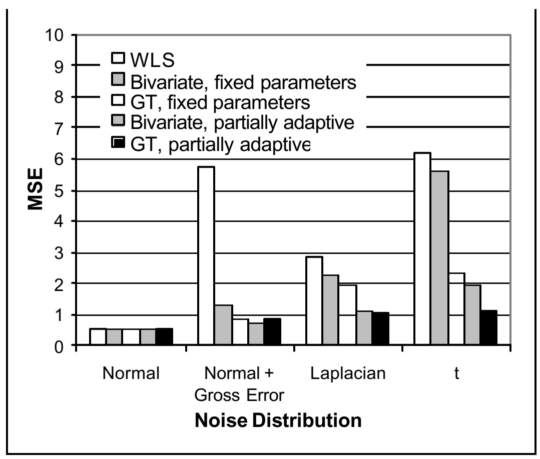

Figure 5. State Estimation Error Comparison of the WLS, Bivariate and GT-based Estimators for Different Noise Profiles

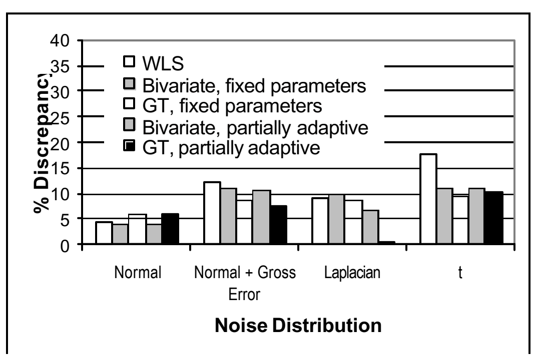

Figure 6. Parameter Estimation Accuracy Comparison of the WLS, Bivariate and GT-based Estimators for Different Noise Profiles

Table 2. GT Distribution Parameter Estimates

Variable	Laplacian (p=1,q→ inf)		t with 2 dof (p=2,q=1)	
	p	q	p	q
Feed Flow to Reactor 1	1.32	50.00	1.53	2.67
Feed Flow to Reactor 2	1.01	46.75	3.52	0.50
Cooling Water Flow of Reactor 1	1.15	47.71	3.13	0.62

To visualise how well the estimators characterize the data, the probability density functions of the distributions assumed by each estimator are plotted in Figure 6 (for the variable feed flow to reactor 1 in Table 2, in the case of t noise), along with the relative frequency histogram of the noise. The histogram represents the actual empirical error distribution. The PDF are plotted based on the distribution parameter values assumed by the estimators, i.e. normal $N(0,\sigma^2)$, where σ =standard deviation of the measurement (in this case = 0.5) for WLS; partially adaptive bivariate normal with parameters p=0.0546, b=5.0000, and σ =1.3240 for the bivariate estimator; and GT with $\{p,q,\sigma\}$= {1.5343, 2.6729, 1.4498} for the GT-based estimator. It is seen that the both GT and bivariate estimator characterizes the data very well. However, the GT shape resembles the data more, as it can assume the different shapes of distribution within the GT family, while the bivariate shape is limited to the shape of two normal PDFs combined.

5. CONCLUSION

The DRPE strategy based on GT distribution strikes a balance between the simplicity of strict parametric approaches where a certain distribution is assumed and the flexibility of non-parametric approaches which involves many unknown parameter approaches and may not perform well for common sample sizes in practice. The GT-based estimator is robust for a range of its distribution parameter values, by adapting these distribution parameter to the data, it is efficient particularly when the underlying error distribution is within the GT family. Within the range of its robust distribution parameter values, the GT family encompasses a wide variety of important statistical distributions, and thus it is a worthy candidate for DRPE as it can handle a wide range of error distributions with simple adaptations of its distribution parameters.

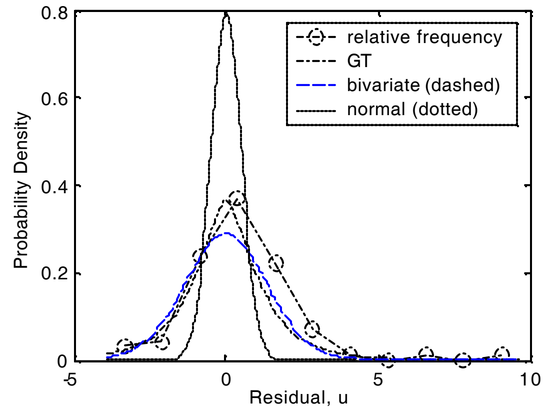

Figure 6. Plots of Relative Frequency Histogram and the Various Error Distributions Assumed by the Different Estimators

6. REFERENCES

Albuquerque, J. S., Biegler, L.T. (1996). Data Reconciliation and Gross-Error Detection for Dynamic Systems. *AIChE J.*, **Vol. 42**, No. 10, pp. 2841-2856.

Arora, N., Biegler, L.T. (2001). Redescending Estimators for Data Reconciliation and Parameter Estimation. Comp. Chem. Eng., **Vol. 25**, pp. 1585-1599.

Butler, R.J., McDonald, J.B., Nelson, R.D., White, S.B. (1990). Robust and Partially Adaptive Estimation of Regression Models. *Rev. Econ. Stat.*, **Vol 72**, No. 2, pp.321-327.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., Stahel, W.A. (1986). *Robust Statistics: The Approach Based on Influence Function.* Wiley, New York.

Huber, P.J. (1981). *Robust Statistics*. Wiley, New York.

Kim, I.W., Liebman, M.J., and Edgar, T.F. (1990). Robust Error-in-Variable Estimation Using Non-linear Programming techniques. *AIChE J*, **Vol. 36**, pp. 985-993.

MacDonald, R.J., Howat, C.S. (1988). *AIChe J*, **Vol. 34**, 1.

McDonald, J.B., Newey, W.K. (1988). Partially Adaptive Estimation of Regression Models via the Generalized T Distribution. *Econometric Theory*, **Vol. 4**, pp.428-457.

Tjoa, I.B., Biegler, L.T. (1991). Simultaneous Strategies for Data Reconciliation and Gross Error Detection of Nonlinear Systems. *Comp. Chem. Eng.*, **Vol. 15,** No. 10, pp. 679-690.

Valko, P., Vadja, S. (1987), An extended Marquardt-type Procedure for Fitting Error-in-Variable Models. *Comp. Chem. Eng.*, **Vol. 11**, pp. 37-43.

Wang, D., Romagnoli, J.A. (2003). A Framework for Robust Data Reconciliation Based on a Generalized Objective Function. *Ind.Eng.Chem.Res.*,**Vol.42**, No.13, pp. 3075-3084.