ChemPlant Technology s.r.o.

Report CPT- 450-22

Date of issue: 12.4. 2022

This report has 64 pages.

Process Data Reconciliation and Monte Carlo Method

Author: Vít Madron

Ústí nad Labem April 2022

HISTORY OF REVISIONS

Revision No	Date	Description	Elaborated by	Approved by

CONTENTS

SUMMARY

1. The main purpose of this report is to verify the rightfulness of application of Data Reconciliation and Validation methods (which were developed on the assumption of models' linearity) to real industrial models which are mostly nonlinear.
2. Three indicators of the influence of model's nonlinearity of Data Reconciliation results were used:
   - Theoretical mean value of the least squares function (which equals to the Degree of Redundancy) was compared with the average of Q_{min} values calculated in MCM simulations
   - Theoretical value of the second central moment (**variance)** of the least squares function (which equals 2 times of the Degree of Redundancy) was compared with the average of Q_{min} variance in MCM simulations
   - Theoretical value of the probability of the Error of I^{st} Kind in testing the presence of a gross errors (which equals 5 %) was compared with the relative number of false Gross Error Detection in MCM simulations (in per cents).
3. The main purpose of Chapter 5 was to verify that MCM methods used in RECON (generation of random variables, etc.) are sound. Calculations revealed that it is needed to make 10,000 MCM repetitions to get reliable results. The MCM analysis of a simple linear model has confirmed that the DRV methodology works and the results' precision agrees with MCM results (Table 5-2). Also the Gross Errors Detectability method gives good results (Table 5-3).
4. The core of the report is in Chapter 6. The spectrum of 12 nonlinear models covers typical DRV tasks we can meet in Chemical and Power Industries. Models' characteristics are shown in Table 6-13. The typical type of nonlinearity is a product of two variables (bilinear models, namely multicomponent and heat balances). The nonlinearity in all cases did not caused significant deviations caused by models' linearization during the DRV solution (Table 6-14).
5. In Section 2.5 was proposed the practical and simple measure of models' nonlinearity by Eq. (2-25). It is the relative improvement of the Least Squares function Q_{difrel} calculated by the Successive Linearization and then improved by the SQP method: $Q_{difrel} = (Qmin_{SL} – Qmin_{SQP}) / Qmin_{SQP}$.

   In Chapter 6 was shown that for most of industrial models the SQP method is not mandatory but in some cases it is required. This decision must be done experimentally, for example by the MCM method.

6. In Chapter 7 were analyzed two bilinear models as concerns the influence of measurement uncertainties on statistical results of DRV. It was concluded that there is no evidence of significant influence of measurement uncertainties on basic statistical characteristics of the data reconciliation process.
7. In Chapter 8 was on 4 examples shown that MCM is a good method for testing models' robustness. Random errors' of measurements were be perturbed up to 5 times of the original measurement uncertainties to test models' robustness.

GLOSSARY AND ABBREVIATIONS

MCM	Monte Carlo Method
Recon	Mass, energy and momentum balancing software with Data Validation and Reconciliation
DoR	Degree of Redundancy
DR	Data Reconciliation
DVR	Data Validation and Reconciliation
GED	Gross Error Detection
NLP	Nonlinear Programming
PDF	Probability Density Function
PF	Perturbation Factor
Qaver	average value of Q_{min}
Q_{min}	the Least Squares sum
Q_{crit}	the critical value of the Least Squares sum
Q_{difrel}	relative difference of Qmin between SL and SQL methods (see Eq.(2-22)
S	Status of Data Quality
S_{aver}	mean (expected) value of the Status
SL	Successive Linearization
SQP	Successive Quadratic Programming
TV	Threshold Value
VQaver	average value of Q_{min} variance (the second central moment)
% GED	% of cases with detected Gross Error
μ	Greek letter *mu* – mean value
ν	Greek letter *nu* – Synonym for Degree of Redundancy (DoR)
σ_i	standard deviation of measurement error
$\sigma_{x'i}$	standard deviation of reconciled value
σ_{vi}	standard deviation of adjustment

1 INTRODUCTION

Process Data Reconciliation (DR) accompanied by related techniques (Gross Errors Detection and Elimination, measurement points placement, etc.) is in practice based on statistical inference methods [1] using the family of probability distributions like Normal (Gauss) or Chi-square. For linear functions the Normal distribution of inputs remains Normal also for functions outputs. On the other hand side, most of models important in practice are not linear and their solution is based on their linearization. In such cases the reconciled values have Normal distribution no more, the same holds for Least Squares functions which are not distributed exactly Chi-square, etc. It is therefore legitimate to state a question: Is it justified to apply DR techniques to industrial models which are mostly nonlinear?

The purpose of this report is to clear the importance of neglecting nonlinearity of models used in practice. Thus formulated problem is not easy to solve analytically and the Monte Carlo simulation [2] can be a good way to tackle this problem.

Monte Carlo Method (MCM) is a mathematical technique used to estimate possible result of an uncertain event. The Monte Carlo Method simulation predicts outcomes based on a set of random input values. It recalculates the results over and over, each time using a different set of random numbers generated according to some probability distribution. The well known **Normal (Gauss) distribution of measurement errors** which is preferred in technical modeling (mass and energy balancing, thermodynamic calculations, etc.) will be used throughout this report.

Application of the Monte Carlo Method (MCM) in studying DR is not new. Probably for the first time MCM was applied by Iordache et all [12] in studying the problem of the Gross Errors Detection test power. Ozyurt and Pike [13] studied by MCM extensively the efficiency of gross errors detection. Bagajewicz and Nguyen proposed to calculate the expected value of accuracy [14,15] by MCM. Syed et all used MCM to verify results from the linearized models of a gas turbine system, especially functioning of Gross errors detection [16]. Cencic and Fruhwirth [17] used the Markov chain Monte Carlo method for modeling non-normal distributions which are results of models nonlinearity and can't be solved analytically. Wingstedt and Saarela used MCM to evaluate error propagation in computation of nuclear plant thermal power [18].

In this report four areas will be studied:

- verification of internal statistical methods used in Recon for MCM
- influence of model nonlinearity on Data Reconciliation results
- influence of measurement uncertainties on reconciled results' precision (confidence intervals)
- robustness of Recon's functionality in the presence of Gross Errors in input data.

The subjects of the MCM will be random errors added to "errorless" values of measured variables (flowrates, temperatures, etc.). With this aid we will simulate the influence of measurement errors on final results.

It is clear that such solution can't give a general answer for all possible models occurring in industrial practice. We have selected 12 model typical for Power generation, Oil refining, Petrochemical and Natural gas distribution industries. In what follows 7 simple models will be used to study MCM applied to typical process industries tasks (mainly unit operations in these areas). At the end 5 industrial size tasks will be analyzed (a coal fired steam generator with auxiliaries, a powerplant supercritical steam cycle, heavy crude vacuum distillation system, a steam generation system in a nuclear powerplant and Natural gas transport and distribution system with hydralics modeling).

The modeling tool used in this report is the mass and energy balancing software with Data Reconciliation and Validation RECON® by ChemPlant Technology, s.r.o., see RECON | ChemPlant Technology - process data information systems, mass and energy balancing software. RECON in its Lite version can be here freely downloaded. The Lite version makes possible calculation of most of examples presented further in this report.

2 MODELING INDUSTRIAL PROCESS SYSTEMS BY RECON

The next Chapter 2 summarizes briefly theory of DR including some more advanced methods like measurement errors propagation and the Power of testing hypotheses about gross errors. There are many good books devoted fully or partially to these subjects [3-10]. There is also practically oriented report [11] available free from Internet at Papers and reports | ChemPlant Technology - process data information systems, mass and energy balancing software. The notation in this report is taken over from the book [4].

2.1 Models

It is universally accepted that any measurement is charged with some error. The measurement error is defined by the following equation.

$$x^+ = x + e \tag{2-1}$$

where x^+ is the measured value

x is the true (unknown) value

e is the measurement error

Most frequently is supposed that e is a random variable with the Normal distribution with zero mean value characterized by the standard deviation σ. In practice the standard deviation is supposed to be related with the measurement tolerance or the maximum measurement error. The measurement uncertainty (maximum measurement error is the term used in Recon) is taken as 1.96 multiple of σ (this stems from the Normal distribution and the probability level 95 %).

Note: The nomenclature here is not unified. The notion measurement *uncertainty* has also the synonym measurement *tolerance.* In Recon used *maximum measurement error* has the same meaning.

Let us start from the mathematical model

$$\boldsymbol{F(x,y,c) = 0} \tag{2-2}$$

where ***F( )*** is the vector of implicit model equations (generally nonlinear)

x is the vector of directly measured variables

y is the vector of directly unmeasured variables

c is the vector of precisely known constants

The starting point for the following solution is the solvability analysis of a set of linear equations in variables representing measured and unmeasured variables. The important simplification of the nonlinear model (2-2) is so-called ***General Linear model***

$$\boldsymbol{A'x} + \boldsymbol{By} + \boldsymbol{a} = \boldsymbol{0} \quad (2\text{-}3)$$

where

x is vector of measured variables

y vector of unmeasured variables

a vector of constants

A' and ***B*** are matrices of constants

The General Linear model can be further simplified by elimination [4] of unmeasured variables to the form containing only measured variables (note that matrices ***A*** and ***A'*** are different):

$$\boldsymbol{Ax} + \boldsymbol{a} = \boldsymbol{0} \quad (2\text{-}4)$$

2.2 Data reconciliation

Eq. (2-2) holds for the true (unknown) values of the variables. If we replace them by the measured values $\boldsymbol{x}^{+}$, the equations need not (and most likely will not) be exactly satisfied:

$$\boldsymbol{F}(\boldsymbol{x}^{+},\boldsymbol{y},\boldsymbol{c}) \neq \boldsymbol{0} \quad (2\text{-}5)$$

whatever be the values of the unmeasured variables (unless the degree of redundancy equals zero).

The basic idea of DR is the adjustment of the measured values in the manner that the reconciled values are as close as possible to the true (unknown) ones. The reconciled values x_i' (marked by apostrophe) result from the relation

$$x_i' = x_i^+ + v_i \quad , \qquad (2\text{-}6)$$

where to the measured values, so-called *adjustments* v_i are added. In the ideal case, these adjustments should be equal to the minus errors, but these are unknown. If, however, we have the mathematical model that must be obeyed by the correct values then the optimal solution is as follows:

The adjustments must satisfy two fundamental conditions:

1) The reconciled values obey Eq. (2-2) – we say that they are consistent with the model

$$\boldsymbol{F(x',y',c) = 0} \qquad (2\text{-}7)$$

2) The adjustments are minimal. Most frequently, one minimizes the weighted sum of squares of the adjustments using the well-known *least squares* method

$$\text{minimize} \quad \sum (v_i/\sigma_i)^2 = \sum [(x_i' - x_i^+)/\sigma_i]^2 . \qquad (2\text{-}8)$$

where $v_i = x_i' - x_i^+$ are so called adjustments.

The inverse values of the standard deviations σ_i^2 – so-called *weights* – then guarantee that more (statistically) precise values are less corrected than the less precise ones (this is a relevant property of the method).

The least squares function (2-6) is used in the case of uncorrelated (statistically independent) errors. In the case of correlated errors a more general criterion is minimized:

$$\text{minimize} \quad \boldsymbol{v^T F^{-1} v} \qquad (2\text{-}9)$$

where $\boldsymbol{v}$ is vector of adjustments and $\boldsymbol{F}$ is the covariance matrix of measurement errors.

The reconciliation proper is an optimization problem requiring computer technique and effective software. In contrast to many other engineering calculations, the DR cannot be carried out manually (using a pocket calculator) even with very simple models.

The mathematics of the solution itself was in the last decades many times described in the literature (e.g. [3-11]) and will not be mentioned in the sequel.

So let us further suppose that at our disposal is the program RECON ready to use for DR. Schematically, it is the Data Reconciliation Engine depicted in the following figure.

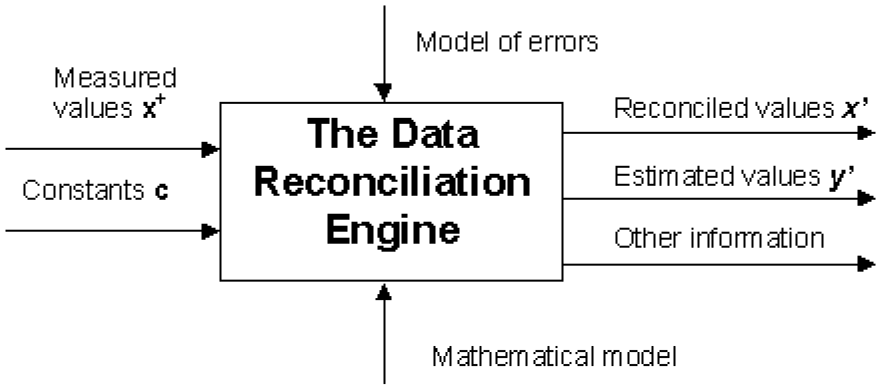


Fig. 2-1: The Data Reconciliation Engine

The model (2-4) is used for DR proper. In the first step the adjustments $\boldsymbol{v}$ are calculated according to the equation

$$\boldsymbol{v = - F_X A^T (A F_x A^T)^{-1} (a + A x^+)} \qquad (2\text{-}10)$$

Reconciled values $\boldsymbol{x'}$ are then calculated from the equation

$$\boldsymbol{x' = x^+ + v} \qquad (2\text{-}11)$$

by substitution from Eq. (2-10).

2.3 Statistical properties of results

Adjustments $\boldsymbol{v}$ have the normal distribution $N(\boldsymbol{0,F_v})$ and the covariance matrix of adjustments $\boldsymbol{F_v}$

$\boldsymbol{F_v} = \boldsymbol{F_x A^T (A F_x A^T)^{-1} A F_x}$ (2-12)

The Quadratic form of adjustments (2-8) or (2-9) is the random variable with $\chi^2_{(1-\alpha)}(\nu)$ distribution with ν degrees of freedom. Values of $\chi^2_{(1-\alpha)}(\nu)$ for probability $(1-\alpha)$ are tabulated in statistical tables.

Between covariance matrices of measurement errors $\boldsymbol{F}$, *adjustments* $\boldsymbol{F_v}$ *and reconciled values* $\boldsymbol{F_{x'}}$ holds the important relation

$\boldsymbol{F} = \boldsymbol{F_v} + \boldsymbol{F_{x'}}$ (2-13)

For variances of measurement errors, adjustments and reconciled values therefore hold

$\sigma_i^2 = \sigma_{vi}^2 + \sigma_{x'i}^2$ (2-14)

Square roots of variances (standard deviations) of reconciled values are important for estimating confidence intervals for results. On assumption of normal distribution of measurement errors it holds that with the probability 95 % the intervals

$< x_i' - 1.96\ \sigma_{x'i}\ ;\ x_i' + 1.96\ \sigma_{x'i} >$ (2-15)

cover the (unknown) true values of individual variables.

Reconciled data are more precise in the statistical sense, if compared with the measured ones. The enhanced precision can be quantified with the aid of the standard deviation of the reconciled value, which is always smaller than the standard deviation of the measurement error.

$\sigma_{x'} < \sigma$ (2-16)

The measure of the precision improvement is so-called *adjustability* defined as

$a = 1 - \sigma_{x'} / \sigma$ (2-17)

The adjustability characterizes the reduction of the standard deviation and thus also the uncertainty of the result, if compared with the primary measurement. If for example the adjustability of the reconciled value is 0.5, the uncertainty has been reduced by half. Adjustability 0.75 means reducing the uncertainty by a quarter, and so on. The greater the adjustability is, the greater is also the reduction of the uncertainty.

2.4 Detection of gross errors

The most frequently used method for Gross Errors Detection (GED) is the test based on the value the least squares function (2-8) or (2-9). The Quadratic form of adjustments (2-8) or (2-9) is the random variable with $\chi^2_{(1-\alpha)}(\nu)$ distribution with ν degrees of freedom. Values of $\chi^2_{(1-\alpha)}(\nu)$ for probability (1-α) are tabulated in statistical tables.

If the value of the minimal value of the least squares function is denoted as Q_{min},

$$Q_{min} = \boldsymbol{v}^T \boldsymbol{F}^{-1} \boldsymbol{v}, \qquad (2\text{-}18)$$

with probability (1-α) the value of Q_{min} will be less than the **critical value of the χ^2 distribution** with ν degrees of freedom.

$$Q_{min} < \chi^2_{(1-\alpha)}(\nu) = Q_{minCrit} \qquad (2\text{-}19)$$

Number of degrees of freedom ν is in DR solutions called **Degree of Redundancy (DoR).** In most cases it holds that

DoR = *Number of model equations – Number of unmeasured variables*

Probability level (1-α) is usually supposed in technical sciences to be 0.95 (95 %) and this value will be used also throughout this report). All this holds on assumptions that only random errors with the Normal distribution are present.

Recon uses for GED slightly modified approach. The *Status of Data quality S* is defined as

$$S = Q_{min} / Q_{minCrit} \qquad (2\text{-}20)$$

Then the Eq. (2-19) reads

$S < 1$ (2-21)

If *S* less than one, no gross error is detected.

The *S* definition has the advantage for an end DR user who does not need to know critical values for *Qmin* at different degrees of freedom. In words, a gross error is detected when the Status of Data Quality is equal or greater than 1. Mean (expected) values of *S* are presented in Table 3.1 in the next Chapter.

It may be useful to note that the probability α is the **expected probability of the *Error of I^st^ kind* (a Gross Error is detected even if it is not present).** In this report is supposed that α is 0.05. This means that we can expect 5 % of cases a gross error is detected even if it is not present.

Gross errors *detectability*

Gross errors *detectability* means that a gross error of some size will be detected with some probability. This problem is solved by so called *threshold values* which are characteristic for every measured redundant variable.

Let's recall the Eq. (2-1) defining a random error and let's modify it to the form

$$x^+ = x + e + d \quad , \qquad (2\text{-}22)$$

where *d* is a gross error (which is a constant).
One has to begin with testing the gross error presence hypothesis.

As any statistical test, also the χ^2 test has its *power characteristic* :

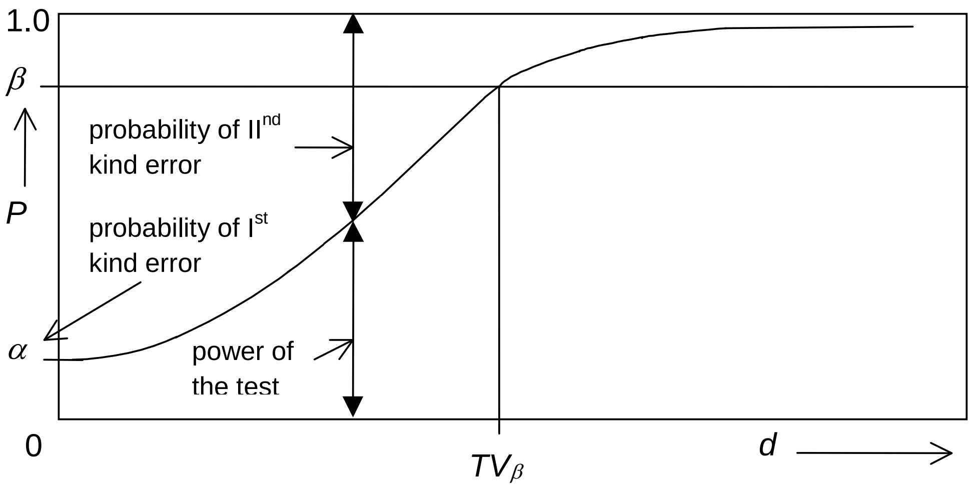


Fig. 2-2: The *power characteristic* of the χ^2 test

On the x - axis, we have the magnitude of the gross error d, on the y - axis the probability P of the gross error detection. The value given by the power characteristic for an adjustable measured variable equals the significance level α of the test assuming the absence of gross error (d=0), and it approaches 1 for high values of the gross error ($d \to \infty$).

The power characteristic represents though complete, still too complicated information for the application in practice (imagine hundreds of such lines in a real size problem). More simple is the characteristic of measured variables by means of a single number, so-called *threshold value* (*TV*) for the gross error detection.

TV_β is the value of gross error that will be detected with probability β (we'll further assume $\beta = 0.9$). TV_β is a characteristic value for any measured adjustable variable. The smaller TV_β , the better. TV_β is called the *threshold value*.

The threshold value is computed from the equation

$$q_i = \delta_\beta(\nu,\alpha)/[a_i(2-a_i)]^{1/2} \qquad (2\text{-}23)$$

where q_i is dimensionless threshold value TV/σ

$$q_i = TV_i/\sigma_i \qquad (2\text{-}24)$$

and $\delta_\beta(\nu,\alpha)$ is a constant, characteristic for the significance level α of the *chi-square* test, degree of redundancy ν and probability of the gross error detection β. For more details, see the literature [4], p. 177.

Values of $\delta_\beta(\nu,\alpha)$ for $\alpha = 0.05$, $\nu = 1,2,\ldots,500$ and $\beta = 0.90, 0.95$ and 0.99 are presented in [19]

Let us notice that for a measured variable, the threshold value is a simple function of its *adjustability* defined by Eq. (3.5-2); see also the following figure.

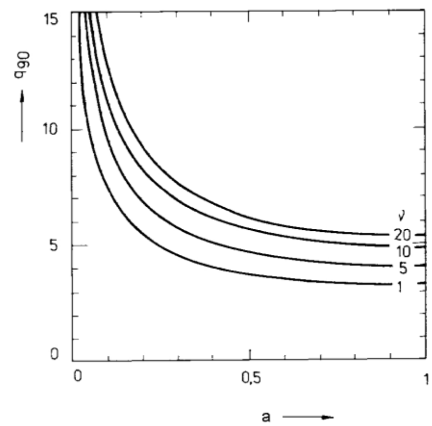


Fig. 2-3: Dimensionless threshold value *q* as function of the degree of redundancy ν and adjustability *a* (for α=0.05 and β=0.9)

From this diagram, one can derive certain simple conclusions:

- The greater the adjustability is, the greater is also the probability that the gross error will be detected (low value of threshold error)
- For adjustability smaller than 0.01, the probability of gross error detection is very small and decreases further rapidly
- The minimum threshold value equals 3.24 times the standard deviation of the measurement (this in the case of ν = 1 and adjustability = 1 , where *q* equals the minimum value 3.24). Considering that the maximum measurement error is taken as 1.96 times the standard deviation, the minimum threshold value results as 1.65 times the maximum measurement error. From this finding follows that the method for gross error detection is not omnipotent even under optimal conditions and is effective only for gross errors significantly greater than supposed measurement uncertainty.

2.5 DR solution by Recon

There exist two basic methods of Data Reconciliation applied to nonlinear models:

- Successive Linearization (SL)
- Nonlinear programming (NLP).

SL method is based on linearization of the model by the first order Taylor expansion. After the least squares solution on such linearized model is solved, this process is repeated from this new point. The iterative process is ended after equation values are zeroed (equations residuals are below some specified values). In Recon is watched also condition of minimal increments of measured and nonmeasured variables in individual iterations.

The SL method is very fast like other Newton-like methods but the problem is that for nonlinear models it does not find the exact Least Squares minimum. The distance from the LS minimum depends on the model nonlinearity and on the distance of measured values from the LS minimum. Recon therefore combines the SL method which finds the first solution and then applies the NLP (the ChemPlant's proprietary version of the Sequential Quadratic Programming - SQP).

In practice, for most models the SQP step is not needed as the SL method finds the solution which is very close to the global minimum. The SQP step is the option which can be used in DR with difficult models. See also the discussion about Example 2.2. below.

Let's illustrate DR on two very simple examples:

Example 2.1: Linear model

The model is very simple. There are two measured variables X1 and X2 and the model is (let's imagine two flowmeters on one pipe).

X1 – X2 = 0

Let the measured values are

X1 = 1

X2 = 0.5

Both flowmeters have the same uncertainty. The solution is shown in the next figure:

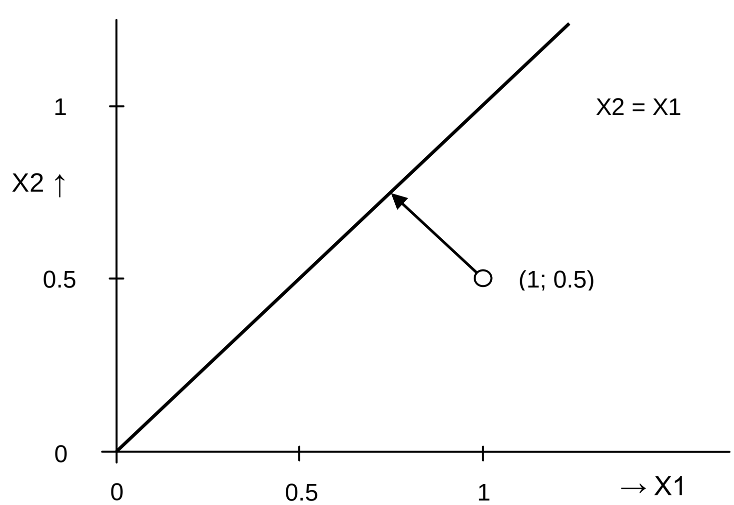


Fig. 2.4: DR - Linear model

The solution **is found in one step**:

X1 = 0.75

X2 = 0.75

The solution (arrow) is perpendicular to model line X1 – X2.

Example 2.2: Nonlinear model

This model is Parabola

X1*X1 – X2 = 0 or $X2 = X1^2$

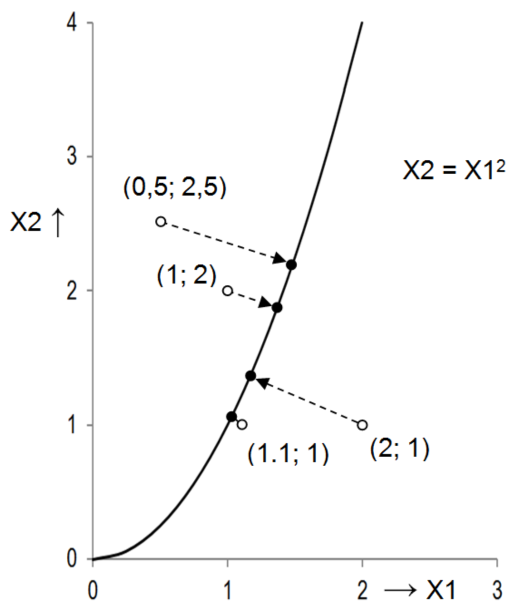
The uncertainties for both variables are 0.1.

There are 4 sets of measured values which are in the different distance from the model curve (Medium distance 1, Medium distance 2, Far distance, Near distance). See the next Table:

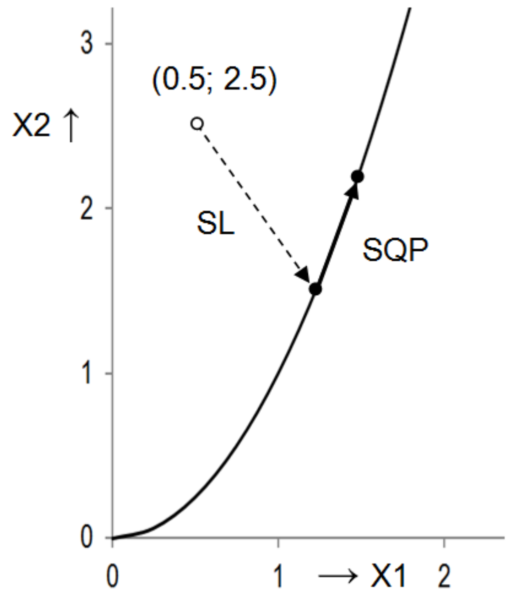
Tab. 2.1: Data reconciliation of the parabola model

Data set	X1 Meas.	X2 Meas.	Qmin SL	Qmin SQP	X1 SL	X1 SQP	X2 SL	X2 SQP
Near distance	1	1.1	0.745	0.744	1.039	1.040	1.080	1.081
Medium distance	1	2	5.932	5.836	1.348	1.365	1.818	1.864
Far distance 1	2	1	3.233	3.170	1.116	1.171	1.246	1.371
Far distance 2	0.5	2.5	585	406	1.225	1.473	1.502	2.169

See also the next Figure:



a)



b)

Fig. 2.5: DR - Nonlinear model (parabola). a) final SL+SQP solution for 4 data sets, b) Detail of SL and SQP solution for the Far distance 2 data.

In Fig. 2.3 a) the arrows show the final DR solution by the SQP method. Fig. 2.3. b) shows the solution details – the SL method with the following SQP for the Far distance 2 data (0.5; 2.5). The first SL solution finds the solution which lies on the model curve. Then is applied the SQP method which finds the solution with the minimum Least Squares sum. In the next Recon report is the course of iterations.

```
Task: _PARABOLA05-25 (parabola)

I T E R A T I O N S

Iter          Qeq            Qx              Qy              Qmin
-----------------------------------------------------------------
START     2.2500E+00
   1      1.2656E+00     7.9510E-01      0.0000E+00      9.7240E+02
   2      1.2656E-01     1.8601E-01      0.0000E+00      6.2352E+02
   3      1.8631E-03     2.3181E-02      0.0000E+00      5.8557E+02
   4      4.3235E-07     3.5158E-04      0.0000E+00      5.8501E+02
   5      6.0338E-02     3.2496E-01      0.0000E+00      4.2258E+02  SQP
   6      3.3825E-04     9.7037E-03      0.0000E+00      4.0711E+02
   7      7.0272E-04     2.9327E-02      0.0000E+00      4.0578E+02  SQP
   8      4.8957E-08     1.1296E-04      0.0000E+00      4.0561E+02
   9      1.7391E-06     2.0310E-03      0.0000E+00      4.0560E+02  SQP
  10      8.1015E-12     2.7937E-07      0.0000E+00      4.0560E+02
```

Legend:

Qeq mean residual of equations
Qx mean increment of measured variables in iteration
Qy mean increment of non-measured variables in iteration
Qmin least-square function

Let's note that the measured data X1 and X2 are far from the model (parabola), the Qmin = 405.6 which is much higher that than the critical value for one degree of freedom (3.84). The limits on Qeq and Qx were set at 0.001. The SL process required 4 iteration to reach the solution with Qmin = 585.01. The first SQP run required 2 iterations to reach Qmin = 407.11. The improvement by 2 following SQP runs is only symbolic (405.60).

Let's discuss differences in Qmin (which is important for gross errors detection) and also the reconciled values of X1 and X2 proper. More details are in Tab. 2.1. above. It can be seen that for the Near distance point (1;1.1) the difference of Qmin between SL and SQL methods is negligible. The same hold for reconciled values, which differs only in the last valid digit. For Medium distance point differences between SL and SQL are in the range of several per cents of reconciled values which are not negligible. The Far distance point (0.5; 2.5) gives SL and SQL results significantly different for Qmin and also for reconciled values. For completeness is presented the course of iterations of the DR process for the SL method followed by SQP:

Discussion about Example 2.2 results:

1. Differences of results between SL and SQL methods depend on model nonlinearity and also on the distance of measured data from reconciled values (on data adjustments).
2. From another point of view, it is good to have some simple indicator of model nonlinearity which includes also the distance of measured data from the model. This distance depends also on uncertainties of individual measured values. We propose the following relative difference of Qmin denoted as Q_{difrel}:

$$Q_{difrel} = (Qmin_{SL} – Qmin_{SQP}) / Qmin_{SL} \qquad (2\text{-}25)$$

It is clear that for linear models the value of Q_{difrel} must be zero. Values of Q_{difrel} for nonlinear data sets in this Example are shown in the next Table:

Table 2.2: Q_{difrel} for Example 2.2

Data set	Q_{difrel}
Near distance	0.001
Medium distance	0.016
Far distance 1	0.020
Far distance 2	0.437

3. Probably it is not possible to set some exact limit of Q_{difrel} for which the SL method should be followed by the SQP method. It was shown that even for small values of Q_{difrel} differences in reconciled values by SL or SQP were not negligible. So we recommend to check for some time Q_{difrel} either by the MCM or in a model on-line implementation. In practice most of models have Q_{difrel} well below 0.01, see the next Table:

Table 2.3: Values of Q_{difrel} for models tested in this report (Chapter 6)

Subsection	Q_{difrel}	Subsection	Q_{difrel}
6.2.1	0.0002	**6.3.5**	0.0000
6.2.2.	0.0039	**6.4.1**	0.0197
6.3.1	0.0017	**6.4.2**	0.0001
6.3.2	0.0007	**6.4.3**	0.0294
6.3.3	0.0000	**6.4.4**	0.0000
6.3.4	0.0002	**6.4.5**	0.0001

But in two industrial size cases the $\mathbf{Q_{difrel}}$ was higher than 0.01 (the coal fired boiler with accessories and the heavy crude oil vacuum distillation system). The SQP solution is not always essential but in some cases it should be used.

3 RANDOM NUMBERS AND PROBABILITY DISTRIBUTIONS

This short Chapter describes how measurements with “random” errors are in RECON generated.

3.1 Random errors with the Uniform distribution

Uniform distribution

The distribution of a random variable is called *uniform (rectangular)*, if the probability density is constant on the whole interval of values the variable can assume. Thus if the range of values of the random variable is the interval <a, b> then the probability density function (PDF) equals.

$f(x) = 1 / (b - a)$ for $x \in \langle a, b \rangle$ (3-1)

$f(x) = 0$ otherwise

and the distribution function is

0 for $x \in < a$

$F(x) = (x - a) / (b - a)$ for $a \leq x \leq b$

1 for $x > b$

Both functions are depicted on Fig. 3.1. The basic characteristics satisfy

$E(x) = (a + b) / 2$ (3-2)

$D(x) = (b - a)^2 / 12$ (3-3)

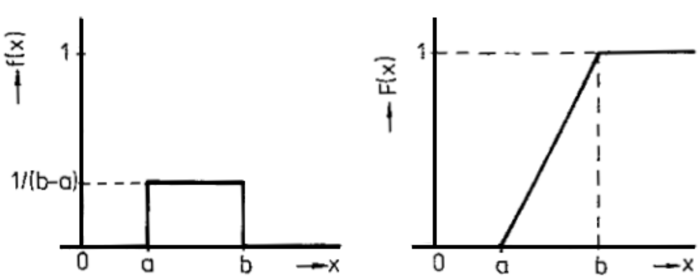


Fig. 3.1: Uniform (rectangular) distribution [probability density function *f (x)* and distribution function *F (x)*]

Uniformly distributed errors play some role in industrial practice (for example it is the distribution of errors when some unmeasured level fluctuating in some interval in a tank is neglected in the mass balance). This distribution is also important for generating other distributions by software generators of pseudorandom numbers (namely generating the Normal distribution). The function generating the Uniform distribution for *a* = 0 and *b* = 1 is available as the standard function in the MS Visual Studio developing package used in RECON (function RND(x)).

3.2 Random errors with the Normal (Gauss) distribution

Normal (Gauss) distribution

The *normal distribution* is the most important distribution of a continuous random variable; under certain circumstances, also some other distributions can be approximated as normal. The probability density of the normal distribution is given by the function.

$$f(x) = \frac{1}{\sigma(2\pi)^{1/2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \tag{3-4}$$

The function is characterized by two parameters, μ and σ, where μ equals the mean and σ the standard deviation of the random variable. The normal distribution is written briefly as $N(\mu, \sigma^2)$. The probability density function is shown in Fig.3.2.

If $\mu = 0$ and $\sigma = 1$, one speaks of the *standard normal distribution* $N(0,1)$.

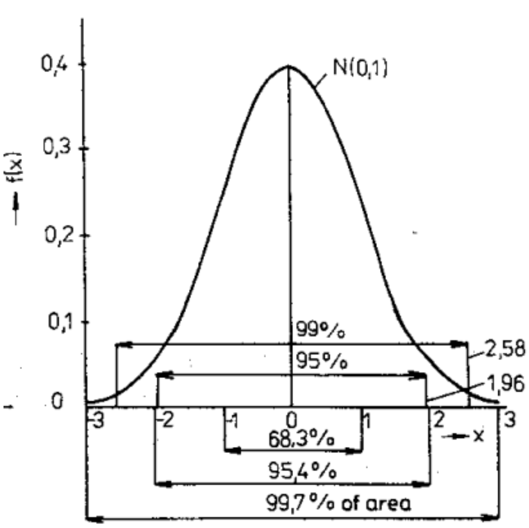


Fig. 3.2: Standard Normal (Gauss) PDF

For generating the Normal distribution Recon uses the Box – Muller method in its "polar" form [20].

The calling of the Normal distribution generation function contains two parameters:

- measured value (mean value of the distribution)
- standard deviation of the measurement error (sigma).

3.3 Chi - square distribution

Let us have υ random variables $U_1, \ldots, U_\upsilon$, mutually uncorrelated, each of them having the distribution *N*(0, 1). The random variable χ^2 defined as the sum of squares of the random variables

$$\chi^2 = U_1^2 + \ldots + U_\upsilon^2 \qquad (3\text{-}5)$$

has the *chi-square* distribution with υ degrees of freedom, denoted by $\chi^2(\upsilon)$. The diagrams of the probability densities of the χ^2 distributions for several degrees of freedom are shown in Fig. 3.3.

The mean E [ ] and variance D [ ] satisfy relations

$$E\,[\chi^2(\upsilon)] = \nu \qquad (3\text{-}6)$$

$$D\,[\chi^2(v)] = 2\nu \qquad (3\text{-}7)$$

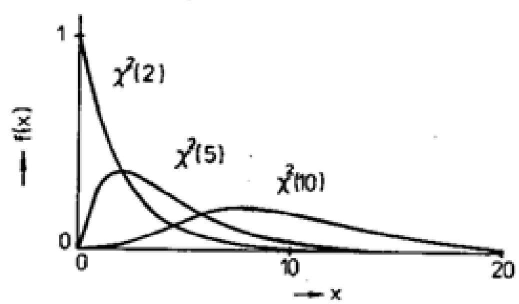


Fig. 3.3: Examples of Probability density functions of the χ^2 distribution.

The quantiles are given in Tab. 3.1.

Table 3.1: Quantiles of χ^2 distribution Q_{crit}

ν	Q_{crit} (95%)	S_{aver}	ν	*Qcrit* (95%)	S_{aver}
1	3.84	0.260	**16**	26.30	0.608
2	5.99	0.334	**17**	27.59	0.616
3	7.82	0.384	**18**	28.87	0.623
4	9.49	0.421	**19**	30.14	0.630
5	11.07	0.452	**20**	31.41	0.637
6	12.59	0.477	**21**	32.67	0.643
7	14.07	0.498	**22**	33.92	0.649
8	15.51	0.516	**23**	35.17	0.654
9	16.91	0.532	**24**	36.15	0.664
10	18.31	0.546	**25**	37.65	0.664
11	19.58	0.562	**26**	38.89	0.669
12	21.03	0.571	**27**	40.11	0.673
13	22.36	0.581	**28**	41.34	0.677
14	23.69	0.591	**29**	42.56	0.681
15	25.00	0.600	**30**	43.77	0.685

Explanation for Table 3.1:

ν No of Degrees of Redundancy = mean (expected) value of Chi-square distribution

Qcrit critical value chi-square distribution for 95 % confidence

S_{aver} mean (expected) value of the Status of Data quality). The Critical value of S_{aver}= 1 irrespective on ν

In Chapter 2, Eq.(2-20) was defined the Status of Data Quality *S*. The mean (expected) value of *S* is

$$E[S(v)] = E[Q_{min}] / Q_{crit} = \nu / Qcrit \qquad (3\text{-}8)$$

4 RECON'S MONTE CARLO FUNCTIONALITY

4.1 The Base case Data Set

MCM requires some prerequisites. The first one is the Base Case model which represents the errorless set of measured data (and of course first guesses of unmeasured variables which will be results of modeling). It is required that the Base Case data set fits exactly the model (all model equations must be zeroed). After that random errors can be added to the Base Case measured data for the MCM Simulation.

This state can be achieved by the menu *Updating guesses* of unmeasured variables and by replacing the original measured values by reconciled values. This is achieved by the menu *Calculate/Update guesses.* See the following Example 4.1:

Example 4.1: Creating the Base Case data set

Let's take the Recon Demo example MC-2.

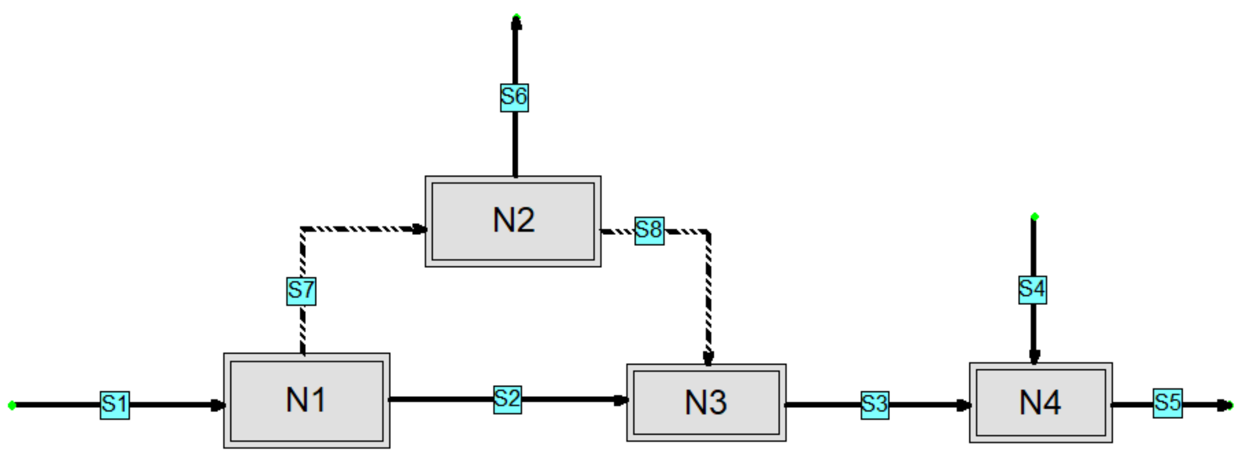


Fig. 4-1: Mass balance flowsheet (4 nodes, 6 measured and 2 unmeasured streams)

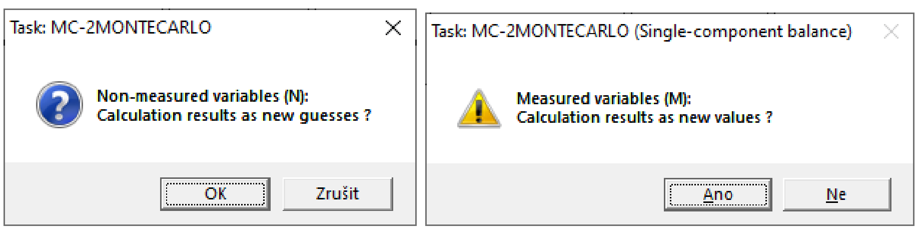
The original example results are

```
RECON 11.9.7-Pro  [ChemPlant Technology s.r.o.]
 Task: MC-2 (Single-component balance)

 I T E R A T I O N S
 Iter            Qeq            Qx              Qy              Qmin
 -----------------------------------------------------------------
 START     1.4944E+01
    1      3.5527E-15    3.0571E-01      2.7931E+01      1.3081E+00
    2      3.5527E-15    2.0136E-15      5.1227E-16      1.3081E+00

M A S S   F L O W R A T E S
 Name             Type      Inp.value      Rec.value      Abs.error
 ------------------------------------------------------------------
 S1               MC          100.100         99.287          1.300  KG/S
 S2               MN           41.100         41.100          1.644  KG/S
 S3               MC           79.000         79.359          1.239  KG/S
 S4               MC           30.600         30.048          2.533  KG/S
 S5               MC          108.300        109.407          2.632  KG/S
 S6               MC           19.800         19.927          0.755  KG/S
 S7               NO           10.000         58.187          2.096  KG/S
 S8               NO           10.000         38.259          2.058  KG/S
```

In this example measured values are reconciled and unmeasured variables are corrected from their first guesses to proper values. After that we can use the menu *Updating guesses.* Two message boxes appear on the screen.



Accept OK/Yes for both questions. After the new calculation the following results appear:

```
RECON 11.9.7-Pro  [ChemPlant Technology s.r.o.]
 Task: MC-2MONTECARLO (Single-component balance)

 I T E R A T I O N S
 Iter            Qeq            Qx              Qy              Qmin
 -----------------------------------------------------------------
 START     2.5121E-15
    1      0.0000E+00    0.0000E+00      3.5527E-15      0.0000E+00

 M A S S   F L O W R A T E S

 Name             Type      Inp.value      Rec.value      Abs.error
 ------------------------------------------------------------------
 S1               MC           99.287         99.287          1.299  KG/S
 S2               MN           41.100         41.100          1.644  KG/S
 S3               MC           79.359         79.359          1.240  KG/S
 S4               MC           30.048         30.048          2.509  KG/S
 S5               MC          109.407        109.407          2.616  KG/S
 S6               MC           19.927         19.927          0.759  KG/S
 S7               NO           58.187         58.187          2.095  KG/S
```

```
 S8                NO            38.259          38.259           2.059   KG/S
```

You can see that the input data completely fit the model and no reconciliation is needed. This is the Base Case ("errorless") data set suitable for MCM.

Note: Realize that by this operation you have irreversibly lost the original measured data (you can try it on the model file copy).

4.2 MCM Simulation

After creation of the Base Case data set and calculation of the task is possible to do MCM Simulations. This is enabled by the menu *Calculate/Monte-Carlo analysis*. The following panel appears:

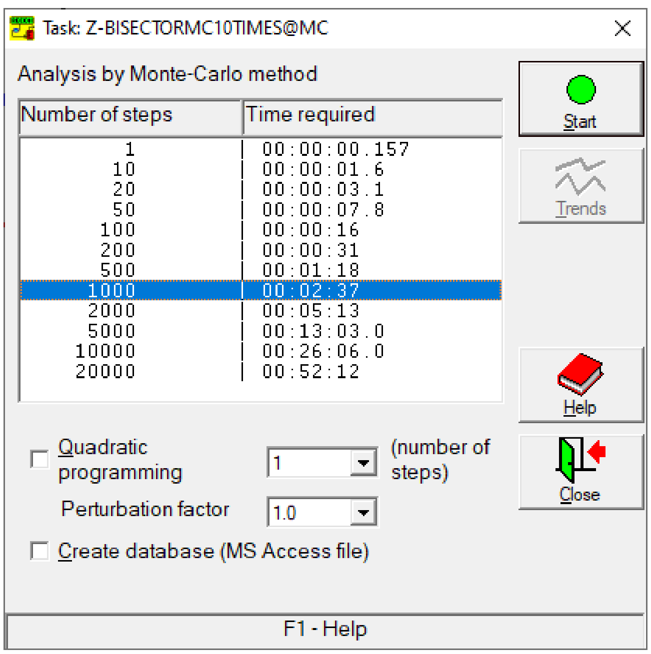


Fig. 4-2: Panel for configuring MCM Simulation

The following functions are available:

- Selecting number of steps (simulations)
- Using the SQP method
- Selecting of the Perturbation Factor in the range <0.5 – 5>
- Creating the MS Access Database for archiving results
- Seeing trends MCM variables.

The Perturbation Factor (PF) serves for changing the magnitude of errors added to measured data from the Base Case data set. PF = 1 means that random errors are generated from the original set of measured data sigmas. Higher values of PF are used for the simulation of large gross errors (the original sigmas are multiplied by PF, this serves mainly for testing models' robustness).

All results can be saved to the Recon's MS Access Database. Trends of variables can be seen after pressing the *Trends* button. The result of the model presented in Fig. 4.1 follows:

```
ANALYSIS BY MONTE-CARLO METHOD
==============================

      Step         Status (S)          Iter.count     Remark
  -------------------------------------------------------------------
         1             0.170                   4
         2             0.433                   4
         3         4.1466E-3                   4
         4             0.157                   4
         5             0.478                   4
         6             0.057                   4
         7             0.200                   4
         8         3.6051E-3                   4
         9             0.418                   4
        10             1.984 !                 5     Gross Error
  -------------------------------------------------------------------
     S-AVG             0.391 Average Status
     S-MAX             1.984 Maximum Status          Step 10
     S-BAD (>1)       10.00 % Gross Error detected

  Qmin-AVG             1.500 Average Qmin
 Qmin-CRIT             3.840 Qmin critical
  Qmin-VAR             5.079 Qmin variance
       DoR                 1 Degree of redundancy
    S-MEAN             0.260 DoR/Qmin-CRIT
  -------------------------------------------------------------------
```

This is the result of MCM Simulation (10 steps).

- In the *Status* column are values of Statuses of data quality. The last step value is greater than 1 – a gross error was detected! Further are shown the Average and Maximum Status values. The S-BAD shows the number of data sets with detected gross error(s) in per cents. As there was no Gross Error present, this is the case of the *Error if the I^{st} kind.*
- The column *Iter.count* shows number of iterations needed for the task convergence
- The *Time stamp* column shows the Date/Time under which results are saved in the Access database for the further analysis. This functionality is important for example for analyzing cases when the calculation did not converged.
- Further lines have the following meaning:
  - `Qmin-AVG` `Average Qmin`
  - `Qmin-CRIT` `Qmin critical (constant)`
  - `Qmin-VAR` `Average Qmin variance`
  - `DoR` `Degree of redundancy (constant)`
  - `S-MEAN` `DoR/Qmin-CRIT (constant for given DoR)`

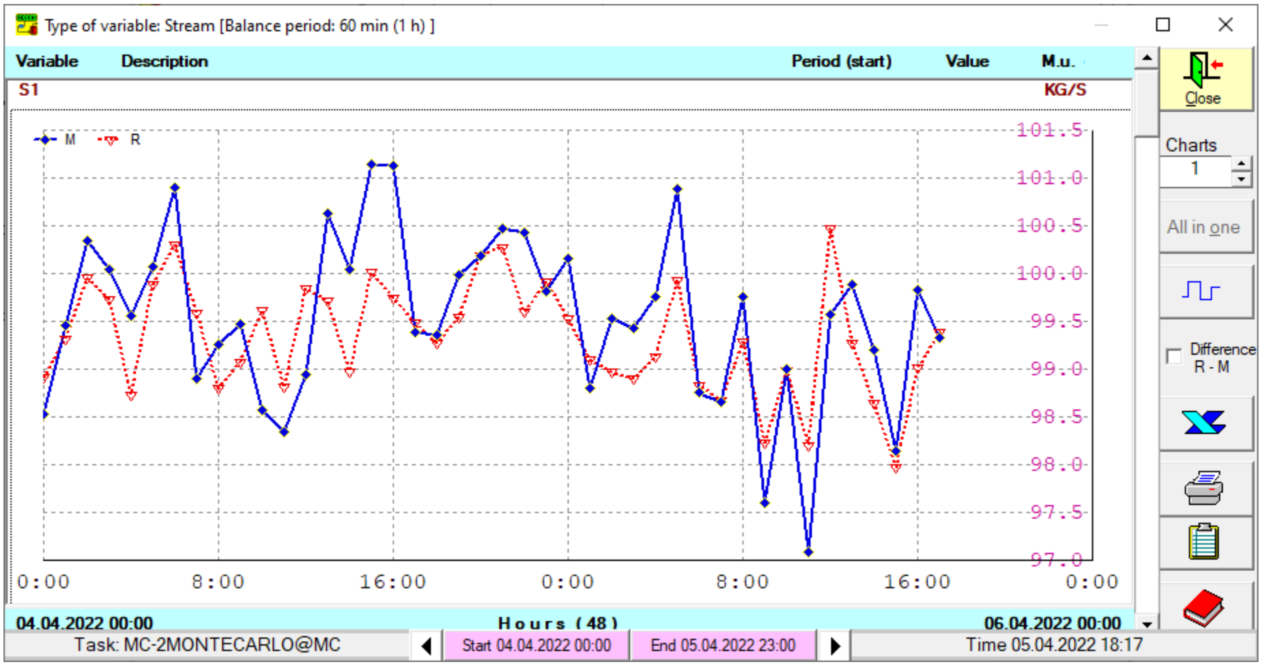


Fig. 4.3: Trend of measured and reconciled values of the stream S1 flowrate (50 simulation runs)

It is possible to download data for the individual simulation steps to Recon a analyze more deeply possible problems in individual steps.

In what follows the influence of models' nonlinearity on DR results will be characterized by the difference between

- Theoretical mean value of the least squares function (which equals to the Degree of Redundancy) and the average of Q_{min} values calculated in MCM simulations (see Eq.(3-6))
- Theoretical mean value of the variance of the least squares function (which equals 2 times the Degree of Redundancy) and the average of Q_{min} variance in MCM simulations (see Eq. (3-7))
- Theoretical value of the probability of the Error of Ist Kind in testing the presence of a gross errors (which equals 5 %) and the relative number of false Gross Error Detection in MCM simulations (in per cents)

5 SIMPLE LINEAR MODEL

The main purpose of this chapter is testing Recon's MCM functions on one simple linear model. Possible differences between values calculated from statistical theory (theoretical expectations) and MCM results proper can have many reasons, namely:

- Generation of random numbers of the uniform (rectangular) distribution in the interval <0 ; 1>
- Generation of random numbers with Normal distribution from random numbers obtained in the previous step
- DR calculations done with limited accuracy
- Limited accuracy of statistical tables (for example critical values of chi-square distribution)
- Limited number trials in MCM simulations.

In the first part will be tested basic model statistics which are

1. Probability of the Ist kind error (GE is detected even if it is not present)
2. Mean value of the Least Squares function Q_{min}
3. Mean value of Q_{min} variance (the Second Central Moment of the distribution).

Next parts concern

4. Predicted uncertainty of results (reconciled and calculated unmeasured variables)
5. Gross errors detectability (probability that a gross error of some size will be detected)

In all cases theoretical (expected) values will be compared with values obtained by MCM (up to 10,000 repetitions (N)).

5.1 Basic model statistics

Let's return to the simple model described in Example 4.1. This model has 6 measured variables and 4 linear equations. There are 2 Degrees of Redundancy. Results of MCM simulation for the Perturbance Factor = 1 are presented in the next table.

Tab. 5-1: Results of MCM simulation for Example 4.1.

	Run									Average		
N	1			2			3					
	GED	Q_{aver}	VQ_{aver}	GED	Q_{aver}	VQ_{aver}	GED	Q_{aver}	VQ_{aver}	GED	Q_{aver}	VQ_{aver}
Expected	**5**	**2**	**4**	**5**	**2**	**4**	**5**	**2**	**4**	**5**	**2**	**4**
10	0.00	2.340	5.385	0.00	1.860	2.981	0.00	1.482	1.607	0.00	1.894	3.324
100	1.00	1.664	2.542	4.00	2.041	3.737	4.00	1.980	2.475	3.00	1.895	2.918
1000	5.50	1.982	4.041	3.60	1.956	3.627	3.90	1.915	3.462	4.33	1.951	3.710
10,000	4.80	1.991	3.924	5.00	1.995	3.987	5.30	2.025	4.210	5.03	2.000	4.040

Here

N is number of simulation runs
Expected in this row are Expected (theoretical mean) values
GED number of cases when a Gross Error was Detected (in %)
Q_{aver} the average value of Q_{min}
VQ_{aver} the average value of Q_{min} Variance.

The expected value of Qmin for DoR = 2 is 2, the expected value of the Variance is 4 (see Eqs. (3-6) and (3-7)). It is clear that the expected value of GED is 5 (probability of the GED test which also equals the probability of the Error of I^{st} kind of this test). Such table will be used for presenting of results throughout this Report).

Results presented in Table 5.1 are visualized in next three figures.

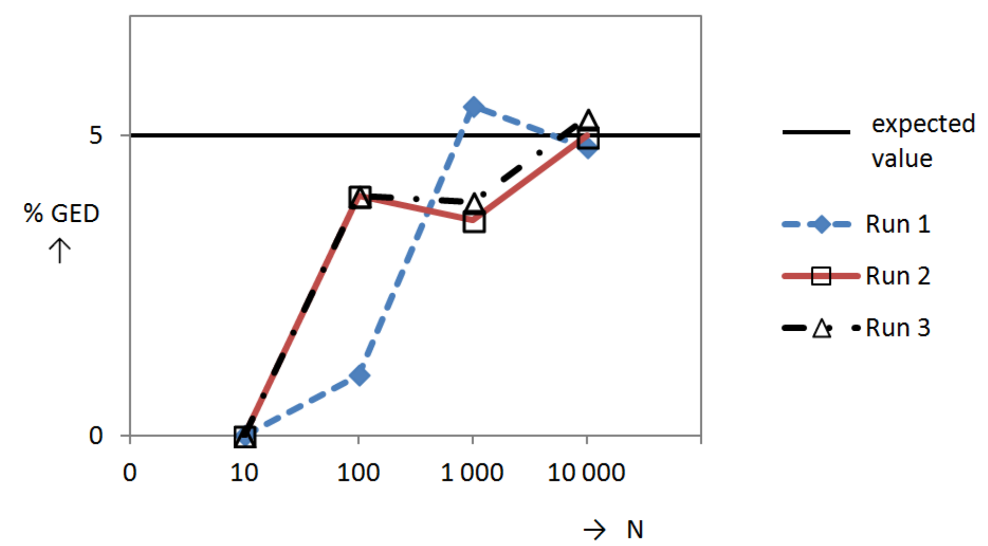


Fig. 5.1: Detected Gross Errors (%) for 3 runs and increasing number of simulations

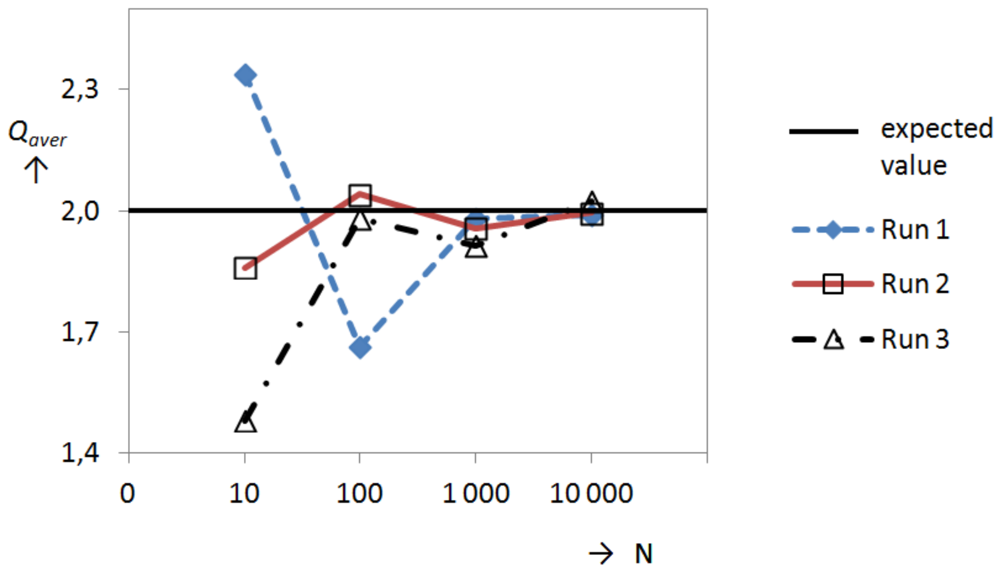


Fig. 5.2: Averages of Q_{min} for 3 runs and increasing number of simulations. The expected value is 2.

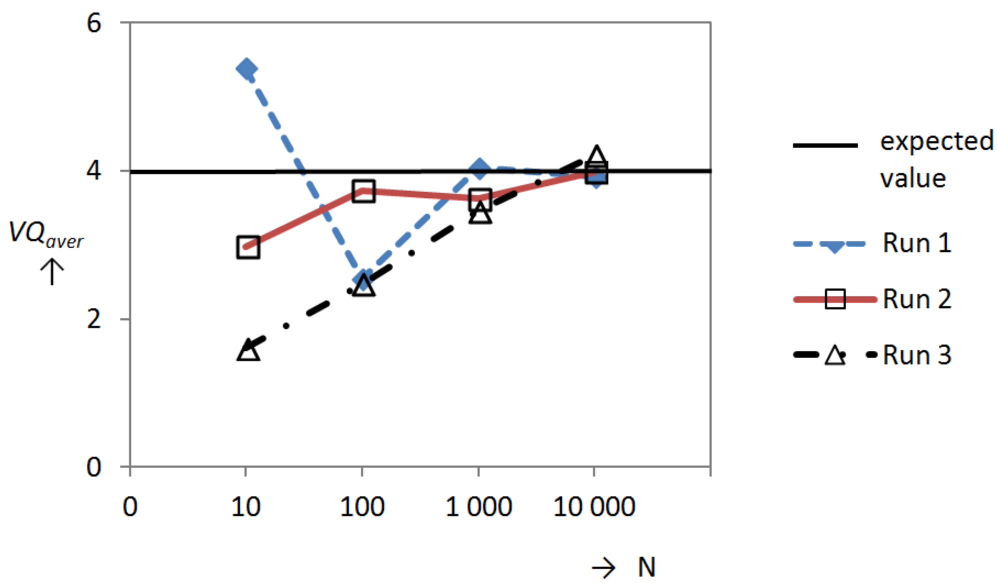


Fig. 5.3: Averages of Q_{min} Variance for 3 runs and increasing number of simulations. The expected value is 4.

From the table and figures above can be seen that with increasing number of repetitions the values of % GED and S_{aver} converge to the expected values. The values obtained from 10,000 simulations are very close to them. Also values obtained for 1000 repetitions may be acceptable (10,000 simulations can be time prohibitive for large industrial models). It is good to realize that our a priori information about measurement uncertainties (maximum errors) in practice are not precise, usually even the second digit is not sure.

We can conclude that this simple linear example confirmed that MCM method contained in Recon gives reasonable results which are in agreement with GED theory (generation of pseudorandom errors is in agreement with chi-square testing by the so called Global test). It is not possible to analyze here the efficiency of MCM (a speed of approaching to the final result which requires the infinite number of repetitions). It is known that this approach is proportional to $1/(\text{No of repetitions})^{1/2}$. With a great simplification we can say that for improving the MCM results' precision by one order we must increase the No of repetitions by two orders. It is well known that MCM is not very fast algorithm.

5.2 Uncertainty of results

The uncertainty of results (reconciled measured values and calculated unmeasured values) is in RECON calculated by the error propagation method. For the Base Case (errorless) data set such results are shown in the next report:

M A S S   F L O W R A T E S

Name	Type	Inp.value	Rec.value	Abs.error	
S1	MC	100.100	100.100	1.308	KG/S
S2	MN	41.100	41.100	1.600	KG/S
S3	MC	79.000	79.000	1.249	KG/S
S4	MC	30.600	30.600	2.510	KG/S
S5	MC	109.600	109.600	2.621	KG/S
S6	MC	21.100	21.100	0.762	KG/S
S7	NO	59.000	59.000	2.066	KG/S
S8	NO	37.900	37.900	2.030	KG/S

You can see that the Input values are identical with reconciled values . The Abs.error column contains uncertainties of results on assumption of normal distribution of errors and 95 % probability level. Between the Absolute Error (AE) and the standard deviation of results holds the simple relation:

$$s = AE/1.960 \qquad (5\text{-}1)$$

where *s* = result's standard deviation. It is therefore easily possible to transform the absolute errors to standard deviations

The MCM proper in this case consists of the following steps:

1. MCM simulation is done for required number of trials N (N=10.000 in our case, index is j))
2. From reconciled and other calculated variables (index i) is for individual variables calculated the sample mean $\underline{X}_i$ (arithmetic average)

$$\underline{X}_i = (\sum X_{ij}) \ / N \qquad (5\text{-}2)$$

3. The sample variance s_i^2 is calculated according to

$$s_i^2 = (\sum(X_{ij} - \underline{X}_i)^2) \ / \ (N - 1) \qquad (5\text{-}3)$$

The sample standard deviations are square roots of sample variances s_i^2.Results are presented in the next table:

Tab. 5-2: Comparison of calculated (predicted) and MCM results

Stream	B.C.value [kg/s]	*s* predicted [kg/s]	*s* MCM [kg/s]	*AE predicted* [kg/s]	*AE MCM* [kg/s]
S1	100.1	0.6662	0.6673	1.306	1.308
S3	79.0	0.6358	0.6372	1.246	1.249
S4	30.6	1.301	1.281	2.549	2.510
S5	109.6	1.361	1.343	2.668	2.632
S6	21.1	0.3836	0.3888	0.752	0.762

where

B.C. value flowrate value Base Case (errorless)
s predicted standard deviation predicted by RECON (method of errors propagation)
s MCM standard deviation obtained from MCM

5.3 Gross Errors detectability

The method for calculating GE detectability was described in Subsection 2.4. The following study starts at the Base Case data (errorless data set which fulfill exactly the model) described in the preceding Subsection 5.2. For this data set we can find Threshold Values TV_β defined in Eq. (2-24). This information is available in Recon's menu *Calculate/Classification*:

```
Task: MC-2AB (Single-component balance)

 REPORT ON CLASSIFICATION OF VARIABLES
 =====================================
All unmeasured variables observable

 R E D U N D A N T   M E A S U R E M E N T S

 Type Variable      Adjustability          Threshold value          Unit
                                    Beta: 90%    Beta: 95%Beta: 99%
 ------------------------------------------------------------------------
  MF  S1                 0.346093        4.802        5.305        6.244  KG/S
  MF  S3                 0.219605        4.648        5.135        6.044  KG/S
  MF  S4                 0.163282        9.951       10.993       12.939  KG/S
  MF  S5                 0.404285        9.951       10.993       12.939  KG/S
  MF  S6                 0.046892        4.802        5.305        6.244  KG/S

 Legend:
  Adjustability   = relative cut of error due to reconciliation
  Threshold value = gross error that will be detected with probability Beta
  Beta            = probability of detecting Gross Error [%]
  MF  Mass flow
```

For example the GE (bias) in the measured flowrate equal to 4.802 kg/s will be detected with the probability 90 %. There were configured 5 tasks with the bias for every of five redundant variables, one by one. The MCM simulation was repeated 10,000 times for every task. Results (for Beta = 90 %) are shown in the next table:

Tab. 5-3: Comparison of calculated (predicted) and MCM results

Stream	B.C.value [kg/s]	Beta 90 % [kg/s]	MCM value [kg/s]	GED%
S1	100.1	4.802	104.902	89.86
S3	79.0	4.648	83.648	89.54
S4	30.6	9.951	40.551	90.20
S5	109.6	9.951	119.551	90.34
S6	21.1	4.802	25.902	89.59

where

B.C. value flowrate value Base Case (errorless)
Beta 90 % Threshold Value for probability 90 %
MCM value biased value of the flowrate for MCM simulation
GED% % of MCM simulations when GE was detected.

It is clear the expected value for the GED% column is 90. The comparison with the table column GED% shows that the MCM results are quite close to the theoretical (expected) value The maximum difference is 0.46 % for variable S3.

6 NONLINEAR MODELS

6.1 Introduction

The linear model (2-3) finds its application mostly in mass balance calculations (Yield Accounting, utilities distribution systems, etc.). Even in such systems there occur frequently needs of including some nonlinear equations. The first question is how to measure model nonlinearity.

For a linear model holds that all first derivatives according to all variables are constant. For example, the mass balance of one node

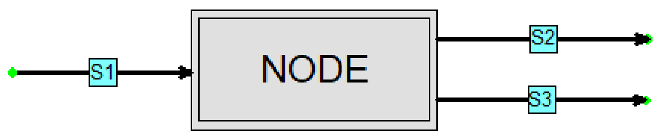


Fig 6.1: Mass balance of one node

with one input stream and two output streams is written (compare with Eq. (2-4).

$$F1 - F2 - F3 = 0 \qquad (6\text{-}1)$$

Matrix of first derivatives of this model according to *F1, F2* and *F3* is

$$\begin{vmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{vmatrix} \qquad (6\text{-}2)$$

This is the incidence matrix of the flowsheet in Fig. 6.1 which contains only constants – this model is linear. Second derivatives of matrix (6-2) are all zero.

The **nonlinearity** of models can be characterized by the Hessian matrix. Let's have a scalar function of a vector **x** of *n* variables $f(\boldsymbol{x})$. The Hessian matrix is the symmetric matrix (*n* x *n*).

$$\mathbf{H}_f = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_1\,\partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_1\,\partial x_n} \\ \dfrac{\partial^2 f}{\partial x_2\,\partial x_1} & \dfrac{\partial^2 f}{\partial x_2^2} & \cdots & \dfrac{\partial^2 f}{\partial x_2\,\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f}{\partial x_n\,\partial x_1} & \dfrac{\partial^2 f}{\partial x_n\,\partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

Or

$$(\mathbf{H}_f)_{i,j} = \frac{\partial^2 f}{\partial x_i\,\partial x_j}.$$

It is clear that for linear functions the Hessian matrix is zero matrix. In other words, the nonzero Hessian is the measure of model nonlinearity.

The simplest case of nonlinear models is the **bilinear** model. Left suppose for example that the enthalpy H of a stream is calculated according to the function

$$H = F\, c_p\, t \qquad (6\text{-}3)$$

where F is the mass flowrate, t is temperature and c_p (constant) is the mean heat capacity of the stream .

The Hessian matrix is then

$$\begin{vmatrix} 0\,;\, c_p \\ c_p\,;\, 0 \end{vmatrix} \qquad (6\text{-}4)$$

The Hessian matrix in this case is nonzero.

The another nonlinear example is the component balance in the form

$$m_1 = F\, c_1 \qquad (6\text{-}5)$$

where m_1 is the mass flow of the component 1 (concentration c_1) and F is the overall mass flowrate.

The Hessian matrix is then

$$\begin{vmatrix} 0 ; 1 \\ 1 ; 0 \end{vmatrix} \tag{6-6}$$

Bilinear models are very frequent in engineering (component and heat balancing). They belong to models with the weak nonlinearity.

There exist more nonlinear models, containing for example logarithmic or exponential functions. Quite dangerous can be the Logarithmic Mean Temperature Difference (LMTD) which is even not defined in the quite common case of equal temperature differences. There exist also complex models for turbine efficiencies, Stodola equations, etc. A significant problem can arise when derivatives are not continuous (for example at phase changes). Luckily some strongly nonlinear calculations frequently need not be included in the main model and can be calculated after the DVR proper is completed (data postprocessing in Recon). This recommendation holds for example for turbine efficiencies and heat transfer coefficients.

Further on we will present in the next three Sections results of MCM simulations for models of increasing nonlinearity and size. Three MCM results will be presented:

GED — number of cases when a gross error was detected (in %). As in all cases no gross error was present, this is the percentage of Errors of the Ist kind. The expected value is 5 %

Q_{aver} — the average value of the Least Squares function. The expected value equals the Degree of Redundancy (see Table 3-1).

VQ_{aver} — the average value of the Least Squares function variance. The expected value equals the Degree of Redundancy times 2. This is the second central moment of the Chi-square distribution

Further in this Chapter we will study 12 models of different types and complexity. In the next table is the summary of basic characteristics of models which are named by their Subsection in this report.

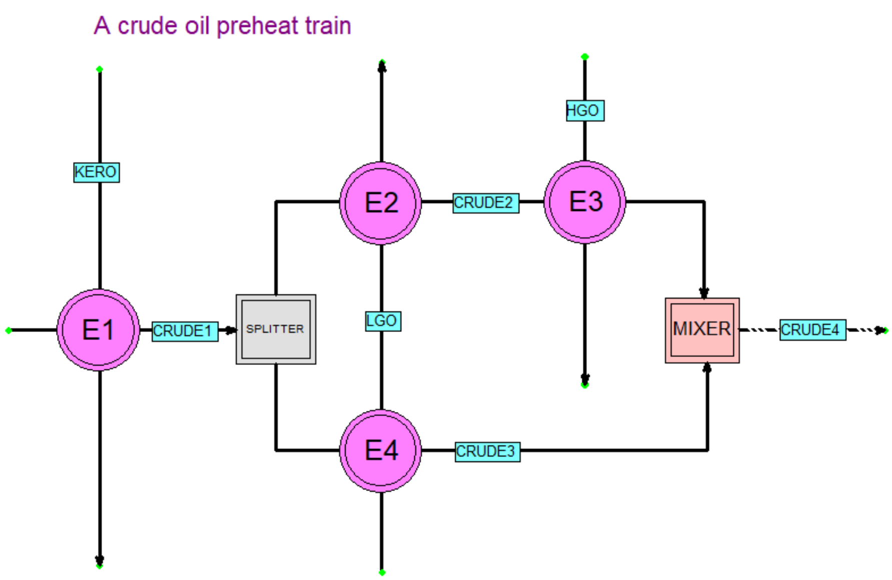
6.2 Bilinear models

In this subsection two small bilinear models will be analyzed:

- A crude oil preheat system
- System of 3 distillation columns (multicomponent balance)

6.2.1 Heat balance – Crude oil preheat

This is the standard Recon’s Demo example E-12.



The crude oil is heated by contact with the kerosene stream, splits and is further heated by the light gas oil and heavy gas oil.

This model has

18 measured variables

7 model equations

5 Degrees of Redundancy.

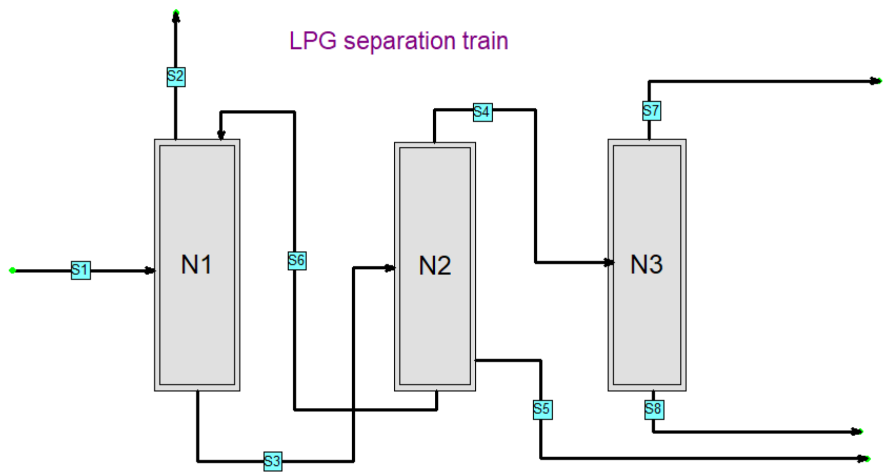
Results are shown in the next table:

Tab. 6-1: Results of MCM simulation, Subsection 6.2.1.

	Run									Average		
N	1			2			3					
	GED	Q_{aver}	VQ_{aver}	GED	Q_{aver}	VQ_{aver}	GED	Q_{aver}	VQ_{aver}	GED	Q_{aver}	VQ_{aver}
Expected	**5**	**5**	**10**	**5**	**5**	**10**	**5**	**5**	**10**	**5**	**5**	**10**
10	10.00	6.174	11.709	0.00	3.592	6.781	20.00	5.427	13.117	10.00	5.064	10.536
100	3.00	5.332	9.169	2.00	4.727	8.197	5.00	4.670	10.120	3.33	4.910	9.162
1000	5.30	4.950	10.203	6.10	4.935	10.562	3.80	4.839	8.351	5.07	4.908	9.705
10,000	5.20	5.005	10.112	5.20	5.016	10.195	5.00	4.967	10.204	5.13	4.996	10.170

6.2.2 Multicomponent balance – LPG separation train

This is the standard Recon's Demo example MC-6.



This model has

38 measured variables (flowrates and concentrations)

23 model equations

5 chemical components

18 Degrees of Redundancy.

Results are shown in the next table:

Tab. 6-2: Results of MCM simulation, Subsection 6.2.2.

	Run									Average		
N	1			2			3					
	GED	Q_{aver}	VQ_{aver}	GED	Q_{aver}	VQ_{aver}	GED	Q_{aver}	VQ_{aver}	GED	Q_{aver}	VQ_{aver}
Expected	**5**	**18**	**36**	**5**	**18**	**36**	**5**	**18**	**36**	**5**	**18**	**36**
10	0.00	17.403	26.020	0.00	13.268	37.549	0.00	21.055	29.407	0.00	17.242	30.992
100	7.00	19.170	37.796	5.00	17.807	33.647	1,00	17.774	29.126	4.33	18.250	33,523
1000	6.30	17.924	40.148	5.60	18.106	36.702	6.20	18.266	38.848	6.03	18,099	38.566
10,000	5.73	18.124	38.932	5.51	18.157	38.871	5.74	18.196	39.050	5.68	18.159	38.951

6.3 Small General models

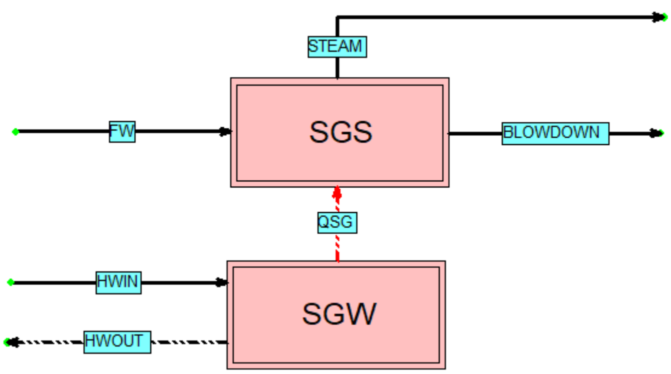
5 small general models will be treated here. They are cutouts from larger industrial models:

- Steam generator in a nuclear powerplant

- Steam to steam heat exchanger in a supercritical boiler
- Component and heat balance of a bisector air preheater
- A simple model of a coal fired boiler with air preheat
- A simple model NG gathering and distribution with hydraulic calculations

6.3.1 Steam generator in a nuclear powerplant

This is the standard Recon’s Demo example E-4.



This model has

10 measured variables (flowrates, pressures and temperatures)

4 model equations

2 Degrees of Redundancy.

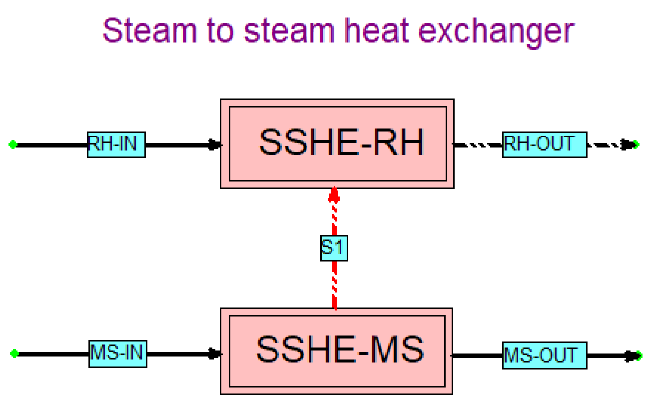
Results are shown in the next table:

Tab. 6-3: Results of MCM simulation, Subsection 6.3.1.

	Run									Average		
N	1			2			3					
	GED	Q_{aver}	VQ_{aver}	GED	Q_{aver}	VQ_{aver}	GED	Q_{aver}	VQ_{aver}	GED	Q_{aver}	VQ_{aver}
Expected	**5**	**2**	**4**	**5**	**2**	**4**	**5**	**2**	**4**	**5**	**2**	**4**
10	10.00	1.858	5.254	0.00	1.860	2.426	0.00	2.221	2.937	3.33	1.980	3.539
100	3.00	1.997	2.927	5.00	2.076	3.751	7.00	2.241	4.753	5.00	2.105	3.810
1000	4.80	2.014	3.777	5.10	2.019	4.253	4.40	1.870	3.949	4.77	1.968	3.993
10,000	5.10	2.006	4.069	5.10	1.993	3.926	5.10	2.023	4.095	5.10	2.007	4.030

6.3.2 Steam to steam supercritical heat exchanger

This exchanger serves for control the reheat steam temperature by contact with the superheated main steam. The pressure of the main steam is supercritical (27 MPa).



This model has

9 measured variables (flowrates, pressures and temperatures)

4 model equations

2 Degrees of Redundancy.

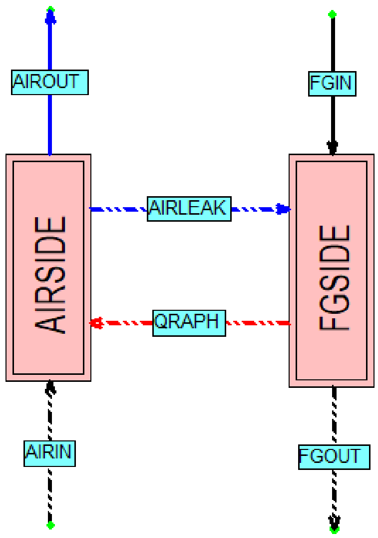
Results are shown in the next table:

Tab. 6-4: Results of MCM simulation, Subsection 6.3.2.

	Run									Average		
N	1			2			3					
	GED	Q_{aver}	VQ_{aver}	GED	Q_{aver}	VQ_{aver}	GED	Q_{aver}	VQ_{aver}	GED	Q_{aver}	VQ_{aver}
Expected	**5**	**2**	**4**	**5**	**2**	**4**	**5**	**2**	**4**	**5**	**2**	**4**
10	0.00	2.343	3.435	0.00	2.042	2.097	0.00	1.953	1.307	0.00	2.113	2.280
100	3.00	1.917	3.830	4.00	1.894	3.972	3.00	1.849	3.345	3.33	1.887	3.716
1000	5.30	2.013	3.840	5.40	1.999	4.050	5.40	1.968	4.309	5.37	1.993	4.066
10,000	5.00	2.019	4.046	4.80	1.958	3.855	4.90	2.000	4.008	4.90	1.992	3.967

6.3.3 Bisector rotary air preheater

This exchanger serves for preheat of air by contact with flue gases. The model combines the multicomponent balance with heat balance. This makes possible to calculate the AIRLEAK flowrate which is important for preheater's diagnostics.



This model has

12 measured variables (flowrates, pressures and temperatures)

12 model equations

6 chemical components

1 Degree of Redundancy.

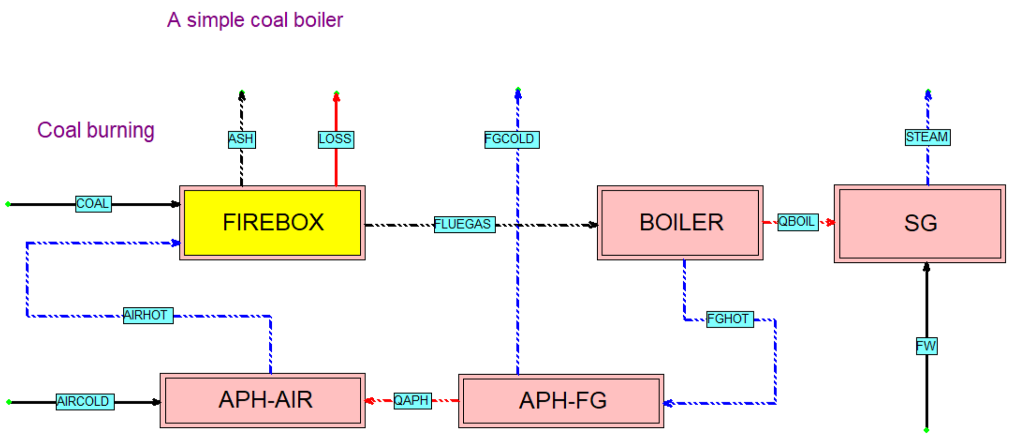
Results are shown in the next table:

Tab. 6-5: Results of MCM simulation, Subsection 6.3.3.

	Run									Average		
N	1			2			3					
	GED	Q_{aver}	VQ_{aver}	GED	Q_{aver}	VQ_{aver}	GED	Q_{aver}	VQ_{aver}	GED	Q_{aver}	VQ_{aver}
Expected	**5**	**1**	**2**	**5**	**1**	**2**	**5**	**1**	**2**	**5**	**1**	**2**
10	0.00	0.757	0.469	10.00	1.381	3.148	20.00	1.612	4.147	10.00	1.250	2.588
100	11.00	1.173	2.805	1.00	0.933	1.442	7.00	1.061	2.137	6.33	1.057	2.128
1000	4.90	1.021	1.948	4.80	1.018	2.188	4.90	0.998	1.767	4.87	1.012	1.968
10,000	4.90	0.993	1.918	4.90	0.998	1.922	5.10	0.997	2.043	4.97	0.996	1.961

6.3.4 Coal boiler with air preheat

A little bit more complex is the model of a coal fired steam generator with air preheat (Demo example E-16). In the FIREBOX is burned coal, the released heat is transferred to the steam generator (SG). Flue gases (FGHOT) transfer the heat (QBOIL) to the air (AIRCOLD).



This model has

22 measured variables (flowrates, pressures and temperatures)

18 model equations

11 chemical components

3 Degree of Redundancy.

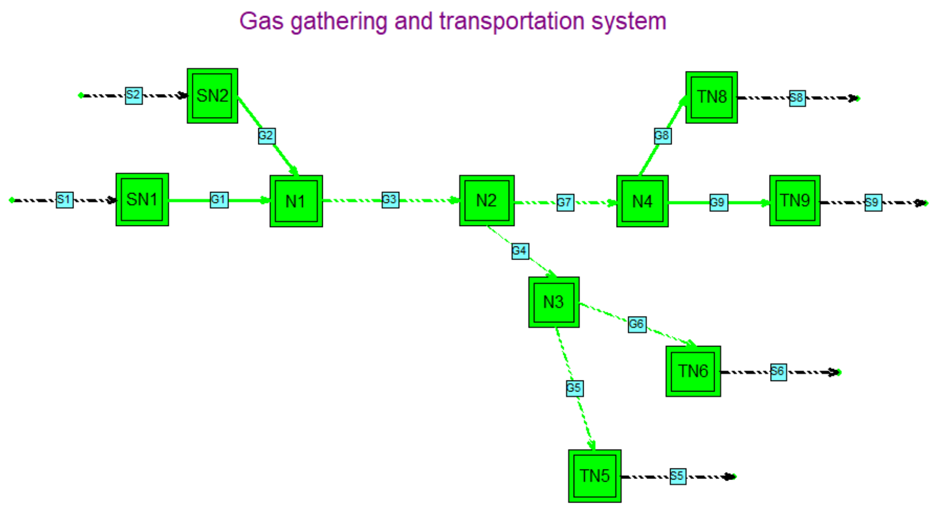
Results are shown in the next table:

Tab. 6-6: Results of MCM simulation, Subsection 6.3.4.

	Run									Average		
N	1			2			3					
	GED	Q_{aver}	VQ_{aver}	GED	Q_{aver}	VQ_{aver}	GED	Q_{aver}	VQ_{aver}	GED	Q_{aver}	VQ_{aver}
Expected	**5**	**3**	**6**	**5**	**3**	**6**	**5**	**3**	**6**	**5**	**3**	**6**
10	10.00	3.047	9.113	0.00	3.100	4.923	0.00	3.060	3.613	3.33	3.069	5.883
100	6.00	2.960	7.502	6.00	2.956	5.746	4.00	3.012	5.087	5.33	2.976	6.112
1000	3.90	2.901	5.354	4.30	2.855	5.185	6.20	3.144	6.665	4.80	2.967	5.735
10,000	5.10	2.991	6.138	4.70	3.020	5.924	5.20	3.021	6.066	5.00	3.011	6.043

6.3.5 Natural gas gathering and transportation with hydraulic calculations

The mass balance of a gas is complemented by hydraulic calculations (momentum balancing which includes pressure drops in individual pipelines.



This model has

10 measured variables (flowrates, pressures and temperatures)

19 model equations

4 Degrees of Redundancy.

Results are shown in the next table:

Tab. 6-7: Results of MCM simulation, Subsection 6.3.5.

	Run									Average		
N	1			2			3					
	GED	Q_{aver}	VQ_{aver}	GED	Q_{aver}	VQ_{aver}	GED	Q_{aver}	VQ_{aver}	GED	Q_{aver}	VQ_{aver}
Expected	**5**	**4**	**8**	**5**	**4**	**8**	**5**	**4**	**8**	**5**	**4**	**8**
10	0.00	3.298	4.884	0.00	3.124	3.780	0.00	2.628	1.107	0.00	3.017	3.257
100	6.00	3.723	9.196	3.00	3.938	5.869	4.00	4.160	7.764	4.33	3.940	7.610
1000	4.70	3.908	8.359	5.50	4.180	8.620	6.80	4.103	9.490	5.67	4.064	8.823
10,000	5.38	4.087	8.516	6.03	4.207	8.870	5.48	4.133	8.501	5.63	4.142	8.629

6.4 Models of industrial size

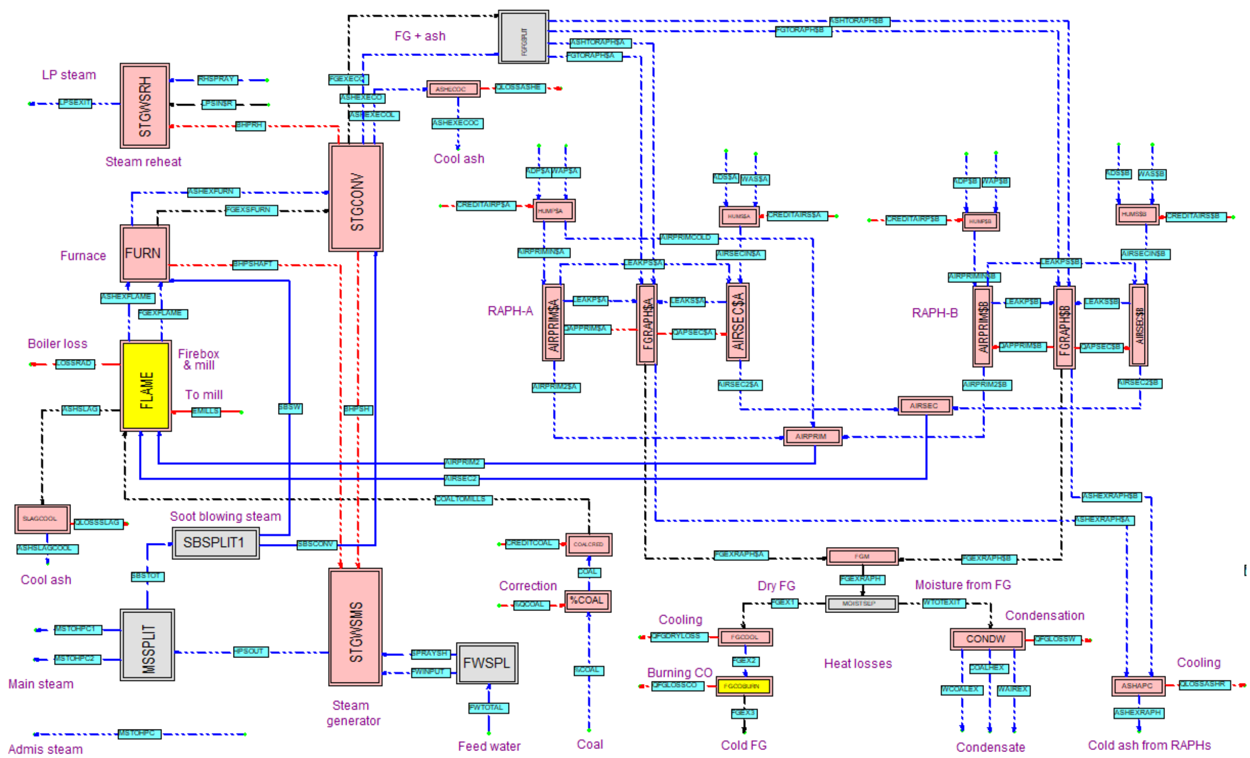
Five models of industrial size will be analyzed:

- Coal fired boiler (powerplant 660 MWe) with accessories (2 trisector air preheaters)

- Steam cycle (steam turbines, heaters, condensers, etc.)
- Vacuum distillation column of heavy crude oil
- Monitoring heat power of a nuclear reactor
- Natural gas transport and distribution system

6.4.1 Coal fired supercritical boiler

This model servers for monitoring of boiler's KPIs (efficiency, losses, etc.). The burning of coal is modeled as the reaction invariant chemical reactor. Besides that the model contains auxiliaries, namely two trisector air preheaters.



This model has

34 measured variables (flowrates, pressures and temperatures)

171 model equations

11 chemical components

7 Degree of Redundancy.

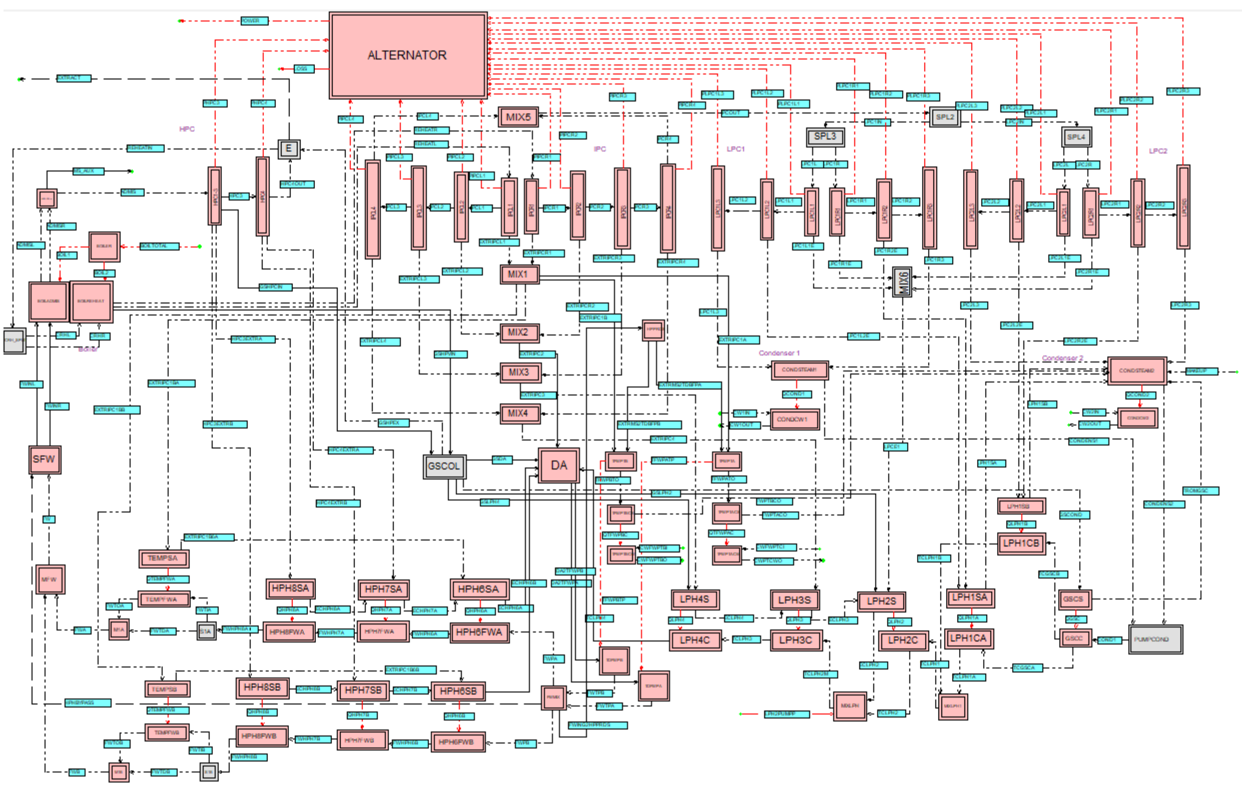
Results are shown in the next table:

Tab. 6-8: Results of MCM simulation, Subsection 6.4.1.

	Run									Average		
N	1			2			3					
	GED	Q_{aver}	VQ_{aver}	GED	Q_{aver}	VQ_{aver}	GED	Q_{aver}	VQ_{aver}	GED	Q_{aver}	VQ_{aver}
Expected	**5**	**7**	**14**	**5**	**7**	**14**	**5**	**7**	**14**	**5**	**7**	**14**
10	10.00	7.963	18.784	0.00	7.642	6.692	0.00	5.908	5.360	3.33	7.171	10.279
100	4.00	6.616	11.624	2.00	6.703	11.170	4.00	6.988	15.017	3.33	6.769	12.604
1000	4.80	7.015	25.895	4.60	7.137	24.727	5.41	8.160	18.215	4.94	7.437	22.946
10,000	5.32	7.924	14.120	5.48	8.158	18.332	5.61	9.005	16.813	5.47	8.362	16.388

6.4.2 Steam cycle of a powerplant

This model servers for monitoring of the Heat Rate of the whole steam cycle. Monitored are also parameters of all substantial equipment (turbines, heaters, condenser, pumps).



This model has

214 measured variables (flowrates, pressures and temperatures)

255 model equations

1 chemical component (H2O)

30 Degrees of Redundancy.

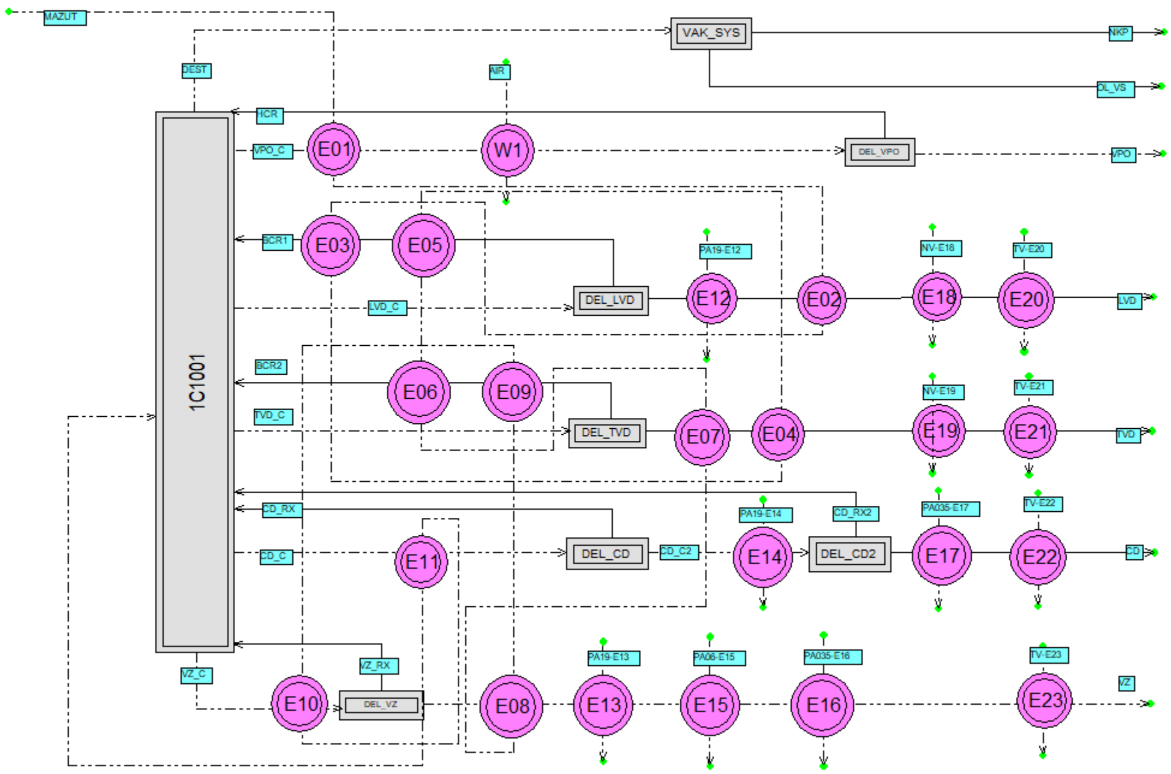
Results are shown in the next table:

Tab. 6-9: Results of MCM simulation, Subsection 6.4.2.

	Run									Average		
N	1			2			3					
	GED	Q_{aver}	VQ_{aver}	GED	Q_{aver}	VQ_{aver}	GED	Q_{aver}	VQ_{aver}	GED	Q_{aver}	VQ_{aver}
Expected	**5**	**30**	**60**	**5**	**30**	**60**	**5**	**30**	**60**	**5**	**30**	**60**
10	10.00	31.174	83.350	0.00	27.105	19.604	10.00	30.178	75.555	6.67	29.486	59.503
100	6.00	30.110	59.301	6.00	30.025	70.027	5.00	31.089	58.133	5.67	30.408	62.487
1000	5.30	29.424	60.656	4.80	29.968	63.336	5.10	29.982	59.796	5.07	29.791	61.263
10,000	6.00	30.220	65.430	5.64	30.256	63.353	5.84	30.289	64.885	5.83	30.255	64.556

6.4.3 Vacuum distillation of heavy crude oil

This model serves for setting up mass and heat balance of a vacuum column including heavy crude preheat and generation of steam. There are 24 heat exchangers of 3 kinds: heat exchange between hydrocarbon streams, water coolers and steam generators.



This model has

58 measured variables (flowrates, pressures and temperatures)

32 model equations

24 heat exchangers (coolers, heaters, steam generators)

9 Degrees of Redundancy.

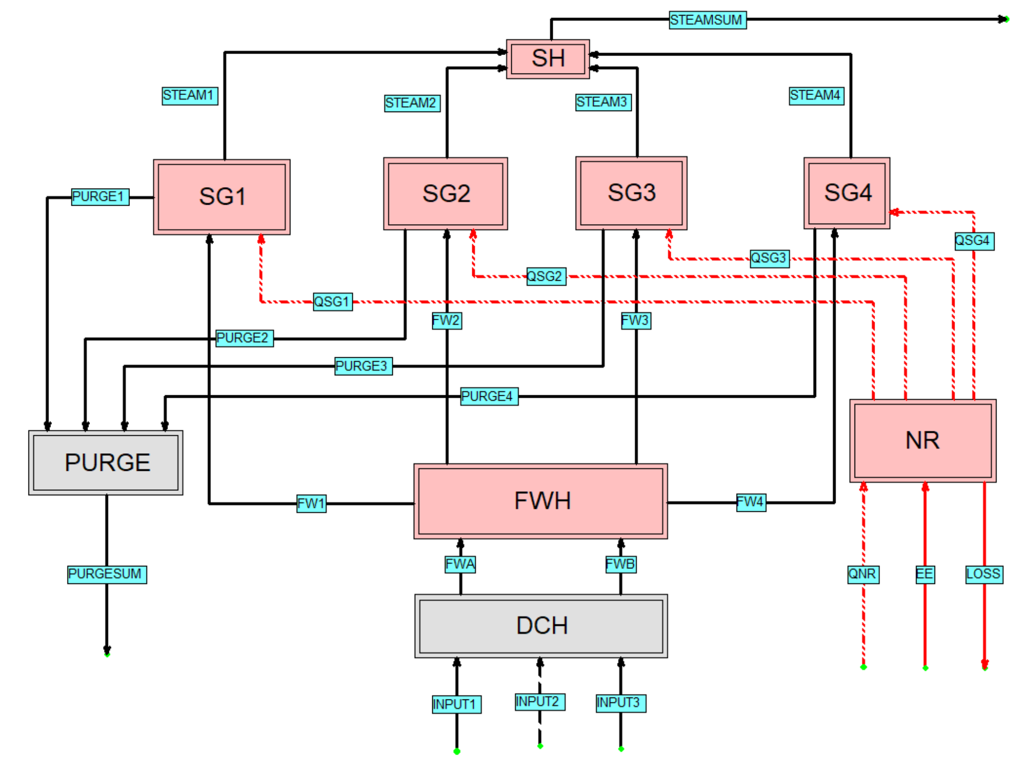
Results are shown in the next table:

Tab. 6-10: Results of MCM simulation, Subsection 6.4.3.

	Run									Average		
N	1			2			3					
	GED	Q_{aver}	VQ_{aver}	GED	Q_{aver}	VQ_{aver}	GED	Q_{aver}	VQ_{aver}	GED	Q_{aver}	VQ_{aver}
Expected	**5**	**9**	**18**	**5**	**9**	**18**	**5**	**9**	**18**	**5**	**9**	**18**
10	20.00	12.358	31.894	0.00	8.334	11.837	20.00	10.312	19.081	13.33	10.335	20.937
100	8.00	9.134	22.464	5.00	9.322	16.266	6.00	9.500	16.654	6.33	9.319	18.461
1000	4.20	8.866	17.180	4.10	8.686	17.143	4.60	9.066	17.273	4.30	8.873	17.199
10,000	5.40	9.100	18.830	5.30	9.042	18.220	5.70	9.100	19.320	5.47	9.081	18.790

6.4.4 *Steam generation system of a NPP*

This model serves for monitoring of a Nuclear Power Plant heat power. The heat balance is based on measured flows of the Feed Water, generated Steam and the Purge streams.



This model has

33 measured variables (flowrates, pressures and temperatures)

15 model equations

4 steam generators

10 Degrees of Redundancy.

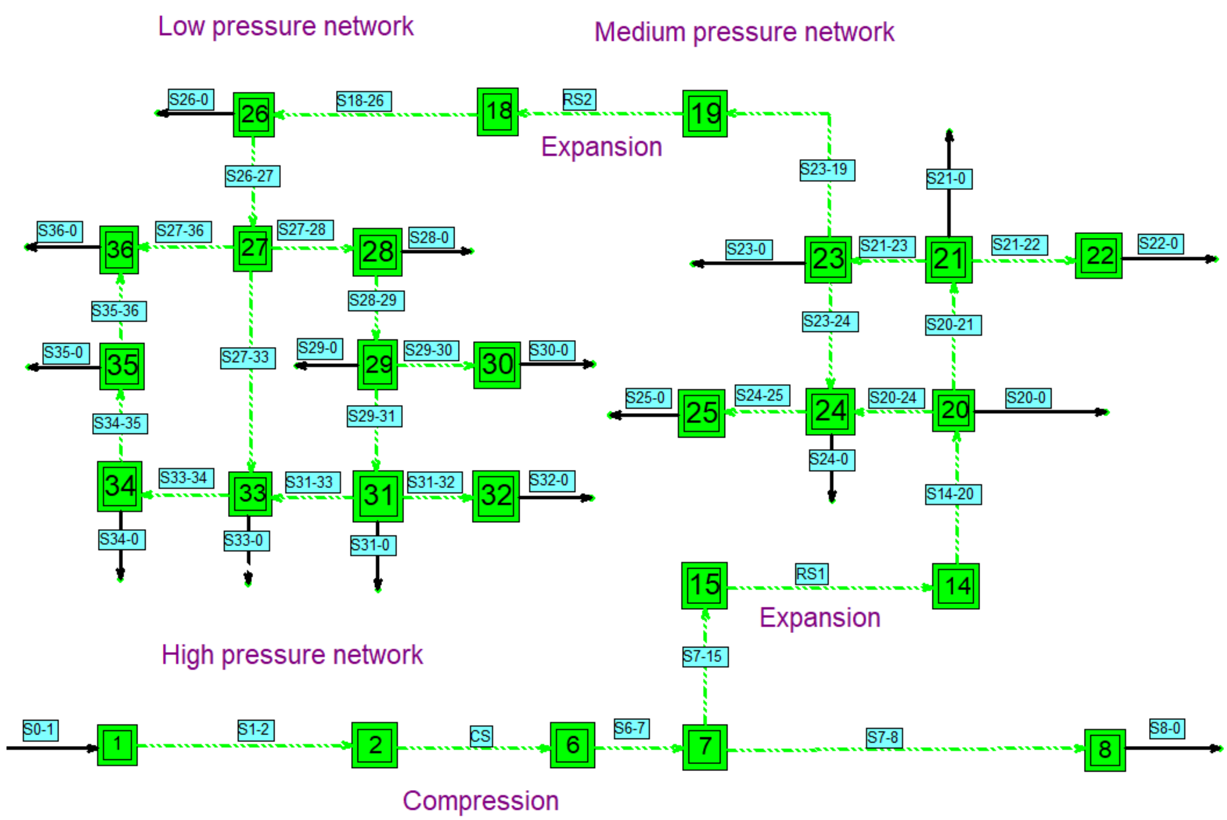
Results are shown in the next table:

Tab. 6-11: Results of MCM simulation, Subsection 6.3.4. Expected S_{aver} = 0.546

	Run									Average		
N	1			2			3					
	GED	Q_{aver}	VQ_{aver}	GED	Q_{aver}	VQ_{aver}	GED	Q_{aver}	VQ_{aver}	GED	Q_{aver}	VQ_{aver}
Expected	**5**	**10**	**20**	**5**	**10**	**20**	**5**	**10**	**20**	**5**	**10**	**20**
10	0.00	7.904	8.979	0.00	8.497	14.359	0.00	10.188	11.213	0.00	8.863	11.517
100	5.00	10.376	19.426	4.00	10.634	19.932	3.00	10,624	16.183	4.00	10.544	18.514
1000	4.70	9.987	20.633	5.20	10.037	20.969	5.30	10.079	21.813	5.07	10.034	21.139
10,000	5.00	10.000	20.070	5.10	9.971	20.610	5.30	10.100	20.610	5.13	10.024	20.430

6.4.5 *Natural gas transport and distribution system*

This model serves for monitoring the mass and momentum balance of the natural gas transport and distribution system. It includes two pressure reduction steps. The mass balance of a gas is complemented by hydraulic calculations (momentum balancing which includes pressure drops in individual pipelines).



This model has

26 measured variables (flowrates, pressures and temperatures)

52 model equations

7 Degrees of Redundancy.

Results are shown in the next table:

Tab. 6-12: Results of MCM simulation, Subsection 6.4.5.

	Run									Average		
N	1			2			3					
	GED	Q_{aver}	VQ_{aver}	GED	Q_{aver}	VQ_{aver}	GED	Q_{aver}	VQ_{aver}	GED	Q_{aver}	VQ_{aver}
Expected	**5**	**7**	**14**	**5**	**7**	**14**	**5**	**7**	**14**	**5**	**7**	**14**
10	0.00	8.050	10.900	0.00	6.774	3.841	0.00	6.876	9.665	0.00	7.233	8.135
100	5.00	7.010	15.670	6.00	7.304	12.461	2.00	6.728	11.728	4.33	7.014	13.286
1000	4.30	6.938	13.752	6.20	7.145	14.521	5.60	7.137	15.312	5.37	7.073	14.528
10,000	5.14	7.056	14.316	5.13	7.039	14.235	4.85	7.057	14.239	5.04	7.051	14.263

6.5 Models' summary

In the next table is the summary of main model parameters:

Tab. 6-13: Main model's parameters Models are named by their Sections in Chapter 6

Sect.	**Task**	N_{meas}	N_{eq}	**DoR**	N_{comp}	F_{unc} [%]	T_{unc} [K]	C_{time} [s]
6.2.1	Crude oil preheat	18	7	5	1	1 – 5	1 – 6	0.1
6.2.2	LPG separation	38	23	18	5	2 – 6	–	0.1
6.3.1	Steam generator	10	4	2	1	2 – 5	1	0.1
6.3.2	Steam to steam exch.	9	4	2	1	3	3	0.1
6.3.3	Air preheater	12	12	1	6	2	1 – 2	0.2
6.3.4	Simple coal boiler	22	18	3	11	1 – 5	1 – 3	0.4
6.3.5	NG gathering	10	19	4	1	5	–	0.1
6.4.1	Coal fired boiler	34	171	7	11	5 – 10	1 – 5	3
6.4.2	PP steam cycle	214	255	30	1	1 – 10	1 – 2%	3
6.4.3	Vacuum distillation	58	32	9	1	1 – 5	2 – 3%	0.1
6.4.4	NPP steam generation	33	15	10	1	2 – 4	1	0.1
6.4.5	NG distribution	26	52	7	1	2 - 10	–	0.2

Legend:

N_{meas} Number of measured values

N_{eq} Number of model equations

DoR Degree of redundancy

N_{comp} Number of chemical components

F_{unc} Uncertainty of measured flowrates

T_{unc} Uncertainty of measured temperatures (mostly in K), two cases in % of ^{o}C

C_{time} typical computing time for 3 iterations of SL (in practice the time can be doubled by the following SQP step). All calculations in this Chapter were done with the SL and then with the SQP method

In the next table are summarized main results of MCM for 12 models from Table -13. They are averages of 3 runs of MCM simulations with 10,000 repetitions.

Tab. 6-14: Main results of Chapter 6

Sect.	Task	GED	Ratio	DoR	Qaver	Ratio	2*DoR	VQaver	Ratio
6.2.1	Crude oil preheat	5.13	1.026	5	4.996	0.999	10	10.170	1.017
6.2.2	LPG separation	5.68	1.136	18	18.159	1.009	36	38.951	1.025
6.3.1	Steam generator	5.10	1.020	2	2.007	1.004	4	4.030	1.008
6.3.2	Steam heat exchanger	4.90	0.98	2	1.992	0.996	4	3.967	0.992
6.3.3	Air preheater	4.97	0.994	1	0.996	0.996	2	1.961	0.980
6.3.4	Simple coal boiler	5.00	1.000	3	3.011	1.004	6	6.043	1.007
6.3.5	NG collection	6.63	1.326	4	4.142	1.031	8	8.629	1.079
6.4.1	Coal fired boiler	5.47	1.094	7	8.362	1.195	14	16.388	1.171
6.4.2	PP steam cycle	5.83	1.166	30	30.255	1.008	60	64.556	1.076
6.4.3	Vacuum distillation	5.47	1.094	9	9.081	1.009	18	18.790	1.044
6.4.4	NPP steam generation	5.13	1.026	10	10.024	1.002	20	20.430	1.022
6.4.5	NG distribution	5.04	1.008	7	7.051	1.007	14	14.236	1.017
	Averages of Ratios deviations from 1	**-**	**0.077**	**-**	**-**	**0.021**	**-**	**-**	**0.041**

Legend:

GED % of cases with detected gross error. Expected value is 5 %

Ratio ratio Average value/Expected (theoretical) value

Qaver Least Squares value. Average of 3 runs of MCM simulations with 10,000 repetitions. The Expected value is DoR

VQaver Variance of Least Squares value. Average of 3 runs of MCM simulations with 10,000 repetitions. The Expected value is 2*DoR

In the last row of the table are averages of deviations from the expected value which is 1 - [ABS(Ratio - 1)]. It is clear that the expected value of this deviation is zero. Results can be also compared with analogical values found in Chapter 5 – Table 5-1. In the next Tab. 6-15 we multiply ratios by 100 to get values in per cents.

Tab. 6-15: Comparison of average absolute deviations in % for linear and nonlinear models

	Deviation GED [%]	Deviation Qaver [%]	Deviation VQaver [%]
Linear model (Table 5-1)	0.6	0.0	1.0
Nonlinear models (Table 6-14)	7.7	2.1	4.1

It can be seen that nonlinear models has in all cases significantly higher ratios. But in practice, the ratios for nonlinear models are still very small. For example, the highest deviation is 7.7 % for the Gross Error Detection ((Error of the Ist kind). This means that the average absolute value was in the interval 5 ± 5*0.077 = 5 ± 0.38. Such distance from the expected value which equals 5 is negligible. Our information about measurement errors is not very precise, we sometimes can’t guarantee even the value of the first digit of the standard error sigma. The same holds also for Qaver and for VQaver.

The conclusion is that even if there is some difference between behavior of typical linear and nonlinear models, this difference is not significant from the practical point of view.

7 INFLUENCE OF MEASUREMENTS' UNCERTAINTIES

As was already shown earlier, aside of model's nonlinearity also measurements' uncertainties play role in statistical treatment of process data. It is clear that the linearization of a curve is justified only in a small vicinity of the measurement point. With increasing measurement errors also errors brought by linearization grows. In other words, for strongly nonlinear function the linearization can be justified only in the small area around the measured value. In this subsection we will study this problem on 2 small bilinear models from the Subsection 6.1. For every model will be tested several model versions differing in increasing values of measured values' uncertainties.

7.1 Crude oil preheat

This is the Demo task E-12 used in the Section 6.2.1. The parameters of model versions are presented in the next table. The uncertainties of flowrates are changed from 1 % to 20 % of the measured value. The uncertainties of temperatures are changed from 1 to 20 K of the measured value For all model versions were calculated 1,000 simulations.

Version	UFlow [kg/s]	UTemp [K]	GED [%]	Q_{aver}	VQ_{aver}
Expected value →	-	-	5	5	10
1	1 %	1	5.8	5.195	10.558
2	3 %	3	5.0	5.002	9.971
3	5 %	5	4.5	4.995	9.582
4	10 %	10	3.9	4.902	9.060
5	20 %	20	4.8	5.194	9.810

where
- Version — Version of uncertainty
- UFlow — uncertainty of flowrates
- UTemp — uncertainty of temperatures
- GED — Gross error was detected (Error of Ist Kind)
- Q_{aver} — average value of Q_{min}
- VQ_{aver} — average value of Q_{min} varaince

7.2 Multicomponent balance – LPG separation train

This is the Demo task MC-6 used in the Section 6.2.2. The parameters of model versions are presented in the next table. The uncertainties of flowrates were changed from 1 % to 20 % of the measured value. For all model versions were calculated 1,000 simulations.

Version	UFlow [kg/s]	GED [%]	Q_{aver}	VQ_{aver}
Expected value →	-	5	18	36
1	1%	5.1	18.028	35.835
2	3%	5.2	18.132	38.770
3	5%	4.4	17.712	37.384
4	10%	5.5	18.221	36.669
5	20%	4.6	18.002	37.970

where
Version Version of uncertainty
UFlow uncertainty of flowrates
UTemp uncertainty of temperatures
GED Gross error was detected (Error of Ist Kind)
Q_{aver} average value of Q_{min}
VQ_{aver} average value of Q_{min} variance

7.3 Conclusions

From two tables in this Chapter is clear that for the two bilinear models can be seen that there is no evident significant influence of measurement uncertainties on basic statistical characteristics of the data reconciliation process.

8 USING MCM FOR TESTING MODEL ROBUSTNESS

For on-line installations of DR systems in industry which are running 24 hours 7 days a week is important their robustness. In an on-line running is too late to solve problems caused by insufficient numerical properties of software, changes in plant hardware configuration, etc. MCM is therefore good for testing DR software before the DR system proper is installed in harsh industrial conditions.

The most simple method for testing DR system robustness are MCM functions of RECON (recall Chapter 4). There is the option of simulating gross errors by multiplying all uncertainties of the model by the Perturbation Factor which is in the range 1 – 5 (recall that PF = 1 means standard random errors). Complete results of MCM can be saved to the MS Access DB from which problematic data sets can be downloaded and analyzed off line.

This chapter is not systematic, only informative. On four previously described tasks (2 small and 2 of industry size) will be shown typical behavior of models in presence of bad data.

The following characteristics of results will presented:

PF	Perturbation Factor
Iter	average number of iterations (SL + SQP)
GED	number of cases when Gross Error was detected [%]
S_{aver}	Average value of the Status of data quality (Status > 1 means that GE was detected)
S_{max}	maximum value of the Status of data quality
$N_{notconv}$	number of cases when calculation did not converged [%]

In 2 cases the number of MCM repetitions was 10,000,- (Sections 8.1 and 8.2), in 2 cases (Sections 8.3 and 8.4) the MCM repetitions were 1000.

8.1 Crude oil preheat

Model from Subsection 6.2.1.

PF	Iter	GED [%]	S_{aver}	S_{max}	$N_{notconv}$ [%]
1	4.32	4.63	0.45	2.44	0.00
2	4.87	73.9	1.81	9.28	0.00
3	4.97	93.80	4.06	23.66	0.00
5	5.01	99.36	11.52	92.19	0.00

8.2 LPG separation

Model from Subsection 6.2.2.

PF	Iter	GED [%]	S_{aver}	S_{max}	$N_{notconv}$ [%]
1	3.00	6.7	0.636	1.54	0.00
2	3.05	98.7	2.591	15.46	0.00
3	3.26	100.0	6.076	77.08	0.00
5	3.67	100.0	18.822	201.59	0.00

8.3 Steam cycle of a powerplant

Model from Subsection 6.4.2

PF	Iter	GED [%]	S_{aver}	S_{max}	$N_{notconv}$ [%]
1	7.66	5.6	0.667	1.21	0.00
2	7.96	100.0	2.94	6.88	0.00
3	8.69	100.0	8.19	23.18	0.50
5	9.65	100.0	19.84	68.97	4.50

8.4 Vacuum distillation of heavy crude oil

Model from Subsection 6.4.3

PF	Iter	GED [%]	S_{aver}	S_{max}	$N_{notconv}$ [%]
1	4.97	5.9	0.543	1.778	0.00
2	5.84	88.9	2.092	6.772	0.00
3	6.16	98.9	4.727	16.508	0.00
5	7.17	100.0	13.801	42.779	0.00

8.5 Conclusions

From results of this Chapter can be seen that models in Sections 8.1, 8.2 and 8.4 are well robust. The model in Section (8.3) (Steam Cycle) has problems with the convergence in the case of MCM Perturbation Factor 3 and 5 (the calculation did not converged in 0.5 % of runs in the case of PF = 3 and in 4.5 % of runs in the case of PF = 5).

The original uncertainties of Flowrates in this case were 1 - 10 % of the measured value and the temperature uncertainties were in the range 1 – 2 % in the Celsius temperature scale (see Table 6-13). In the case of the Perturbation Factor 5 this means very bad input data for calculation (for example flowrate errors up to 50 % of the Base Case value). High temperature perturbations can also lead to phase changes in water/steam streams.

It should be noted here that RECON’s MCM module enables one to save all MCM repetitions data to the MS Access Database. Data can be then manually imported and the cause of problems can be revealed.

9 DISCUSSION AND CONCLUSIONS

The main purpose of Chapter 5 was to verify that MCM methods used in RECON (generation of random variables, etc.) are sound. Calculations revealed that it is needed to make 10,000 MCM repetitions to get reliable results. The MCM analysis of a simple linear model has confirmed that the DRV methodology works and the results' precision agrees with MCM results (Table 5-2). Also the Gross Errors Detectability method gives good results (Table 5-3).

The core of the report is in Chapter 6. The spectrum of 12 nonlinear models covers typical DRV tasks we can meet in Chemical and Power Industries. Models' characteristics are shown in Table 6-13. The typical type of nonlinearity is a product of two variables (bilinear models, namely multicomponent and heat balances). The nonlinearity in all cases did not caused significant deviations caused by models' linearization during the DRV solution (Table 6-14).

In Section 2.5 was proposed the practical and simple measure of models' nonlinearity by Eq. (2-25). It is the relative improvement of the Least Squares function calculated by the Successive Linearization and then improved by the SQP method.

In Chapter 7 were analyzed two bilinear models as concerns the influence of measurement uncertainties on statistical results of DRV. It was concluded there that there is no evidence of significant influence of measurement uncertainties on basic statistical characteristics of the data reconciliation process.

In Chapter 8 was on 4 examples shown that MCM is a good method for testing models' robustness. Random errors of measurement were perturbed up to 5 times of the original measurement uncertainties to test models' robustness.

LITERATURE CITED

[1] Rao, C.R: *Linear Statistical Inference and its Applications.* J.Wiley, NY 1973

[2] Liu, J.S.: *Monte Carlo Strategies in Scientific Computing.* Springer, NY 2004

[3] Mah, R. S. H., 1990, Chemical Process Structure and Information Flows, Butterworths, Stoneham, MA

[4] Madron, F.: Process Plant Performance. Measurement and data processing for optimization and retrofits. Ellis Horwood, Chichester 1992

[5] Veverka, V. V., and Madron, F., 1997, Material and Energy Balancing in the Process Industries: From Microscopic Balances to Large Plants, Elsevier, Amsterdam.

[6] Narasimhan, S., and Jordache, C., 2000, Data Reconciliation and Gross Error Detection: An Intelligent Use of Process Data, Gulf, Houston, TX.

[7] Romagnoli, J. A., and Sanchez, M. C.: Data Processing and Reconciliation for Chemical Process Operations. Academic Press, San Diego 2000

[8] Bagajewicz M. Process Plant Instrumentation. Design and Upgrade (ISBN:1-56676-998-1), CRC Press (2000).

[9] Bagajewicz, M. J., 2010, Smart Process Plants. Software and Hardware Solutions for Accurate Data and Profitable Operations, McGraw-Hill, New York.

[10] Knopf, F. C., 2012, Modeling Analysis and Optimization of Process and Energy Systems, Wiley, Hoboken, NJ

[11] Process data validation in practice. Applications from chemical, oil, mineral and power industries. Report CPT 229-07. Usti nad Labem 2007

[12] C. IORDACHE, R.S.H. MAH, A.C. TAMHANE, "Performance studies of the measurement test for detection of gross errors in process data". AIChE J., vol. 31, (7), p. 1187-1200, 1985.

[13] Ozyurt, D.B., Pike, R.W.: Theory and practice of simultaneous data reconciliation and gross error detection for chemical processes. Computers and Chemical Engineering, 28, 2004 (2004)

[14] Bagajewicz, M. , On a new definition of a stochastic-based accuracy concept of data reconciliation-based estimators. In Proceedings of the 15th European symposium on computer-aided process engineering. Spain 2005

[15] Bagajewicz, M., & Nguyen, D. (2008). Stochastic-based accuracy of data reconciliation estimators for linear systems. Computers and Chemical Engineering, 32(6), 1257–1269 (2008)

[16] Syed M.S. et all: Data reconciliation and suspect measurement identification for gas turbine cogeneration systems. J Eng Gas Turbines and Power, September 2013, Vol. 135: 091701

[17] Cencic, O.: Application of the method of data reconciliation for minimizing uncertainty of the weight function in the multicriteria optimization model. PHD Thesis, Wien Technical University, 2018

[18] Wingstedt, E.and Saarela, O. (2019). Monte Carlo Simulations to Evaluate Error Propagation in Computation of Thermal Power. In 11th Nuclear Plant Instrumentation, Control and Human-Machine Interface Technologies (pp. 721-732). American Nuclear Society (ANS)

[19] Madron,F., Hostalek,M., Stepan,L.: Protection of a Nuclear Reactor Monitoring System against Gross Measurement Errors. International Journal of Nuclear Energy Science and Engineering (IJNESE) Volume 5, 2015

[20] Box–Muller transform - Wikipedia