# Gross Error Detection in Chemical Plants and Refineries for On-Line Optimization

**Xueyu Chen, Derya B. Ozyurt and Ralph W. Pike**

**Louisiana State University**

**Baton Rouge, Louisiana**

**Thomas A. Hertwig**

**IMC Agrico Company**

**Convent, Louisiana**

**Jack R. Hopper and Carl L. Yaws**

**Lamar University**

**Beaumont, Texas**

Workshop on Systems Safety, LAWSS 2003, Jorge A. Aravena, Workshop Program Chair, sponsored by NSF and NASA, Baton Rouge, LA (February 28, 2003)

# INTRODUCTION

- o Status of on-line optimization
- o Theoretical evaluation of distribution functions used in NLP's
- o Numerical results support the theoretical evaluation
- o An optimal procedure for on-line optimization
- o Application to a Monsanto contact process
- o Interactive Windows program incorporating these methods

Mineral Processing Research Institute
web site
www.mpri.lsu.edu

# On-Line Optimization

Automatically adjust operating conditions with the plant's distributed control system

Maintains operations at optimal set points

Requires the solution of three NLP's

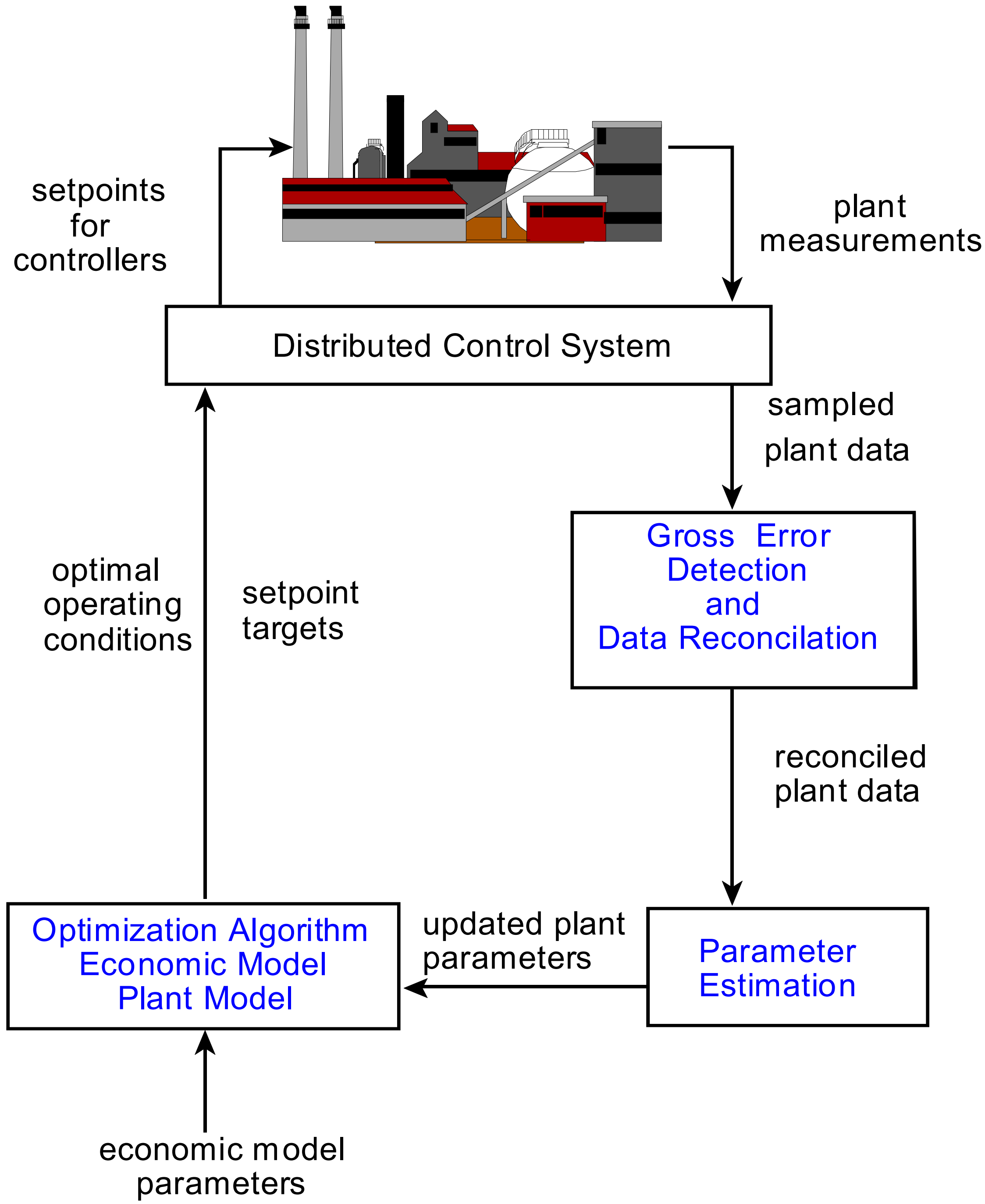
- gross error detection and data reconciliation
- parameter estimation
- economic optimization

## BENEFITS

Improves plant profit by 3-5%

Waste generation and energy use are reduced

Increased understanding of plant operations


setpoints
for
controllers
plant
measurements
Distributed Control System
sampled
plant data
Gross Error
Detection
and
Data Reconciliation
optimal
operating
conditions
setpoint
targets
reconciled
plant data
Optimization Algorithm
Economic Model
Plant Model
updated plant
parameters
Parameter
Estimation
economic model
parameters

## Some Companies Using On-Line Optimization

| United States | Europe |
| --- | --- |
| Texaco | OMV Deutschland |
| Amoco | Dow Benelux |
| Conoco | Shell |
| Lyondel | OEMV |
| Sunoco | Penex |
| Phillips | Borealis AB |
| Marathon | DSM-Hydrocarbons |
| Dow | |
| Chevron | |
| Pyrotec/KTI | |
| NOVA Chemicals (Canada) | |
| British Petroleum | |

## Applications

mainly crude units in refineries and ethylene plants

## Companies Providing On-Line Optimization

Aspen Technology - Aspen Plus On-Line
- DMC Corporation
- Setpoint
- Hyprotech Ltd.

Simulation Science - ROM
- Shell - Romeo

Profimatics - On-Opt
- Honeywell

Litwin Process Automation - FACS

DOT Products, Inc. - NOVA

## Distributed Control System

Runs control algorithm three times a second

Tags - contain about 20 values for each measurement, e.g. set point, limits, alarm

Refinery and large chemical plants have 5,000 - 10,000 tags

## Data Historian

Stores instantaneous values of measurements for each tag every five seconds or as specified.

Includes a relational data base for laboratory and other measurements not from the DCS

Values are stored for one year, and require hundreds of megabites

Information made available over a LAN in various forms, e.g. averages, Excel files.

# Plant Problem Size

| | Contact | Alkylation | Ethylene |
|---|---|---|---|
| Units | 14 | 76 | - |
| Streams | 35 | 110 | ~4,000 |
| Constraints | | | |
| Equality | 761 | 1579 | ~400,000 |
| Inequality | 28 | 50 | ~10,000 |
| Variables | | | |
| Measured | 43 | 125 | ~300 |
| Unmeasured | 732 | 1509 | ~10,000 |
| Parameters | 11 | 64 | ~100 |

# Status of Industrial Practice for On-Line Optimization

Steady state detection by time series screening

Gross error detection by time series screening

Data reconciliation by least squares

Parameter estimation by least squares

Economic optimization by standard methods

## Key Elements

Gross Error Detection

Data Reconciliation

Parameter Estimation

Economic Model
(Profit Function)

Plant Model
(Process Simulation)

Optimization Algorithm

# DATA RECONCILIATION

Adjust process data to satisfy material and energy balances.

Measurement error - **e**

$$\mathbf{e} = \mathbf{y} - \mathbf{x}$$

**y** = measured process variables
**x** = true values of the measured variables

$$\tilde{\mathbf{x}} = \mathbf{y} + \mathbf{a}$$

a - measurement adjustment

# DATA RECONCILIATION

measurements having only random errors - least squares

*Minimize*: $\min_{\mathbf{x}} \; \mathbf{e}^T \Sigma^{-1} \mathbf{e} = (\mathbf{y} - \mathbf{x})^T \Sigma^{-1} (\mathbf{y} - \mathbf{x})$

*Subject to*: $\mathbf{f}(\mathbf{x}) = 0$

$\Sigma$ = variance matrix = $\{\sigma^2_{ij}\}$.

$\sigma_i$ = standard deviation of $e_i$.

$\mathbf{f}(\mathbf{x})$ - process model
- linear or nonlinear

# DATA RECONCILIATION

Linear Constraint Equations - material balances only

$\mathbf{f}(\mathbf{x}) = \mathbf{A}\mathbf{x} = 0$

analytical solution - $\tilde{\mathbf{x}} = \mathbf{y} - \Sigma\mathbf{A}^T(\mathbf{A}\Sigma\mathbf{A}^T)^{-1}\mathbf{A}\mathbf{y}$

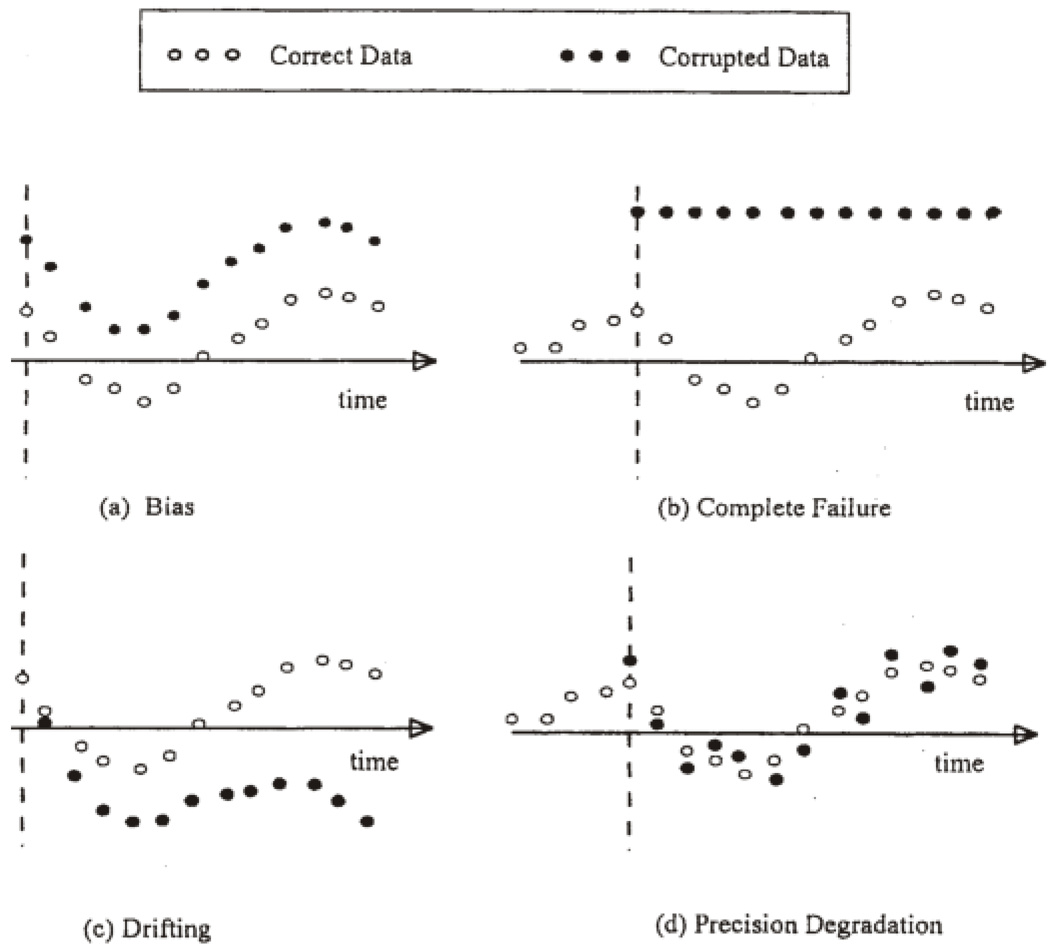
Nonlinear Constraint Equations

$\mathbf{f}(\mathbf{x})$ includes material and energy balances, chemical reaction rate equations, thermodynamic relations

nonlinear programming problem

GAMS and a solver, e.g. MINOS

# Types of Gross Errors




Source: S. Narasimhan and C. Jordache, *Data Reconciliation and Gross Error Detection*, Gulf Publishing Company, Houston, TX (2000)

# Gross Error Detection Methods

Statistical testing

- many methods
- can include data reconciliation

Others

- Principal Component Analysis
- Ad Hoc Procedures - Time series screening

# Combined Gross Error Detection and Data Reconciliation

Measurement Test Method - least squares

$$\underset{\mathbf{x},\,\mathbf{z}}{\textit{Minimize}}: \quad (\mathbf{y} - \mathbf{x})^T \Sigma^{-1} (\mathbf{y} - \mathbf{x}) = \mathbf{e}^T \boldsymbol{\Sigma}^{-1} \mathbf{e}$$

$$\textit{Subject to}: \quad \mathbf{f}(\mathbf{x}, \mathbf{z}, \boldsymbol{\theta}) = 0$$

$$\mathbf{x}^L \leq \mathbf{x} \leq \mathbf{x}^U$$

$$\mathbf{z}^L \leq \mathbf{z} \leq \mathbf{z}^U$$

Test statistic:

if $|e_i|/\sigma_i \geq C$ measurement contains a gross error

Least squares is based on only random errors being present

Gross errors cause numerical difficulties

Need methods that are not sensitive to gross errors

# Methods Insensitive to Gross Errors

## Tjao-Biegler's Contaminated Gaussian Distribution

$$P(y_i \mid x_i) = (1-\eta)P(y_i \mid x_i, R) + \eta\, P(y_i \mid x_i, G)$$

$P(y_i \mid x_i, R)$ = probability distribution function for the random error
$P(y_i \mid x_i, G)$ = probability distribution function for the gross error.
Gross error occur with probability $\eta$

Gross Error Distribution Function

$$P(y|x,G) = \frac{1}{\sqrt{2\pi}b\sigma} e^{\frac{-(y-x)^2}{2b^2\sigma^2}}$$

# Tjao-Biegler Method

Maximizing this distribution function of measurement errors or minimizing the negative logarithm subject to the constraints in plant model, i.e.,

Minimize: $\mathbf{x}$

$$-\sum_i \left\{ \ln\left[ (1-\eta)\, e^{\frac{-(y_i-x_i)^2}{2\sigma_i^2}} + \frac{\eta}{b} e^{\frac{-(y_i-x_i)^2}{2b^2\sigma_i^2}} \right] - \ln\left[\sqrt{2\pi}\sigma_i\right] \right\}$$

| | | |
|---|---|---|
| Subject to: | $\mathbf{f}(\mathbf{x}) = 0$ | plant model |
| | $\mathbf{x}^L \le \mathbf{x} \le \mathbf{x}^U$ | bounds on the process variables |

A NLP, and values are needed for $\eta$ and b

## Test for Gross Errors

If $\eta P(y_i | x_i, G) \ge (1-\eta) P(y_i | x_i, R)$, gross error

$\eta P(y_i | x_i, G)$: probability of a gross error

$(1-\eta) P(y_i | x_i, R)$: probability of a random error

$$|\epsilon_i| = \left| \frac{y_i - x_i}{\sigma_i} \right| > \sqrt{\frac{2b^2}{b^2-1} \ln\left[\frac{b(1-\eta)}{\eta}\right]}$$

# Robust Function Methods

$$\min_{\mathbf{x}} \quad -\sum_i [\, \rho(y_i, x_i) \,]$$

Subject to: $\mathbf{f}(\mathbf{x}) = 0$

$$\mathbf{x}^L \leq \mathbf{x} \leq \mathbf{x}^U$$

Lorentzian distribution

$$\rho(\epsilon_i) = \frac{1}{1 + \frac{1}{2}\epsilon_i^2}$$

Fair function

$$\rho(\epsilon_i, c) = c^2 \left[ \frac{|\epsilon_i|}{c} - \log\left( 1 + \frac{|\epsilon_i|}{c} \right) \right]$$

c is a tuning parameter

Test statistic

$$\epsilon_i = (y_i - x_i)/\sigma_i$$

# Parameter Estimation
# Error-in-Variables Method

## Least squares

$$\min_{\theta} \textit{Minimize}: \quad (\mathbf{y} - \mathbf{x})^T \Sigma^{-1} (\mathbf{y} - \mathbf{x}) = \mathbf{e}^T \Sigma^{-1} \mathbf{e}$$

$$\textit{Subject to}: \quad \mathbf{f}(\mathbf{x}, \theta) = 0$$

$\theta$ - plant parameters

## Simultaneous data reconciliation and parameter estimation

$$\min_{\mathbf{x}, \theta} \textit{Minimize}: \quad (\mathbf{y} - \mathbf{x})^T \Sigma^{-1} (\mathbf{y} - \mathbf{x}) = \mathbf{e}^T \Sigma^{-1} \mathbf{e}$$

$$\textit{Subject to}: \quad \mathbf{f}(\mathbf{x}, \theta) = 0$$

another nonlinear programming problem

# Three Similar Optimization Problems

*Optimize:* **Objective function**
*Subject to:* **Constraints are the plant model**

## Objective function

data reconciliation - distribution function
parameter estimation - least squares
economic optimization - profit function

## Constraint equations

material and energy balances
chemical reaction rate equations
thermodynamic equilibrium relations
capacities of process units
demand for product
availability of raw materials

# Theoretical Evaluation of Algorithms for Data Reconciliation

Determine sensitivity of distribution functions to gross errors

Objective function is the product or sum of distribution functions for individual measurement errors

$$P = \prod p(\epsilon) \propto \sum \ln p(\epsilon) \propto \sum \rho(\epsilon)$$

Three important concepts in the theoretical evaluation of the robustness and precisionof an estimator from a distribution function

## Influence Function

Robustness of an estimator is unbiasedness (insensitivity) to the presence of gross errors in measurements. The sensitivity of an estimator to the presence of gross errors can be measured by the influence function of the distribution function. For M-estimate, the influence function is defined as a function that is proportional to the derivative of a distribution function with respect to the measured variable, $(\partial\rho/\partial x)$

## Relative Efficiency

The precision of an estimator from a distribution is measured by the relative efficiency of the distribution. The estimator is precise if the variation (dispersion) of its distribution function is small

## Breakdown Point

The break-down point can be thought of as giving the limiting fraction of gross errors that can be in a sample of data and a valid estimation of the estimator is still obtained using this data. For repeated samples, the break-down point is the fraction of gross errors in the data that can be tolerated and the estimator gives a meaningful value.

# Influence Function

proportional to the derivative of the distribution function, IF $\propto \partial\rho/\partial x$

represents the sensitivity of reconciled data to the presence of gross errors

## Normal Distribution

$$IF_{MT} \propto \frac{\partial \rho_i}{\partial x_i} = \frac{y_i - x_i}{\sigma_i^2} = \frac{\varepsilon_i}{\sigma_i}$$

## Contaminated Gaussian Distribution

$$IF \propto \frac{\partial \rho_i}{\partial x_i} = \ = \frac{\frac{\varepsilon_i}{\sigma_i}\left\{(1-\eta)e^{\frac{-\varepsilon_i^2}{2}\left(1-\frac{1}{b^2}\right)} + \frac{\eta}{b^3}\right\}}{(1-\eta)e^{\frac{-\varepsilon_i^2}{2}\left(1-\frac{1}{b^2}\right)} + \frac{\eta}{b}}$$

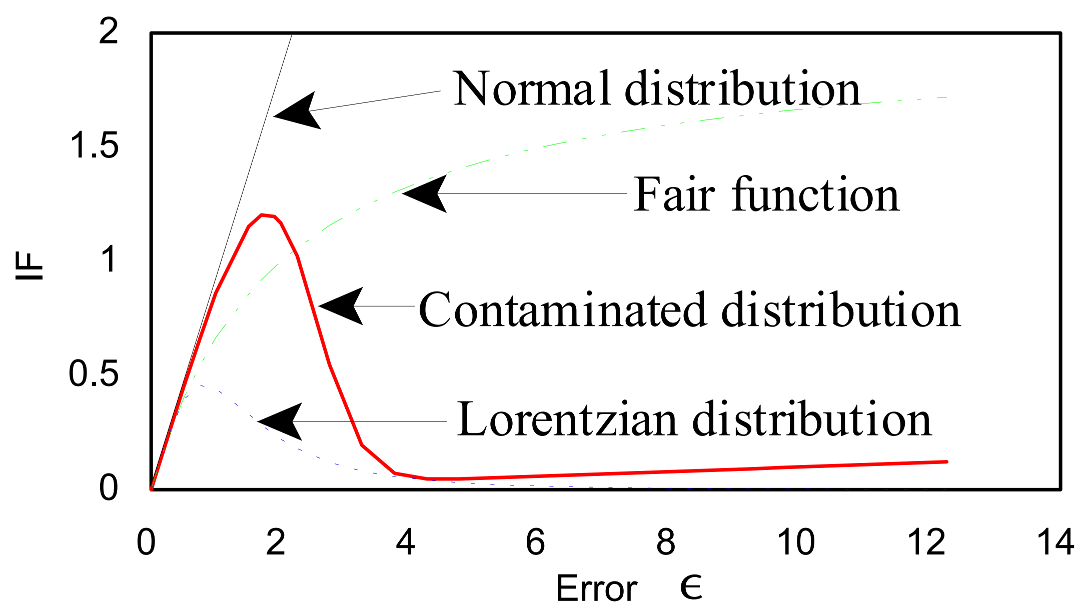
## Lorentzian Distribution

$$IF_{Lorentzian} \propto \frac{\partial \rho_i}{\partial \varepsilon_i} = -\frac{\varepsilon_i}{\left(1 + \frac{1}{2}\varepsilon_i^2\right)^2}$$

## Fair Function

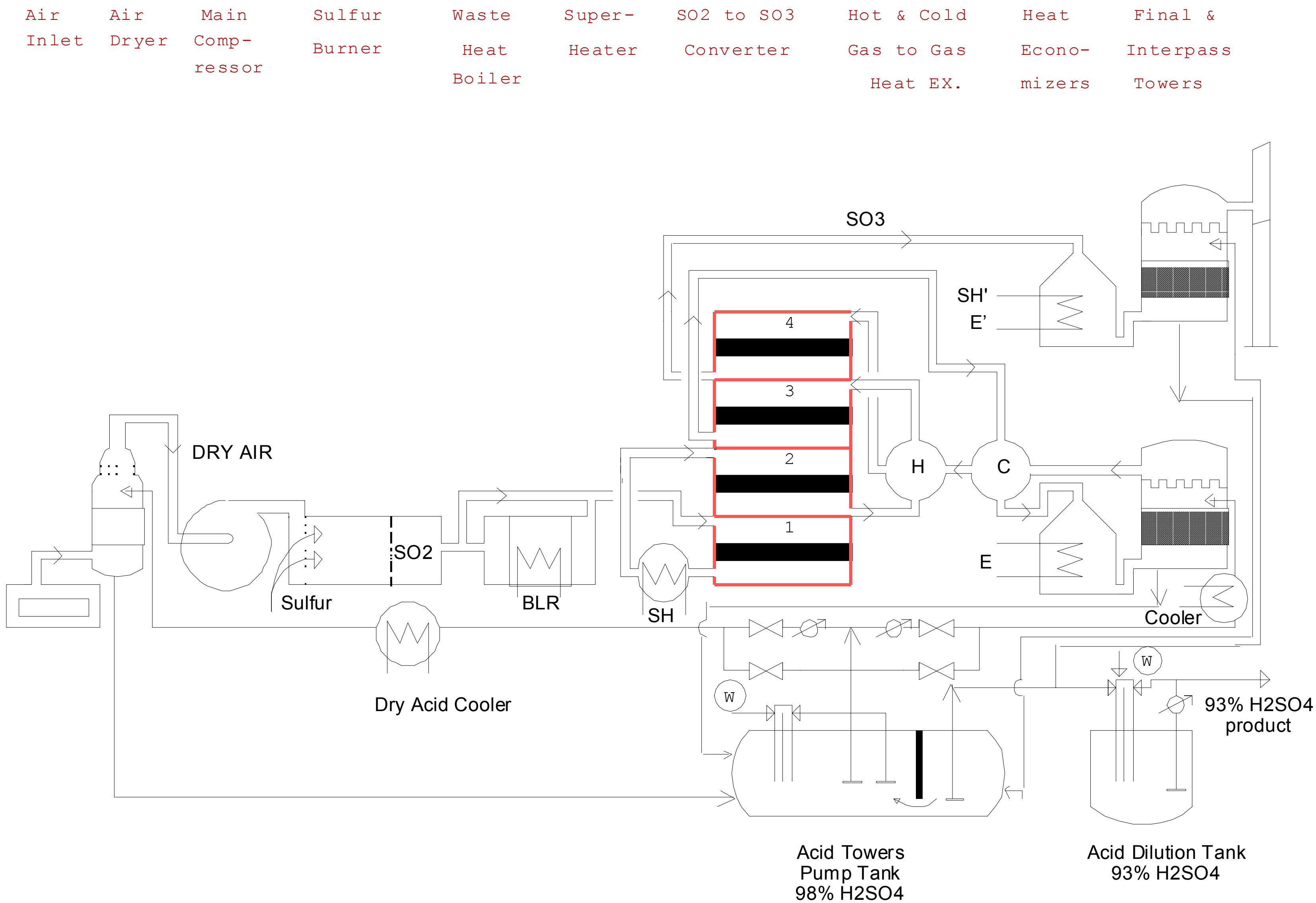
$$IF_{Fair} \propto \frac{\partial \rho_i}{\partial \varepsilon_i} = c^2\left(\frac{1}{c} - \frac{\frac{1}{c}}{1 + \frac{|\varepsilon_i|}{c}}\right) = \frac{1}{\frac{1}{|\varepsilon_i|} + \frac{1}{c}}$$

# Comparison of Influence Functions




Effect of Gross Errors on Reconciled Data - Least to Most

Lorentzian ▸ Contaminated Gaussian ▸ Fair ▸ Normal



Air Inlet
Air Dryer
Main Compressor
Sulfur Burner
Waste Heat Boiler
Super-Heater
SO2 to SO3 Converter
Hot & Cold Gas to Gas Heat EX.
Heat Economizers
Final & Interpass Towers
SO3
SH'
E'
4
3
2
1
H
C
E
DRY AIR
SO2
Sulfur
BLR
SH
Cooler
Dry Acid Cooler
W
W
93% H2SO4 product
Acid Towers Pump Tank 98% H2SO4
Acid Dilution Tank 93% H2SO4

## Numerical Evaluation of Algorithms

Simulated plant data is constructed by

$$\mathbf{y} = \mathbf{x} + \mathbf{e} + a\delta$$

**y** - simulated measurement vector for measured variables

**x** - true values (plant design data) for measured variables

**e** - random errors added to the true values

a - magnitude of a gross error added to one of measured variables

$\delta$ - a vector with one in one element corresponding to the measured variable with gross error and zero in other elements

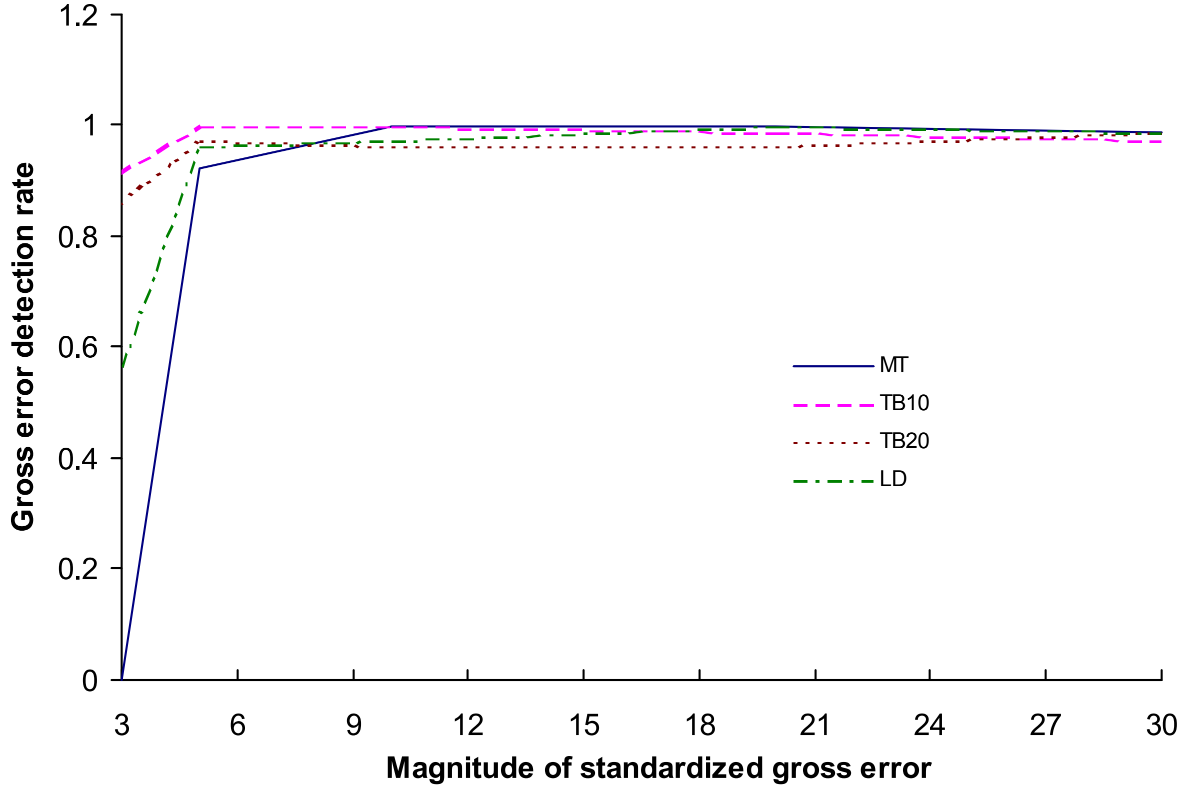
## Criteria for Numerical Evaluation

**Gross error detection rate** - ratio of number of gross errors that are correctly detected to the total number of gross errors in measurements

**Number of type I errors** - If a measurements does not contain a gross error and the test statistic identifies the measurement as having a gross error, it is called a type I error

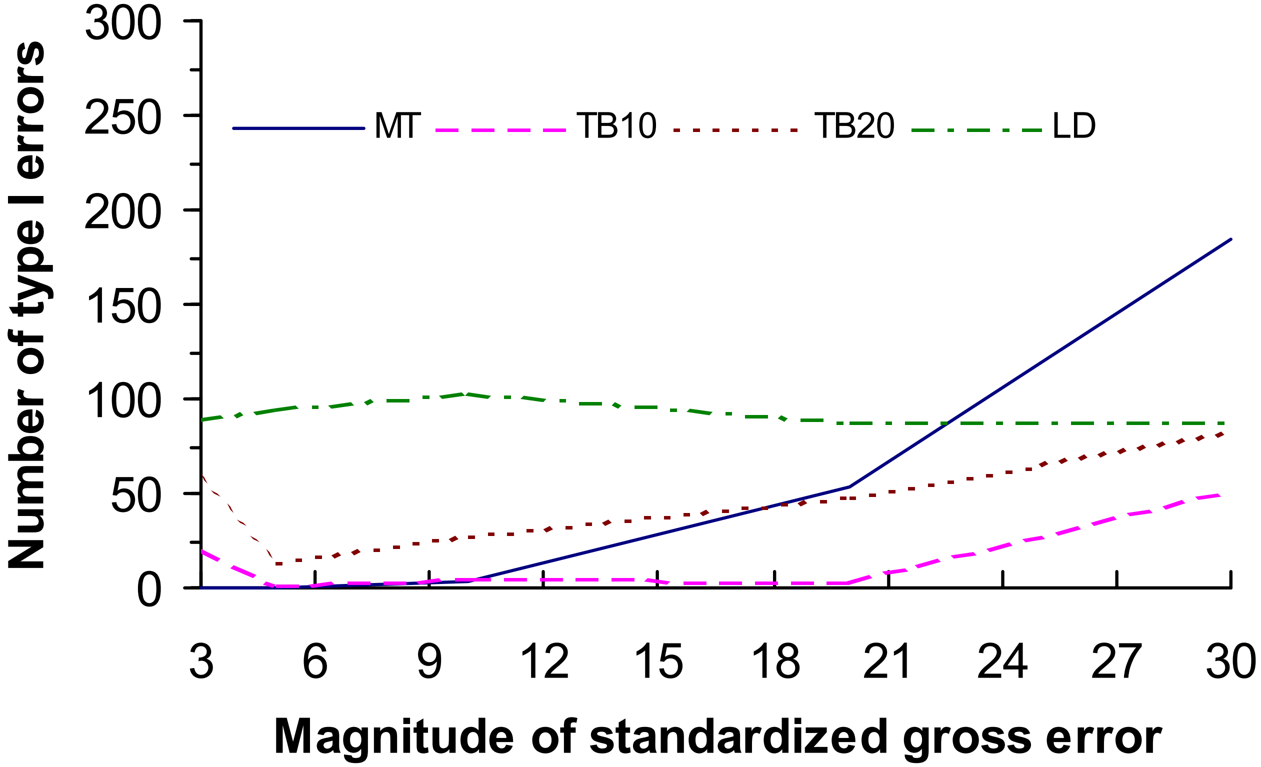
**Random and gross error reduction** - the ratio of the remaining error in the reconciled data to the error in the measurement

# Comparison of Gross Error Detection Rates
# 390 Runs for Each Algorithm

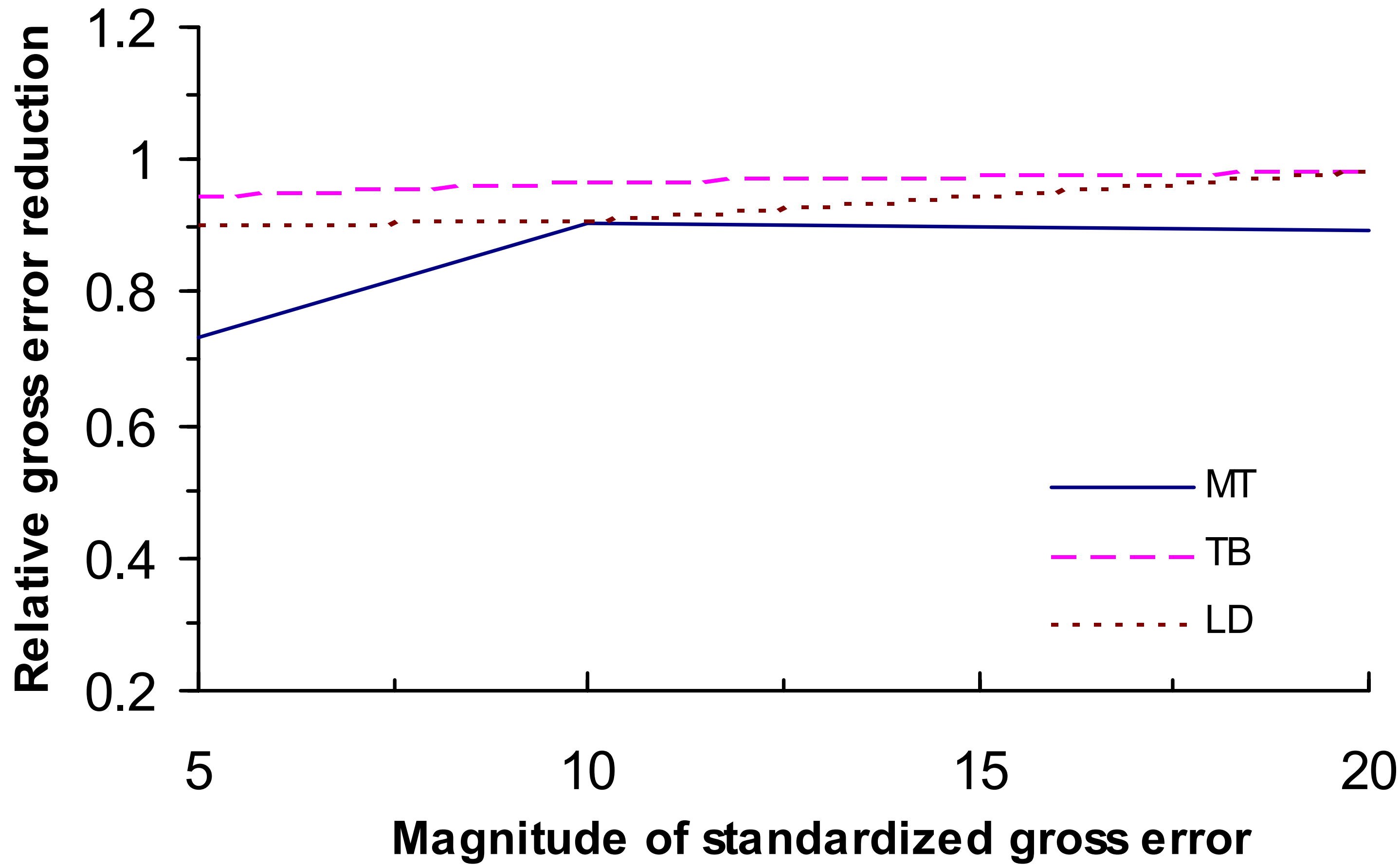


# Comparison of Numbers of Type I Errors

## 390 Runs for Each Algorithm



# Comparison of Relative Gross Error Reductions
# 645 Runs for Each Algorithm

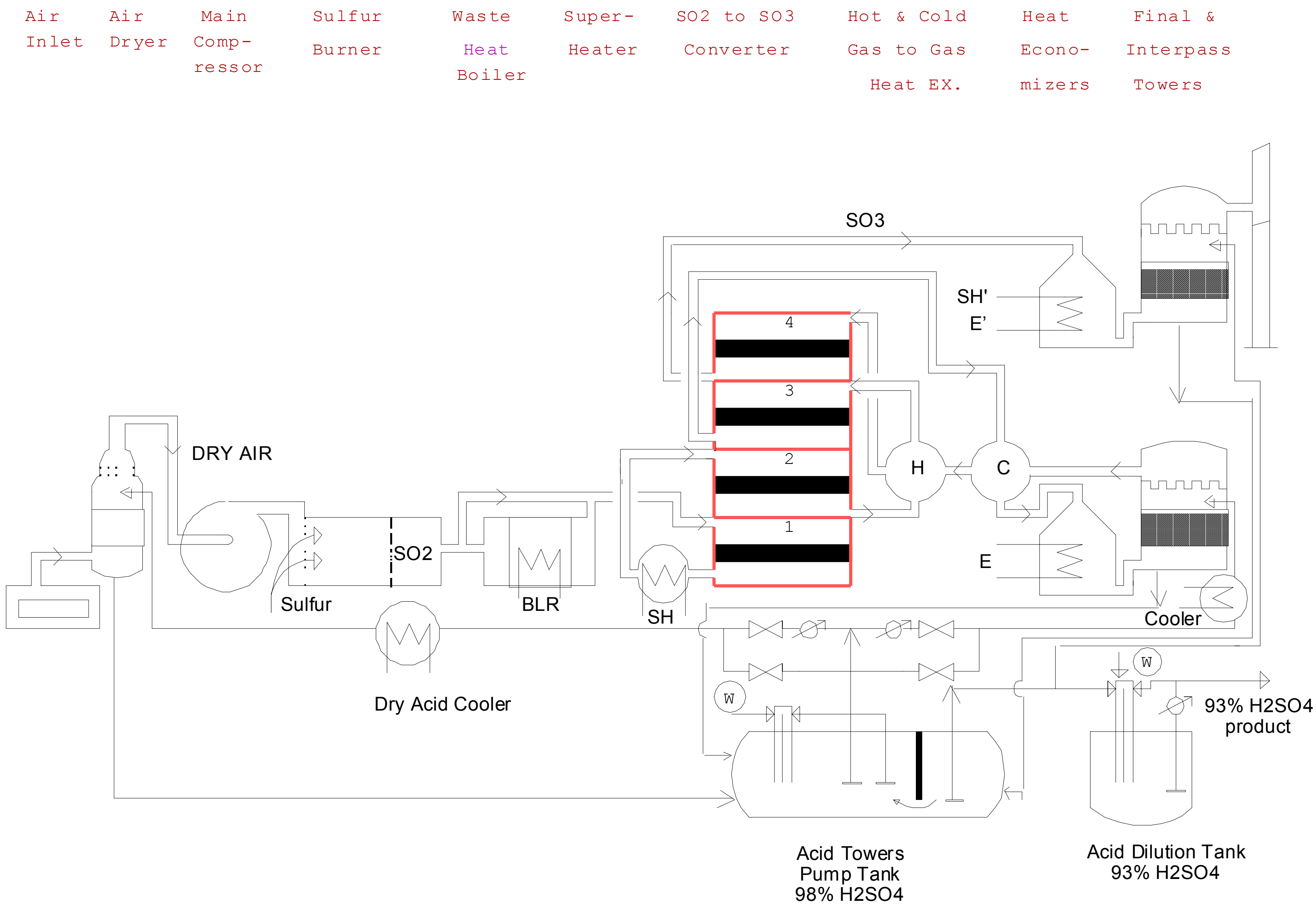


## Results of Theoretical and Numerical Evaluations

Tjoa-Biegler's method has the best performance for measurements containing random errors and moderate gross errors ($3\sigma$-$30\sigma$)

Robust method using Lorentzian distribution is more effective for measurements with very large gross errors (larger than $30\sigma$)

Measurement test method gives a more accurate estimation for measurements containing only random errors. It gives significantly biased estimation when measurements contain gross errors larger than $10\sigma$


Air Inlet
Air Dryer
Main Comp-ressor
Sulfur Burner
Waste Heat Boiler
Super-Heater
SO2 to SO3 Converter
Hot & Cold Gas to Gas Heat EX.
Heat Econo-mizers
Final & Interpass Towers
SO3
SH'
E'
4
3
2
1
DRY AIR
H
C
SO2
Sulfur
BLR
SH
E
Cooler
Dry Acid Cooler
W
W
93% H2SO4 product
Acid Towers Pump Tank 98% H2SO4
Acid Dilution Tank 93% H2SO4

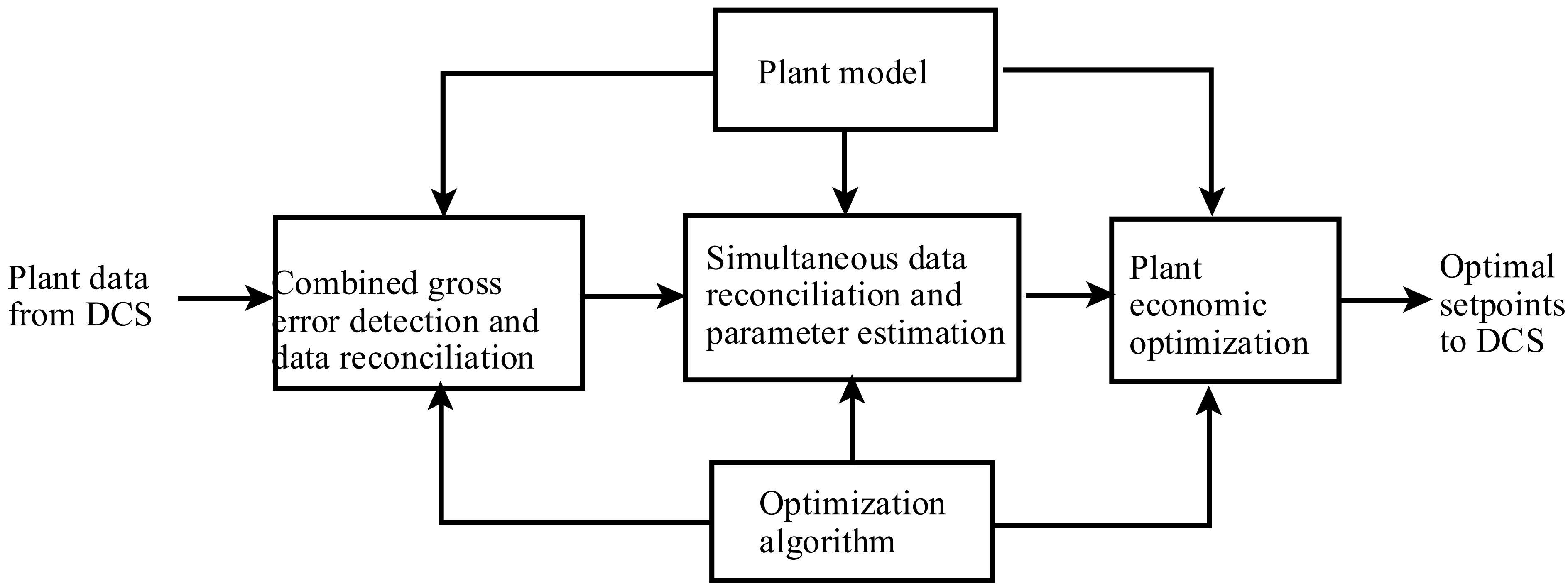
# Economic Optimization

## Value Added Profit Function

$$s_{F64}F_{64} + s_{FS8}F_{S8} + s_{FS14}F_{S14} - c_{F50}F_{50} - c_{FS1}F_{S1} - c_{F65}F_{65}$$

On-Line Optimization Results

| Date | Current ($/day) | Profit Optimal ($/day) | Improvement |
|---|---|---|---|
| 6-10-97 | 37,290 | 38,146 | 2.3% $313,000/yr |
| 6-12-97 | 36,988 | 38,111 | 3.1% $ 410,000/yr |


Plant model
Plant data from DCS
Combined gross error detection and data reconciliation
Simultaneous data reconciliation and parameter estimation
Plant economic optimization
Optimal setpoints to DCS
Optimization algorithm

# Interactive On-Line Optimization Program

1. Conduct combined gross error detection and data reconciliation to detect and rectify gross errors in plant data sampled from distributed control system using the Tjoa-Biegler's method (the contaminated Gaussian distribution) or robust method (Lorentzian distribution).

**This step generates a set of measurements containing only random errors for parameter estimation.**

2. Use this set of measurements for simultaneous parameter estimation and data reconciliation using the least squares method.

**This step provides the updated parameters in the plant model for economic optimization.**

3. Generate optimal set points for the distributed control system from the economic optimization using the updated plant and economic models.

# Interactive On-Line Optimization Program

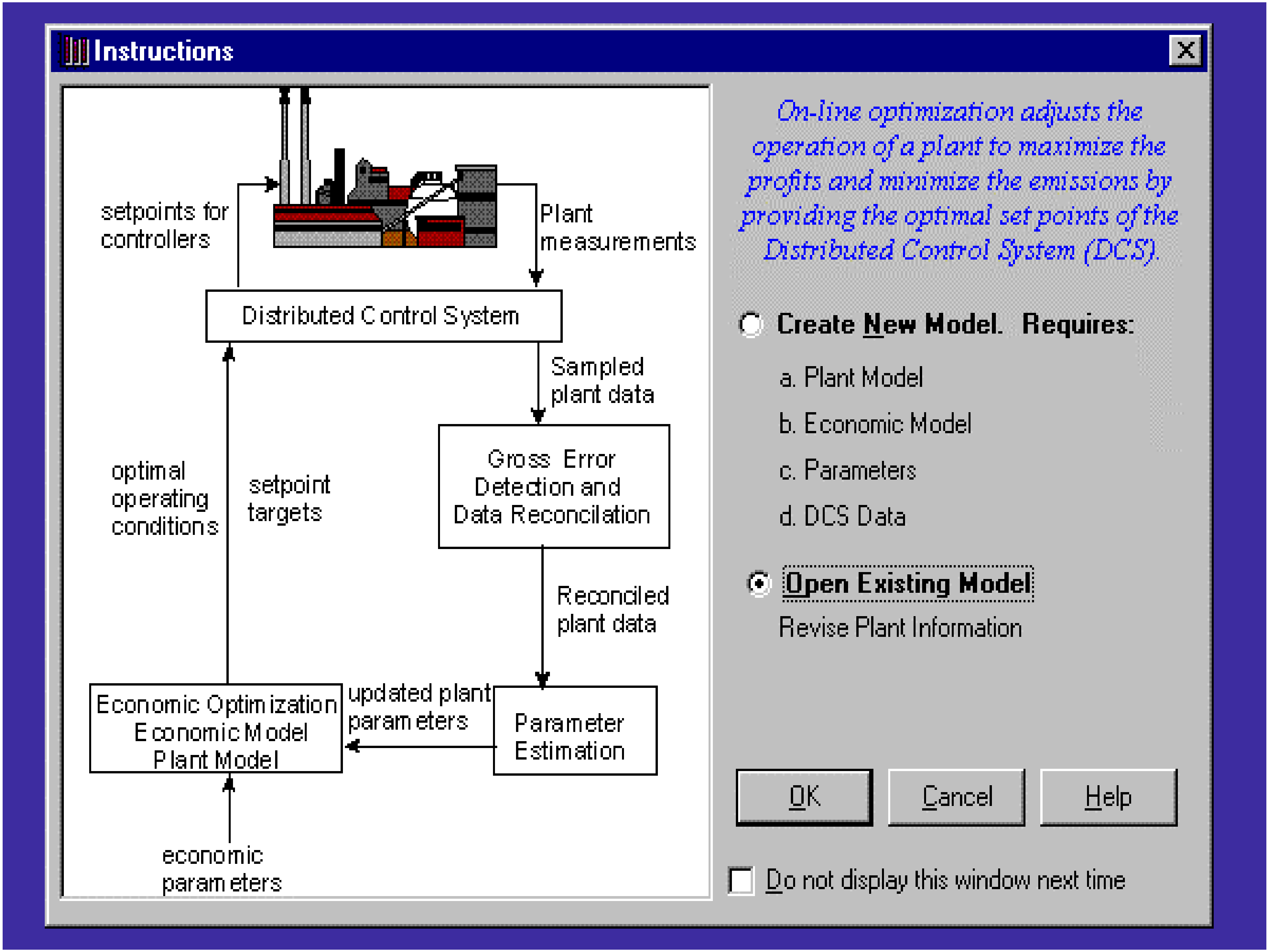
Process and economic models are entered as equations in a form similar to Fortran

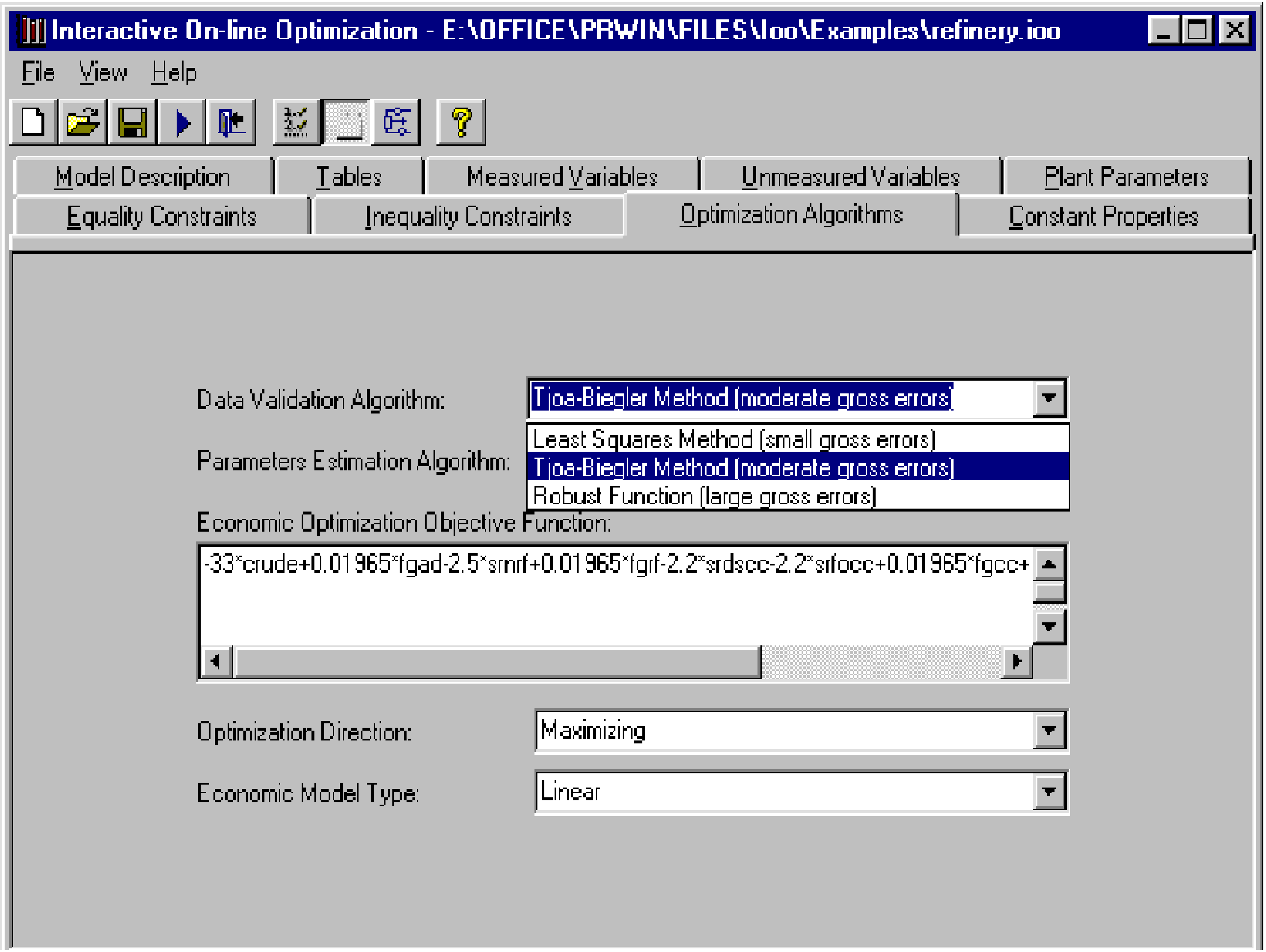
The program writes and runs three GAMS programs.

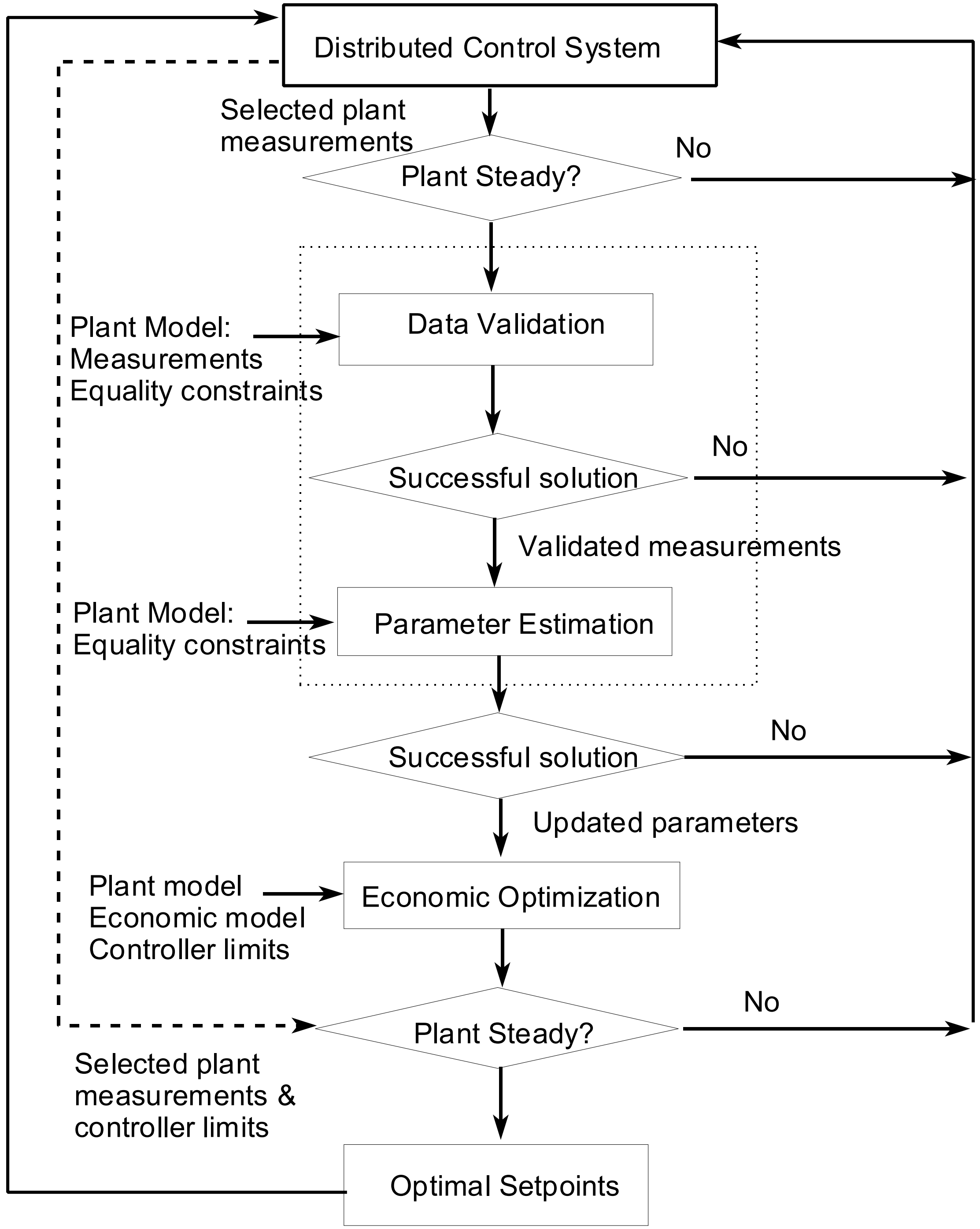
Results are presented in a summary form, on a process flowsheet and in the full GAMS output

The program and users manual (120 pages) can be downloaded from the LSU Minerals Processing Research Institute web site

URLhttp://www.mpri.lsu.edu


Instructions
setpoints for controllers
Plant measurements
Distributed Control System
Sampled plant data
Gross Error Detection and Data Reconcilation
optimal operating conditions
setpoint targets
Reconciled plant data
Economic Optimization Economic Model Plant Model
updated plant parameters
Parameter Estimation
economic parameters
On-line optimization adjusts the operation of a plant to maximize the profits and minimize the emissions by providing the optimal set points of the Distributed Control System (DCS).
Create New Model. Requires:
a. Plant Model
b. Economic Model
c. Parameters
d. DCS Data
Open Existing Model
Revise Plant Information
OK
Cancel
Help
Do not display this window next time


Interactive On-line Optimization - E:\OFFICE\PRWIN\FILES\Ioo\Examples\refinery.ioo
File View Help
Model Description
Tables
Measured Variables
Unmeasured Variables
Plant Parameters
Equality Constraints
Inequality Constraints
Optimization Algorithms
Constant Properties
Data Validation Algorithm:
Tjoa-Biegler Method (moderate gross errors)
Least Squares Method (small gross errors)
Tjoa-Biegler Method (moderate gross errors)
Robust Function (large gross errors)
Parameters Estimation Algorithm:
Economic Optimization Objective Function:
-33*crude+0.01965*fgad-2.5*smrf+0.01965*fgrf-2.2*srdscc-2.2*srfocc+0.01965*fgcc+
Optimization Direction:
Maximizing
Economic Model Type:
Linear



Distributed Control System
Selected plant measurements
Plant Steady?
No
Data Validation
Plant Model:
Measurements
Equality constraints
Successful solution
No
Validated measurements
Parameter Estimation
Plant Model:
Equality constraints
Successful solution
No
Updated parameters
Economic Optimization
Plant model
Economic model
Controller limits
Plant Steady?
No
Selected plant measurements & controller limits
Optimal Setpoints

# Some Other Considerations

Redundancy

Observeability

Variance estimation

Closing the loop

Dynamic data reconciliation
and parameter estimation

# Summary

Most difficult part of on-line optimization is developing and validating the process and economic models.

Most valuable information obtained from on-line optimization is a more thorough understanding of the process

② Acknowledgments

Support from

Gulf Coast Hazardous
Substance Research Center

Environmental Protection
Agency

Department of Energy