ChemPlant Technology s.r.o.

Report CPT- 479-25

Date of issue: 6. 3. 2025

This report has 86 pages.

MAKE PROCESS DATA VALIDATION AND RECONCILIATION SIMPLE AND MORE EFFICIENT

DVR FOR DUMMIES

Author: František Madron

DOI: 10.13140/RG.2.2.22760.17920

Ústí nad Labem March 2025

ChemPlant Technology, s.r.o.
V Zahrádkách 999/35
400 01 Ústí nad Labem
Czech Republic

Phone: +420 474 527 221
E-mail: support@chemplant.cz
Web Site: www.chemplant.cz

HISTORY OF REVISIONS

Revision No	Date	Description	Elaborated by	Approved by

CONTENTS

GLOSSARY AND ABBREVIATIONS

a	adjustability of reconciled variables – see Eq (3-18)
DoR	Degree of Redundancy
DR	Data Reconciliation
DRE	Data Reconciliation Engine
DVR	Data Validation and Reconciliation
GE	Gross Error
GED	Gross Errors Detection
i	indices – i-th variable
KPI	Key Performance Indicator
MCM	Monte Carlo Method
MSSP	Monitoring System Self-Protection (Section 4.6)
OLM	On-line Monitoring
OS	Operating State
PS	Parametric Sensitivity
q_i	dimensionless threshold value, see Eq (4-7)
Q_{crit}	critical value of the Least Squares sum (synonym for $\chi^2_{(1-\alpha)}(\nu)$)
Q_{min}	the Least Squares sum
RS	Reference State
S	Status of Data Quality
TSM	Taylor Series Method
TV	Threshold Value
$\mathbf{W}$	covariance matrix of measurement errors
$\boldsymbol{W}_{\boldsymbol{x'}}$	covariance matrix of reconciled values
$\boldsymbol{W}_{\boldsymbol{v}}$	covariance matrix of adjustments
x_i	i-th measured variable
α	Greek letter Alpha – statistical parameter; *significance level*; probability of Ist kind error ; $(1 – \alpha)$ is the *confidence level*; $\alpha = 0.05$ throughout this report
β	Greek letter Beta – probability, see Fig. 4.1
δ	Greek letter Delta – statistical constant, see Eq. (4-6)
ζ_i	Greek letter *Zeta* – parametric sensitivity, synonym for PS
ν	Greek letter *Nu* – synonym for Degree of Redundancy (DoR)

χ^2	Greek letter *Chi*, χ^2 distribution
$\chi^2_{(1-\alpha)}(\nu)$	critical value of the χ^2 distribution with ν degrees of freedom
σ	Greek letter *Sigma*, standard deviation of a random variable
σ^2	variance of a random variable
σ_i	standard deviation of measurement error of *i*-th variable
$\sigma_{x'i}$	standard deviation of reconciled value of *i* th variable
σ_{Vi}	standard deviation of adjustment of *i*-th variable

1 INTRODUCTION

Methods of process Data Validation and Reconciliation (DVR) are continuously developed in process and power industries (chemicals, oil & gas processing, minerals processing, power generation and distribution) since sixties of the past century.

DVR is the method for modelling industrial processes on the basis of mass, energy and momentum balancing, thermodynamic calculations, KPIs evaluation and optimization based on industrial process data

Important is not only DVR proper but also related techniques like the optimal instrumentation placement, process data driven simulation and some others. There exist hundreds of good papers about DVR which were compiled into several textbooks [2 – 9].

There are three main areas of using DVR in process industries:

1. Mass balancing (linear models) which are used in so called Yield Accounting. Such balancing systems are nowadays standard applications in refineries, tank farms and petrochemical complexes. The frequency of balancing is low, usually one day.
2. On Line Monitoring (OLM) of continuous processes in process industries, including power stations. The frequency of DVR evaluation ranges from minutes to hours. Such systems can be used for example as the decision support for operators, long term monitoring of process economy or as a part of higher level of the process control (Real Time Optimization, Digital Twins, etc.).
3. Performance Tests of producing units after commissioning or after revamps. In the case of such ad hoc tests the plant instrumentation is usually complemented by special portable and temporarily installed instruments. For this area is typical the long-term preparation of measurement, the possibility to repeat it in the case of some problems and some other specialties which are different from classical OLM.

This report concerns mainly OLM area which covers also Yield Accounting as its simpler branch.

Note 1.1: There exists another branch of DVR – estimation of electric power networks, see for example the book [17]. This area of DVR was developed since seventies of the past century independently on the main stream of DVR in Process Industries. It is interesting that statistical methods and results in both areas are very similar■

The purpose of this report is to:

- present shortly the basic theory behind DVR
- mention some new trends and methods which are still not commonly used by the DVR community in practice
- present several short examples illustrating the impact of DVR on monitoring, performance analysis and optimization of operating plants

The mathematics of DVR will be limited to the minimum as there exist several books on this subject. The centroid of this text is the DVR philosophy and the application of the common sense applied to results of DVR provided by some clever software. I will not describe detailed algorithms of a DVR solution.

This report is not a manual how to create a new DVR software. The application of DVR in the harsh industrial environment requires a good professional software (the most valuable details of such software are proprietary and are never published). The main practically oriented text is accompanied by numbered *Notes* printed in petite fonts containing some peculiarities of DVR, historical conjunctions, and the like. Notes can be skipped during the first reading without a loss of continuity of the main text.

Chapters 2 – 4 describe DVR fundamentals and methods. Chapter 5 shows how DVR can help in other engineering activities connected with the overall plant improvement – in other words, how to recast better information about a plant into money. Chapter 6 is about DVR Data Management which has some special features different from other areas of data processing.

DVR techniques will be illustrated throughout the text by (a) the simple linear model of a mass balance, and (b) the simple nonlinear model of a heat exchanger. Input data and basic results of these simple tasks are presented in Appendices 2 and 3. This report also contains Chapter 7 which describes the more complex case study of monitoring the Nuclear Reactor Thermal Power.

All model examples mentioned above can be configured and calculated with the aid of the DVR software Recon which can be downloaded freely from https://www.chemplant.cz/inpage/downloads/ (the Lite version).

The subtitle of the present report does not concern abilities of possible readers. Its target is to show that good DVR results can be reached by very simple and straightforward way (only simple is perfect). The selection of techniques used in this report and the literature cited is limited. The selection is based on my professional life experience with dozens of DVR industrial projects and the long term 24/7 on-line management of two large industrial DVR systems. It is based also on my experience gained during more than 30 DVR courses and seminars held on four continents.

Let's cut to the chase: DVR in Practice.

I will appreciate comments about this report. I especially like the negative feedback.

frantisek.madron@chemplant.cz

2 WHY TO RECONCILE PROCESS DATA?

Next figures show typical situations in process industries where measurement errors cause inconsistencies which should be resolved:

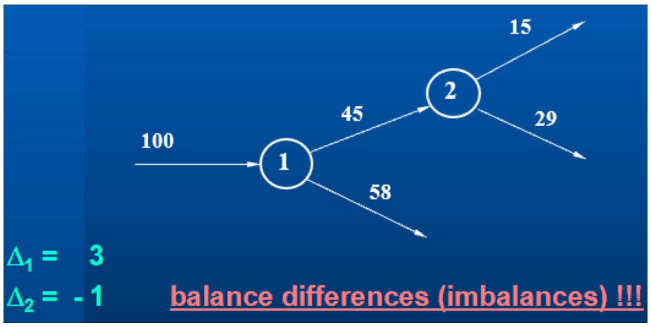


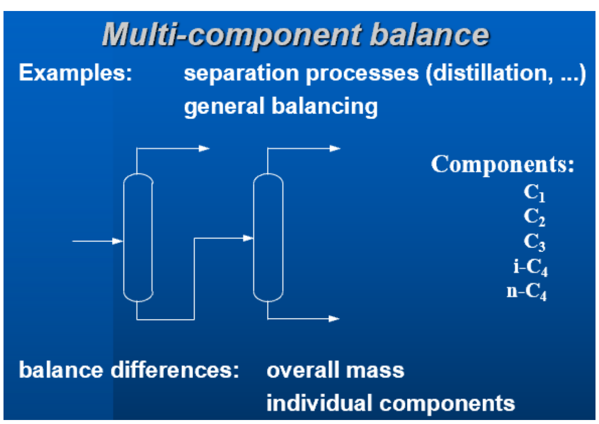
Fig. 2.1: Mass balance

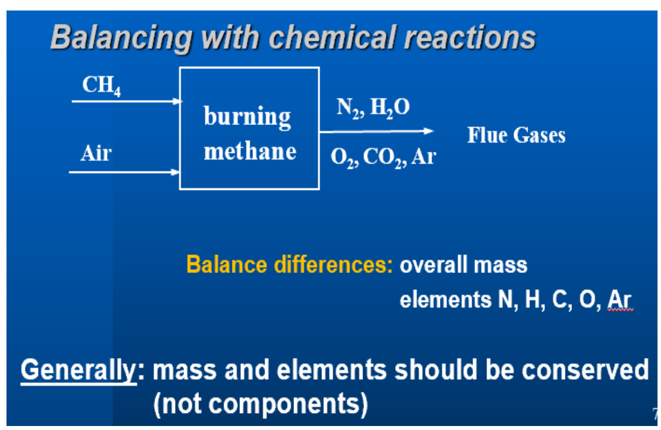
In this case we can see so called single-component balance – we can imagine for example a water distribution system. There are 2 balancing nodes (1,2) and 5 measured streams. We suppose that the system is completely tight. For every node we can calculate the *imbalances* which are differences of nodes' *inputs* and *outputs*:

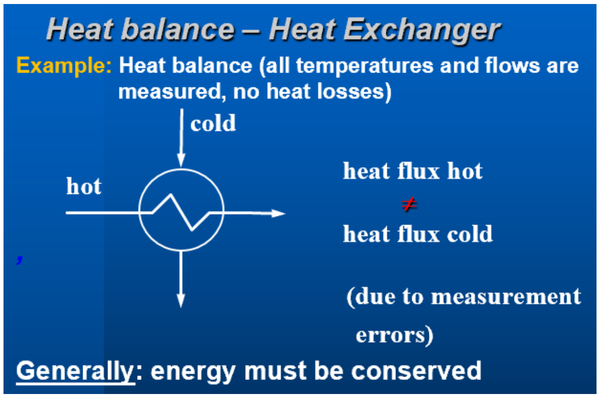
Δ_1 = 45 +58 – 100 = 3

Δ_2 = 15 +29 – 45 = -1

The imbalances (sometimes called improperly "losses") are caused by measurement errors which are inevitable part of every measurement. Theoretically, the imbalances should be zero because there holds the First Law of mass conservation. Below are some other examples typical for process industries:







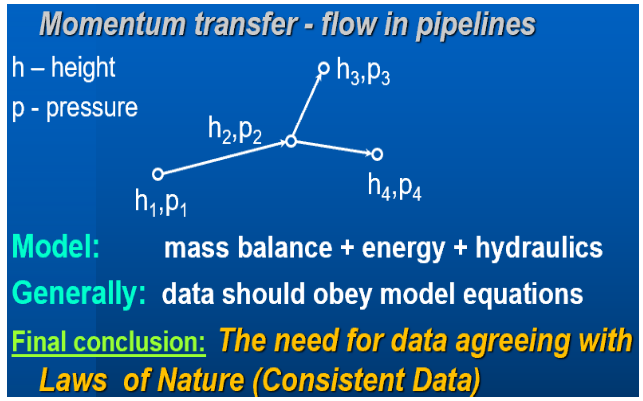


Fig. 2.2: Examples of imbalances

1. Multi-component balance (separation of light hydrocarbons)
2. Chemical reactor – burning of methane
3. Energy (heat) balance of a heat exchanger
4. Momentum transfer – a flow in pipelines. This model requires measurement of flowrates, pressure, temperature and geodetic height.

Generally, there is the need of consistent data which agree with laws of nature (and also laws of accounting). Such *reconciled* (*adjusted*) *data* can be easily incorporated into corporate business models. As will be seen later, there are also other benefits of Data Reconciliation (DR).

Note 2.1: Data reconciliation by the least squares method is not new. It was discovered probably independently by A.M.Legendre (published 1806 and introduced its name – method of least squares), R.Andrain (1808) and K.F.Gauss (1809) who used it allegedly in his works already in 1802. All first applications were from area of astronomy and geodesy. The massive use of DR in the pre-computer era (between two wars) was in geodesy and cartography (reconciliation of measurement of angles and distances in triangular networks and leveling measurements needed in cartography). Method of least squares became then a standard method of data processing with many papers presented every year. The first applications in process industries were reported in early sixties in area of mass balancing (linear models of mass balance typical for crude oil processing). Since then, further problems were solved in hundreds of research papers (nonlinear component and energy balance models, dynamic balancing of nonstationary processes, gross errors detection and identification, optimal placement of instruments, and others). DVR is nowadays the standard matured technique used in industrial data processing■

3 MODELING INDUSTRIAL PROCESS SYSTEMS

Next Chapters 3 and 4 summarize 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 [2-9]. The notation is taken over mostly from the book [3] which is the first book devoted fully to Data Validation and Reconciliation. Book [3] can be freely seen at the ResearchGate Web.

3.1 Models

DVR is based on two types of models:

(a) models of measurement errors

(b) physical models of industrial systems.

3.1.1 Measurement errors

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

$$x^+ = x + e \tag{3-1}$$

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. The standard deviation is supposed to be related with the *uncertainty* of the measured value. In technical practice is usually supposed that the uncertainty equals 1.96 times the standard deviation of the measurement error σ. This follows from the Normal distribution and the confidence level 95 %.

In practice we deal with vector $\boldsymbol{x}$ of measured variables x_i, i = 1,2,…,I with vector $\boldsymbol{\sigma}$ of σ_i . In the general case the model of errors is formed by the covariance matrix. The square covariance matrix (say $\boldsymbol{W}$) contains elements $W_{ij} = \sigma_i\sigma_j$. On the diagonal of the matrix (i=j) are so called variances of errors σ_i^2.

The frequently asked question is: **Where to find values of σ or uncertainties?** In [10,11] are defined two *types* of uncertainty: *Type A* uncertainty is estimated on the basis of measured data. *Type B* uncertainty is estimated by other methods (information from instrumentation vendors, published information, theoretical analysis of the measurement process, etc.) The *Type B* uncertainty is typical for application of DVR in the industrial practice.

Note 3.1: There are many papers about on-line estimation of error's covariance matrices from measured data. The main problem in this approach is that there are mixed two different things – variances and covariances of measurement errors and variances and covariances of real measured signals which are mostly caused by process control systems. In using such methods I am very conservative■

Covariances among errors are difficult to estimate. In general, it is supposed that measurement errors are composed of a number of *elementary* errors (this is also the basis of assumption of normal distribution of measurement errors based on the *Central Limit* Theorem). Covariances originate when one or more elementary errors participates in measuring of two or more variables simultaneously. The estimation of error covariances can be based mostly on theoretical analysis. To give an example, let's see Fig. 3.3 where is the triple measurement of the flowrate. There is one orifice with 3 independent pressure sensors measuring 3 pressure differences. The errors of 3 flowrates calculated by this measurement system are correlated. All flowrates have the common error caused by the error of the orifice proper which causes the covariance among 3 flowrates.

Note 3.2: In general, finding the plausible estimates of measurement uncertainties is not easy and it requires some experience and basic knowledge of measurement techniques and measuring instruments. It is my opinion that frequently even the first digit of the supposed uncertainty is not valid. There are also other areas of measurement theory which are not completely known, like random errors distributions. Anyway, the concern about this issue should not be exaggerated. In the phase of the model implementation you will be confronted with real data and possible issues can be revealed and cleared■

3.1.2 Physical models

Besides the model of measurement errors (3-1), DVR needs also the mathematical model of the industrial process itself. As was already stated earlier, the most common is the model based on First Laws of nature (mass, energy and momentum balances) complemented by further thermodynamic calculations and empirical submodels.

Models can be classified also as stationary and nonstationary [3]. Typical nonstationary models are balances of tank farms with variable inventories or balances of batch processes. Nonstationary models require some special data treatment. You can find more about balancing of nonstationary processes in Section 6.2.

Example 3.1: Model of a Heat exchanger

The heat exchanger in Fig. 3.1 serves for exchanging heat between the COLD and HOT streams.

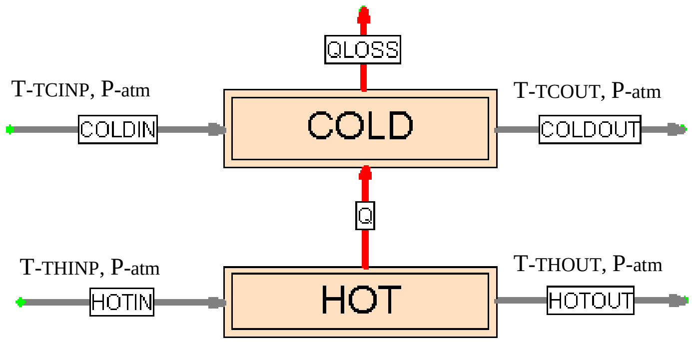


Fig. 3.1: The mass and heat balance scheme of the heat exchanger

The model has 2 nodes (COLD and HOT sides of the exchanger), 4 mass streams and 2 heat fluxes (exchanger heat flux Q and heat loss QLOSS). The mass flows are measured at the inlets to the exchanger (outlets are unmeasured), measured are also all input and output temperatures. The pressure is atmospheric. Enthalpies of water streams are calculated as functions of temperature and pressure according to IAPWS IF-97 method. The heat loss from the shell to the environment is approximately known (estimated).

The model generates altogether 4 balance equations – 2 mass balances and 2 energy balances around both of the nodes. The model has three unknowns – the heat flux Q through the exchanger (red energy stream) and two unknown flowrates at the outlets from both nodes. The equations of the model are:

(1) $F_{HOTIN} - F_{HOTOUT} = 0$

(2) $F_{COLDIN} - F_{COLDOUT} = 0$

(3) $F_{HOTIN}*ENT(T_{HINP},Patm) - Q - F_{HOTOUT}*ENT(T_{HOUT},Patm) = 0$

(4) $F_{COLDIN}*ENT(T_{CINP},Patm) + Q - F_{COLDOUT}*ENT(T_{COUT},Patm) - QLOSS = 0$

The fifth equation of the model is the definition of the Heat Transfer Coefficient HTC for the countercurrent heat exchanger.

(5) $Q - HTC*A*LMTD(T_{HOTIN}, T_{HOTOUT}, T_{CINP}, T_{COUT}) = 0$

where	F* are	flowrates,
	T*	temperatures
	ENT(T*,P*)	water specific enthalpy function
	HTC	Heat Transfer Coefficient
	A	heat transfer area
	LMTD	Logarithmic Mean Temperature Difference

There are the following vectors in the model:

- 5 model equations
- 6 measured variables (2 flowrates, 4 temperatures, heat transfer area, heat loss flux)
- 4 unmeasured variables (2 flowrates, heat flux of the exchanger Q, HTC)

There are present process variables (temperatures, pressure, flowrates), equipment parameter A, fluxes of energy and the model parameter HTC■

The model described above can be symbolically written in the form

$$\boldsymbol{F(x,y,c)} = \mathbf{0} \qquad (3\text{-}2)$$

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 (not present in the Example 3.1)

Typical measured variables ***x*** are directly measured (field) data like flowrates, temperatures, etc. Vector ***y*** contains usually unmeasured process variables but also thermodynamic variables (like specific enthalpies), parameters of models (heat transfer coefficients, turbine efficiencies, etc.), KPIs, etc.

Note 3.3: Here should be noted that practically all statistical theory available holds strictly for linear models only [1]. The exit from this trap is the linearization of models by the Taylor Series Method. After the linearization, **results hold for the original nonlinear model only approximately**. The approximation depends on the model nonlinearity and also on the distance between true and measured values (size of measurement errors). The detailed analysis can be found in [18]■

The important simplification of the nonlinear model (3-2) is so-called ***General Linear model*** (3-3) which can be obtained by linearization of the model (3-2) by the Taylor Series Method [10]:

$$\boldsymbol{A'x} + \boldsymbol{B'y} + \boldsymbol{a'} = \mathbf{0} \qquad (3\text{-}3)$$

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** [3] of unmeasured variables ***y*** to the form containing only measured ones. This model can be used for DR proper:

$$\boldsymbol{Ax} + \boldsymbol{a} = \boldsymbol{0} \qquad (3\text{-}4)$$

where

a is vector of constants
A is the matrix of constants

Note 3.4: Physical models are generally classified as *linear* and *nonlinear*. Linear models were the first models where DVR theory was developed and which were applied in practice (mass balance used mainly for Yield Accounting in crude oil refineries and petrochemical complexes). The solution of linear models was possible on the basis of the graph theory which has simplified the solution significantly. This was important at the time of advent of computers. Later came bilinear models (component balances including chemical reactions and heat balances) where no complete graph solution was found. Historically many techniques developed for special kinds of models (DR proper or gross errors treatment) were developed. At present I prefer the **general nonlinear equation-oriented modelling** which can be complemented by inequalities applied to individual variables or their functions■

3.2 Classification of variables, Redundancy and Observability

Redundancy of measured data is the basis of DR. Without redundant data DR is not possible. In general, there are two kinds of redundant data:

- Point redundancy caused by multiple instruments measuring one variable
- Spatial (model) redundancy caused by model equations among measured variables.

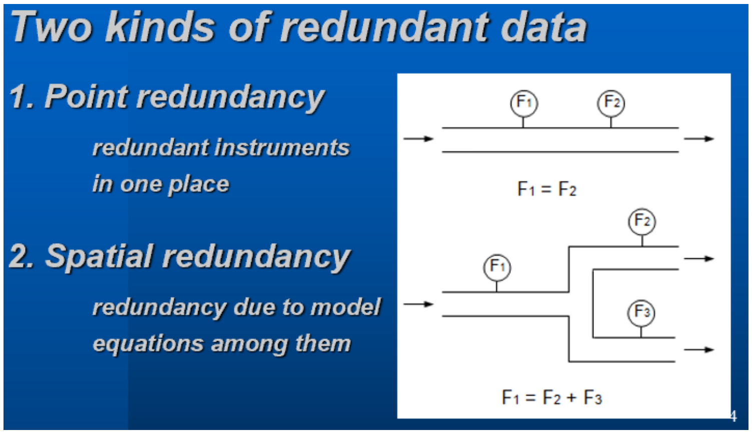


Fig. 3.2: Two kinds of redundancy

In the past, the point redundancy was not common in practice due to expensive instrumentation costs. Nowadays we can meet it quite frequently, especially in critical control applications (supercritical power stations, nuclear power plants).

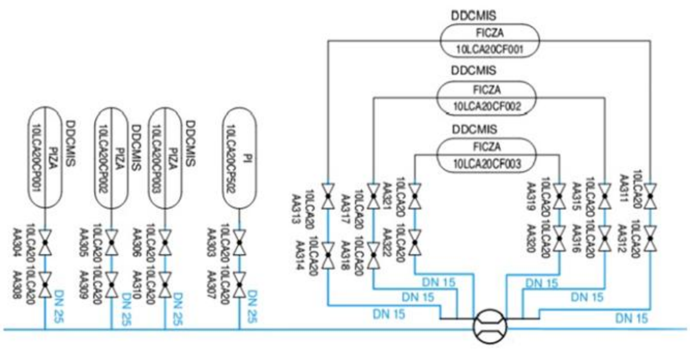


Fig. 3.3: Example of multiple instruments (triple flow measurement and quadruple pressure measurement)

Note 3.5: There can be at least two ways how to cope with multiple instruments [19]. The first method uses all measured data in the main model. This is quite simple. If we have for example three measured temperatures at one place, T1, T2 and T3, we can simply use T1 in the main model and then add two model equations T1 = T2 and T1 = T3. This increases redundancy of the whole model by 2. The second possibility is to evaluate three measured temperatures separately and to create one representative value which can be used in the main model. Both methods have their pros and cons [19], I prefer the second one■

Spatial redundancy will be demonstrated by the following simple example. Simultaneously will be explained also the notion of variables' *observability*.

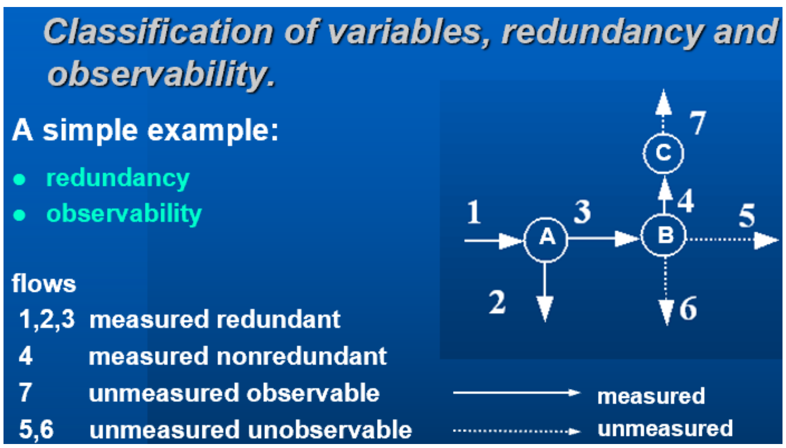


Fig. 3.4: Example of variables' classification

The schema for mass balance in Fig. 3.4 contains 3 nodes (A,B,C) and 7 streams. There holds 3 balance equations among flowrates F*:

(1) $F1 - F2 - F3 = 0$

(2) $F3 - F4 - F5 - F6 = 0$

(3) $F4 - F7 = 0$

The first equation is overdetermined as all flows incident with the node A are measured. Any one stream flow can be calculated from the remaining two. Streams F1, F2 and F3 are therefore *redundant*.

Unmeasured stream 7 can be calculated from the last equation. F7 is so called *observable*. F4 does not contradicts with other measured flowrates. It is called *nonredundant.*

In the second equation there are two unknows. This means that there is no unique solution for F5 and F6. These two streams are called *unobservable.*

The complete classification of variables is shown in the next figure.

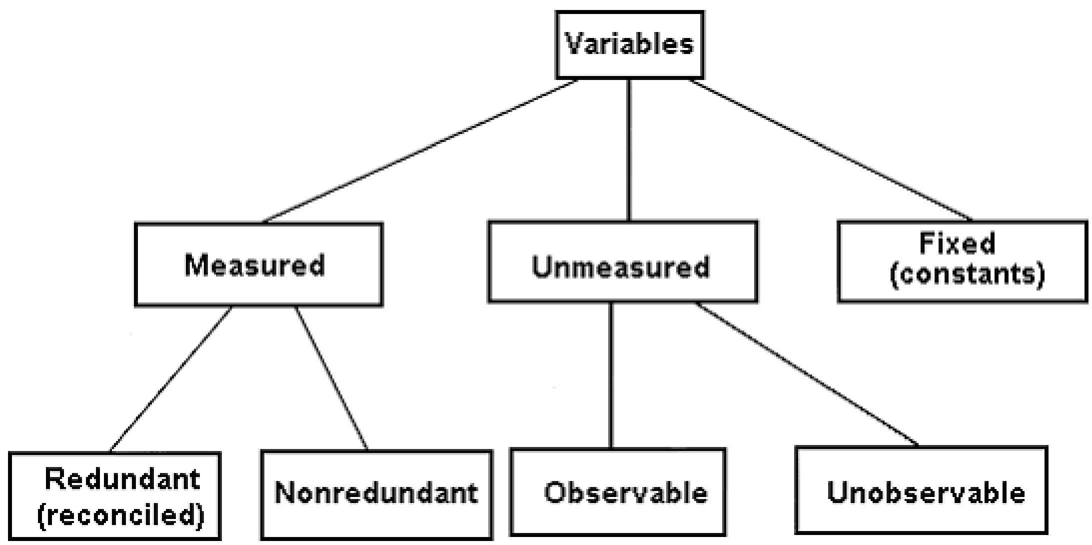


Fig. 3.5: Classification of variables

In addition to the variables measured with some error and the unmeasured ones, we sometimes introduce also a special kind of measured values that are called *errorless*. They are thus a priori known variables, physical constants or variables often obtained by very precise measurement where the error can be neglected. Another kind of such data (disputable) can be custody transfer flows agreed by cooperating parties. If they belong to the redundant ones, their adjustment by the reconciliation is *not* admitted. These variables thus have the character of *constants* during the whole reconciliation process. They are sometimes called *fixed*.

Further properties of the measured variables follow from results of reconciliation. The redundant ones are then adjusted and they are thus also called *adjustable*. The remaining ones are called *nonadjustable*. In one balance scheme, one can meet with a whole spectrum of variables. While in one place (around one node), all measurements are redundant, on the contrary in another place measurements are absent and certain variables are unobservable.

The system with all unmeasured variables observable is called *fully observable*. If some DVR system is used regularly in industrial environment (for example as a part of OLM), it should be fully observable. For typical models, which are nonlinear, the unobservability can cause significant theoretical and also practical problems.

The classification shown in Fig. 3.4 was done by the common sense. **In real industrial DVR tasks with hundreds or thousands of nodes and streams more systematic approaches must be used.**

Note 3.6: The classification of variables brings some practical rules (which hold for linear fully observable linear systems with uncorrelated measurement errors):

- If some unmeasured unobservable variable is measured, it becomes measured nonredundant
- If some measured variable is nonredundant and it becomes unmeasured, it is unobservable
- If some measured variable is nonredundant, it is not reconciled.

For nonlinear systems and systems with correlated measurement errors these rules don’t hold generally. For example, even for linear models, in the case of correlated errors the nonredundant variables can be reconciled. For nonlinear models can happen that if the redundant variable is put among unmeasured ones, it can become unobservable. But such cases are rare in practice■

Note 3.7: Sometimes we can meet the notion *estimability*. Estimability concerns the full observability of the system. The set of estimable variables is the union of the set of directly measured variables and the set of unmeasured observable variables■

Note3.8: The classification presented in Fig. 3.5 is useful but it is a little bit academic concept. In practice it should be complemented by two kinds of information:

1. Uncertainties of calculated unmeasured variables. Quite frequently uncertainties of unmeasured calculated variables are so high that their values are useless (the uncertainty can be sometimes higher than the calculated value itself).

2. Redundancies of some measured variables can be week. This means that their influence on other measured variables is negligible. This can be characterized by their *adjustability* which will be defined later in Section 3.4■

3.3 Measured data reconciliation

Eq. (3-2) holds for the true (unknown) values of variables. If we replace them by the measured values $\boldsymbol{x^+}$, the equations need not (and most likely will not) be exactly satisfied:

$$\boldsymbol{F(x^+,y,c) \neq 0} \tag{3-5}$$

whatever will be the values of the unmeasured variables.

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

$$x_i' = x_i^+ + v_i \quad , \tag{3-6}$$

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. (3-2) – we say that they are consistent with the model

$$\boldsymbol{F(x',y',c) = 0} \tag{3-7}$$

2) The adjustments are minimal. Minimized is the quadratic form

$$\text{minimize} \qquad \boldsymbol{v^T W^{-1} v} \tag{3-8}$$

where $\boldsymbol{v}$ is the vector of adjustments v_i ($v_i = x_i' - x_i^+$) and $\boldsymbol{W^{-1}}$ is the inverse of the covariance matrix of measurement errors. In the case of uncorrelated (statistically independent) errors $\boldsymbol{W}$ is diagonal and the expression (3-8) has the form of weighted sum of squares

$$\text{minimize} \qquad \sum (v_i/\sigma_i)^2 = \sum (x_i' - x_i^+)^2/\sigma_i^2. \tag{3-9}$$

The inverse values of measurement errors variances σ_i^2 – so-called *weights* $1/\sigma_i^2$ - then guarantee that more (statistically) precise values are less corrected than the less precise ones (this is a relevant property of the method). This is the well known Method of Least Squares (or Generalized Least Squares in the case of expression (3-8)).

The reconciliation proper is the 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 for very simple models. The mathematics of the solution itself was in the last decades many times described in the literature (e.g. [2-9]) and will not be mentioned in the sequel.

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

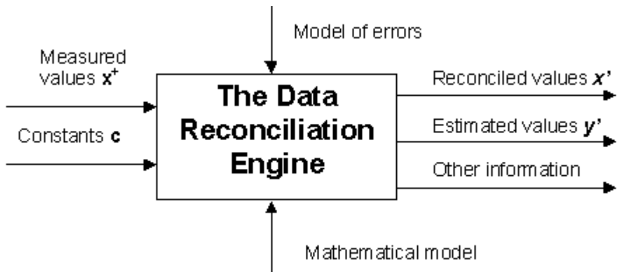


Fig. 3-6: The Data Reconciliation Engine (DRE)

DRE thus transforms the input measured data (vector $\boldsymbol{x}^+$) to the reconciled $\boldsymbol{x}'$. Further in addition, it computes the directly unmeasured variables $\boldsymbol{y}'$ and provides also other information, which will be needed in further sections.

We can write symbolically

$$\boldsymbol{x}' = \boldsymbol{h}_1(\boldsymbol{x}^+) \qquad (3\text{-}10)$$

$$\boldsymbol{y}' = \boldsymbol{h}_2(\boldsymbol{x}^+) \qquad (3\text{-}11)$$

It is important to realize that all results of DVR are transformations of measured values $\boldsymbol{x}^+$. The symbolic functions $\boldsymbol{h}_1(\boldsymbol{x}^+)$ and $\boldsymbol{h}_2(\boldsymbol{x}^+)$ are available via DRE and can be used for calculations of reconciled values and their characteristics.

Note 3.9: The whole DR process (model linearization, elimination of unmeasured variables and DR proper applied to submodel (3-4)) requires efficient software. There exist many methods how to do it [3-9]. There are two main ways how to find the minimum of (3-8). The first is so called Successive Linearization (SL) where the linearized model (3-3) is used in the iterative way until the model (3-7) is zeroed (residuals of equations reach required minimum (values close to zero). This relatively simple and fast algorithm has one drawback – **zeroing the model (3-7) does not guarantee reaching the true minimum of the least squares sum (3-8),** see [5], p.137. This fact is frequently overlooked. The second way is to use some of Nonlinear Programming (NLP) methods, for example Successive Quadratic Programing, which not only zeroes (3-7) but also reaches the real minimum of (3-8). Models in Process Industries are not too much nonlinear and for a routine DVR calculations the SL method is sufficient. **In special situations (e.g. GE identification) finding the exact minimum can be important [18].**
It can be argued that for nonlinear models there can be more than one minimum, but with models used in practice (multicomponent balances, pipeline hydraulics, steam cycles in power stations or classical steam generators) I have never met such case■

Further on will be very briefly described the process of DVR completion. We suppose that the real minimum of the criterion (3-8) or (3-9) was found. The model (3-2) can be linearized by the Taylor Series Method at the solution point to the form of Eq. (3-3), unmeasured variables can be eliminated [3] and we get the model form (3-4). The linearized model is suitable for calculating statistical characteristics of results (covariance matrices of adjustments, reconciled values and calculated unmeasured variables). There are also other important results like parametric sensitivities, info about propagation of errors, etc., which will be described in the next section.

Note 3.10: As was already stated, statistical theory is mostly limited to linear models [1]. There are two main methods that can solve this issue [14]. The first possibility is to linearize nonlinear models by the well known Taylor Series Method (TSM). This solution is quite simple, the only drawback is that this solution provides only approximate results due to approximation of nonlinearities by linear functions. In the case of DVR this can be frequently accepted as the differences between measured and reconciled data are supposed to be small. The second method is MCM – the Monte Carlo Method [18]. MCM is based on repeated calculations which simulate the effect of random errors. In this way can be respected not only nonlinearities of models but also different distributions of random errors. The main disadvantage of MCM is that it is quite time demanding. To get reliable results, thousands of simulations must be done [18]. In daily practice like On-line Monitoring, TSM can be applied. MCM can have its place in detailed studies in the stage of DVR system development■

3.4 Statistical properties of results

This Section describes statistical properties of DVR results (their uncertainty and other substantial characteristics). The mathematics of the solution is suppressed to the minimum.

The Quadratic form of adjustments (3-8) or (3-9) is the random variable with χ^2 distribution with ν degrees of freedom. Values of $\chi^2_{(1-\alpha)}(\nu)$ for probability (1-α) are tabulated in statistical tables.

Between covariance matrices of measurement errors $\boldsymbol{W}$, *adjustments* $\boldsymbol{W_v}$ *and reconciled values* $\boldsymbol{W_{x'}}$ holds the **important** relation

$$\boldsymbol{W} = \boldsymbol{W_v} + \boldsymbol{W_{x'}} \qquad (3\text{-}14)$$

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

$$\sigma_i^2 = \sigma_{vi}^2 + \sigma_{x'i}^2 \qquad (3\text{-}15)$$

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

$$<x'_i - 1.96\ \sigma_{x'i}\ ;\ x'_i + 1.96\ \sigma_{x'i}> \quad (3\text{-}16)$$

cover the (unknown) true values of individual variables.

Reconciled data are more precise in the statistical sense, if compared with the measured ones (this follows from Eq. (3-15)). The enhanced precision of reconciled values 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.

$$\sigma_{x'} < \sigma \quad (3\text{-}17)$$

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

$$a = 1 - \sigma_{x'} / \sigma \quad (3\text{-}18)$$

The **important variable 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. The greater the adjustability is, the greater is also the reduction of the uncertainty.

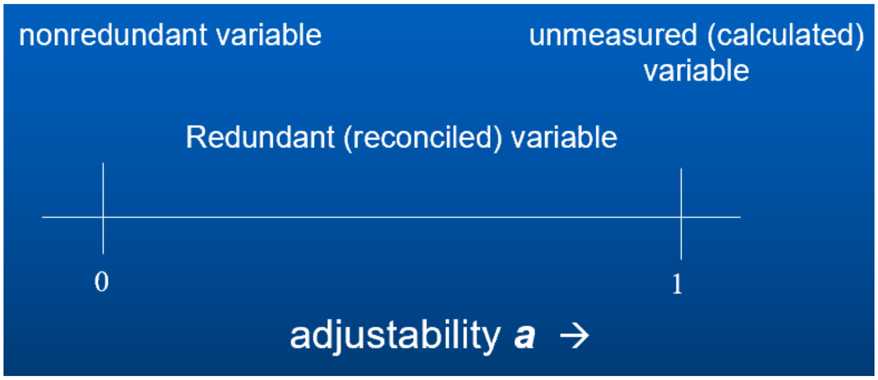


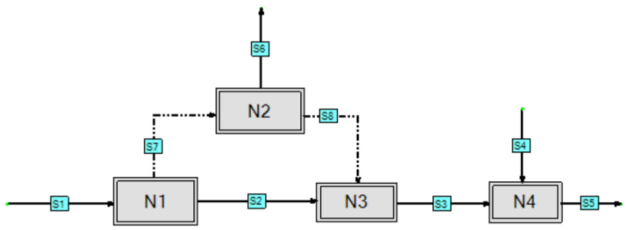
Fig. 3-7: Adjustability

Fig. 3-7 shows graphically relation among nonredundant variables, redundant variables and observable unmeasured variables. Nonredundant variables have adjustability zero, redundant variables have adjustability in the interval (0 ; 1) and an unmeasured observable variable has adjustability 1 (its original value has unlimited uncertainty).

Note 3.11: Adjustabilities will be used in the next Chapter in area of gross errors identification. They play role also in general process modeling. Adjustabilities are the measure of relations among measured variables. **If some redundant variable has adjustability approaching zero, it is not probable that it could be calculated from other variables with some reasonable uncertainty**. As example can serve the measured pressure of water in models based on heat balance at low pressures. If we set such pressure among unmeasured variables, model will calculate probably some nonsense. See also Notes 3.6 and 3.8■

Example 3.2: Mass balance model – classification of variables and Data Reconciliation

The simple model of mass balance is described in the Appendix 2 (see the Appendix 2 for more details about input data and results).



.

Excerpt from results of data reconciliation

```
Task: MASSBALL (Single-component balance)

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

 Name             Type      Inp.value      Rec.value          Uncertainty
 -------------------------------------------------------------------------
 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

Legend:
Type of variables: MC – Measured and Redundant, MN – Measured and Nonredundant. NO Nonmeasured
and Observable.
```

The report about classification of variables follows:

```
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

 ---------------------------------------
  MF  S1                 0,350775  KG/S
  MF  S3                 0,216093  KG/S
  MF  S4                 0,172182  KG/S
  MF  S5                 0,392338  KG/S
  MF  S6                 0,046343  KG/S

 Legend:
  Adjustability   = relative cut of error due to reconciliation
  MF  Mass flow
```

This report informs us that:

- all unmeasured flowrates are observable
- there are 5 redundant measured flowrates
- stream S2 is not among redundant flowrates – it is nonredundant
- report contains also adjustabilities.

For example, for the stream S1 the adjustability equals ca 0.35. This means that after DR the uncertainty is lowered by ca 35 %■

3.5 Parametric sensitivities and propagation of measurement errors

Let's recall equations (3-10) and (3-11):

$$\boldsymbol{x'} = \boldsymbol{h_1}(\boldsymbol{x^+}) \tag{3-10}$$

$$\boldsymbol{y'} = \boldsymbol{h_2}(\boldsymbol{x^+}) \tag{3-11}$$

Eqs. (3-10) and (3-11) can be, for one general variable *z,* approximated by the Taylor Series Method expanded at point $\boldsymbol{x'}$ in the symbolic form

$$z = h(\boldsymbol{x}) \cong h(\boldsymbol{x'}) + \sum \partial h(\boldsymbol{x'})/\partial x_i' \Delta x_i \tag{3-19}$$

The partial derivatives $\partial h(\boldsymbol{x'})/\partial x_i$ are so called *parametric sensitivities* (PS) of reconciled values and calculated unmeasured variables on measured values.

For the variance of variable *z* we have (in the case of uncorrelated measurement errors)

$$\sigma^2_z \cong \sum (\partial h(\boldsymbol{x})/\partial x_i)^2 \sigma_i^2 \tag{3-20}$$

The terms on the right-hand side of Eq. (3-20) are always nonnegative and represent the *contributions* of individual measured variables to the variance of the result. We can now form the vector of the relative contributions (**shares**) $\boldsymbol{s} = (s_1, s_2, \dots, s_I)^T$ by dividing (3-20) by its left-hand side. Shares represent the **percentual share** of individual measured variables on the variance of the result [3]:

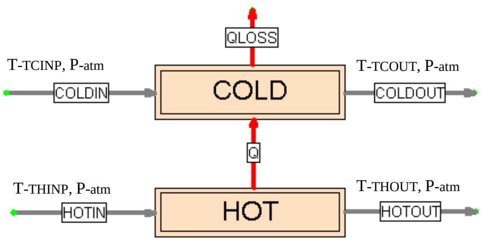
$$s_i = 100\,[(\partial h(\boldsymbol{x})/\partial x_i)^2 \sigma_i^2]/\sigma^2_z \tag{3-21}$$

The sum of shares (vector elements of **s**) is clearly 100 %. The value of the vector indicates, for which of the measurements it makes sense to strive for making them more precise and on the contrary, which of them are irrelevant from the viewpoint of the measurement system optimization.

It remains to note that from the viewpoint of minimizing the result uncertainty, deciding is its standard deviation, which is the square root of the variance. Minimizing the variance thus, indeed, leads to minimizing the result uncertainty; however, the relative importance of individual variables is partially deformed by the nonlinear relation between the standard deviation and variance. The vector of shares itself is thus to be interpreted as the first information for further optimization steps supported by detailed calculations.

Example 3.3: Heat exchanger model – Data Reconciliation, parametric sensitivities and propagation of measurement errors

The simple model of the mass and energy balance is described in the Appendix 3 (see the Appendix 3 for details of input data and complete results).



Excerpt from results of data reconciliation

```
 Task: One heat exchanger

 A U X I L I A R I E S

 Name                Type      Inp.value      Rec.value      Abs.error
 -------------------------------------------------------------------
HTC                  NO          500,000        581,751         25,949  W/m2/K
```

Parametric sensitivities of HTC

From results of DR follows that the calculated value of HTC equals 581.751 W/m2/K with the uncertainty 25.949 W/m2/K. In the next report are parametric sensitivities of HTC to individual measured variables:

```
REPORT ON PARAMETRIC SENSITIVITY
 ================================

 Type Variable    Description
 ---------------------------------------------------
  V   HTC         Heat transfer coefficient [W/m2/K]

 GIVEN VARIABLE IS SENSITIVE TO:
```

```
Type Measured variable    Sensitivity  Unit
-------------------------------------------
 HF  QLOSS                      0,035  [1] / [KJ/S]    heat loss to the environment
 MF  COLDIN                     2,835  [1] / [KG/S]    cold stream in
 MF  HOTIN                     15,217  [1] / [KG/S]    hot stream IN
 P   atm                  -3,27475E-4  [1] / [KPA]     atm. press.
 T   TCINP                      0,701  [1] / [C]       water cold input
 T   TCOUT                     14,274  [1] / [C]       water cold output
 T   THINP                      4,437  [1] / [C]       water hot input
 T   THOUT                    -19,358  [1] / [C]       water hot output
 V   A                         -2,909  [1] / [1]       Heat transfer area [m2]

Legend:

 HF  Heat flow
 MF  Mass flow
 P   Pressure
 T   Temperature
 V   Auxiliaries
```

This report contains parametric sensitivities (PS) of HTC to all measured variables. For example, PS of HTC to mass flowrate of the input cold water COLDIN is 2.835 W/m2/K per 1 kg/s of the water flow. This means that the error in the flow measurement 1 kg/s will cause error in HTC equal to 2.835 W/m2/K■

Propagation of measurement errors in calculation of HTC

In the next report is the vector of shares for HTC:

```
REPORT ABOUT PROPAGATION OF ERRORS
==================================

Type Variable    Description
---------------------------------------------------
 V   HTC         Heat transfer coefficient [W/m2/K]

THE VARIANCE OF GIVEN VARIABLE IS CAUSED MAINLY BY:

Type Measured variable      Share
---------------------------------
 MF  HOTIN                      9 %  hot stream IN
 T   TCOUT                     30 %  water cold output
 T   THINP                      3 %  water hot input
 T   THOUT                     56 %  water hot output

 Sum                           97 %

Legend:
 MF  Mass flow
 T   Temperature

Remark
------
 The sum should be close to 100
```

This report contains shares of 4 measured variables (flow HOTIN and temperatures TCOUT, THINP and THOUT). The variance (σ_i^2) of HTC is mostly affected by variances of these 4 measured variables which together cause 97 % of variance HTC. In the report are presented only variables whose shares are greater than 1 %. The most important is measurement of TCOUT and THOUT. These measurements are the bottleneck of improving HTC measurement precision. These measurements should be carefully maintained and improved (new more precise thermometers, installation of multiple instruments)■

3.6 Modeling – conclusions

1. The most important source of information about measurement errors (their uncertainty) are data from instrumentation vendors, standards of measuring methods and theoretical analysis of the measurement process. These errors have mostly systematic character (bias). See also the introductory part of Chapter 4.
2. The General implicit nonlinear model (3-2) is the basis for solution of DVR tasks. The solution should be based on some nonlinear optimization method as the most common Successive Linearization method can yield biassed results.
3. The linearized model (3-3) can serve as the basis for further studies like statistical properties of results, parametric sensitivities, etc.
4. Data reconciliation is also the basis for classification of variables as redundant, nonredundant, observable and unobservable. **DVR is the only reliable method for analyzing large industrial complexes as concerns sufficiency of their instrumentation systems**.
5. **Very important notion is the *adjustability* of measured variables**. Adjustability plays role in several DVR activities (modeling, GE detection and localization)
6. Parametric sensitivities based on model's linearization are useful for analysis of measurement errors propagation and also for optimization of operating plants.

4 DETECTION, IDENTIFICATION AND ELIMINATION OF GROSS ERRORS

More about errors:

Let's recall the Subsection 3.1.1 where the measurement error was defined by Eq. (3-1). It was supposed that error has the character of a random variable with zero mean and normal distribution. If the measurement is repeated in time, we can meet the second kind of error – so called *systematic error* (*bias*). The error is the sum of the systematic (for example the constant error) and of the random error. Systematic errors are typical for measuring instruments which prevail in the process industries. Random errors prevails only in area of laboratory analyses (analysis proper and manual sampling).

The term *Gross Error* (GE) means the measurement error which is highly improbable as being a random error, for example it is greater than three times the random error σ (the probability of such random error for normally distributed errors is less than 0.003). The cause of a GE can be random (single occurrence) but also systematic, caused for example by malfunction of some measuring instrument.

The process of DR is based on one model where all variables, measured and unmeasured are tied together. This means that one measured value corrupted by some big error can influence resulting values of many other measured and unmeasured variables. This is the well known effect of GEs *smearing*. **The protection of the DVR process against gross errors is therefore essential**.

Besides measurement errors there can exist also errors in the physical model proper (wrong calculation of physical properties, wrong physical model equations, incomplete stoichiometry of chemical reactions, etc.). Localization of such errors is difficult, in practice model errors should be revealed during the model building.

The special and important kind of model errors are so called *leaks*. Detection of leaks from a plant to environment is quite frequent theme in DR papers from academic strata. In practice substantial leaks of materials to environment are easily visible, at least in the case of a bulk chemical processing, oil refining or large power stations. The real problem are so called *internal leaks* between plant subsystems. There exist usually number of bypasses of individual apparatuses needed for their isolation in the case of their damage, typical examples can be heat exchangers. The tightness of individual bypasses during the normal plant operation depends on quality and maintenance of valves. In the case of power generation this issue has the special name – *the steam cycle isolation*. Another example can be the problem of balancing tank farms. There are usually many connections between tanks, especially tanks with the same materials, connections needed for products blending, and the like. All these internal leaks need not be large, all depends on valves' tightness and pressure differences. DVR methods need not be efficient for leaks detection, usually the other methods are more efficient (for example monitoring temperature on pipes which should be normally cold).

Dealing with GEs is usually done in three steps:

1. *Detection* of the presence of (one or more) GEs
2. GEs *identification* which means finding a set of suspect measurements which are causing the issue
3. GEs *elimination*.

Further on will be described the individual steps.

4.1 Introductory screening

This step should precede more advanced techniques based on DVR which will be described later. It is based mainly on the common sense. Every measured variable has its *feasible range* where the measured values can occur. Here are some examples:

- the flowrate can't excess the pump capacity
- temperatures can be limited by phase equilibria (boiling water, etc.)
- the mass of fluid in a tank can be only in the interval <0 ; full>
- pressure has its feasible range
- pressure in the pipe should decrease in the direction of the flow.

In practice the instrument's damage is frequently manifested abruptly, with clearly visible unacceptable value. In the beginning it is therefore useful to give to every measured variable its **feasible range** with regular checking. This is very simple and modest precaution which surely pays off. The introductory screening should precede the more sophisticated methods of GEs search.

4.2 Data analysis on the basis of DVR models

The frequently asked question is:

I have introduced the artificial GE greater than the measurement uncertainty and this GE was not detected by my DVR system. How is it possible?

In the beginning it should be stated that many DVR users believe that it is possible to find all gross measurement errors present in the data set. As will be seen in the next, this is not possible. This is the theme of the present Section. In general, GEs belong to two groups:

1. Gross errors in redundant measured values which contradicts with other measured values. Such gross errors can be detected, but only with some probability
2. Gross errors of measured variables which are not redundant and can't be detected during DRV process at all.

There is also the issue of directly unmeasured but calculated variables (process variables, model parameters and KPIs). Results of these variables can be devalued by gross errors in measured values, redundant and nonredundant. In what follows will be answered the following questions:

1. What is the probability to detect a gross error of some size (GE detectability)
2. How will GEs influence values of main results (targets of the overall measurement), like KPIs, Heat Rates, etc.
3. How to design a system protecting main results against GEs.

The most frequently used method for Gross Errors Detection (GED) is the test based on the value the least square function (3-8) or (3-9). The Quadratic form of adjustments (3-8) or (3-9) is the random variable with χ^2 (ν) distribution with ν degrees of freedom.

Values of χ^2(ν) 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},

$$Q_{min} = \boldsymbol{v}^T \boldsymbol{W}^{-1} \boldsymbol{v} \quad , \qquad (4\text{-}1)$$

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.

$$Q_{min} < \chi^2_{(1-\alpha)}(\nu) \qquad (4\text{-}2)$$

$\chi^2_{(1-\alpha)}(\nu)$ is called the *critical value* of Q_{min} (Q_{crit}). Number of degrees of freedom ν is in DR called **Degree of Redundancy (DoR).** In most cases for DoR in the case of the fully observable system 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 text). All this holds on assumptions that only random errors with the Normal distribution are present.

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

$$S = Q_{min} / \chi^2_{(1-\alpha)}(\nu) \qquad (4\text{-}3)$$

Then the inequality (4-2) reads

$$S < 1 \qquad (4\text{-}4)$$

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

The S definition has the advantage for an end DVR 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.

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.

Note 4.1 The GE detection based on inequalities (4-2) or (4-4) is called the Global Test. Besides the Global test there also exist GE testing based on so called *maximum normalized adjustments* (MNA) described in [2]. This test has several drawbacks and will not be described in what follows. Frequently it is not understood well. While all normalized adjustments have distribution N(0,1), MNA distribution is different (for details see [2]. p.422 – 423). This assertion looks strange but it is true. MNA test main disadvantages are:

- the test is only approximate. Its exact power is not known. See also [2], p. 423.
- while the chi-square test (4-2) is numerically robust, the MNA test is prone to numerical problems (ratio of two very small numbers in the case of almost nonadjustable variables)

On the other side, the normalized adjustments are useful in the stage of GE identification■

Note 4.2: Important can be the average value of Q_{min} in daily operation. The mean value of χ^2 distribution equals ν. The average value of Q_{min} **significantly** below ν means that the uncertainties of measurements are overestimated and should be revised. The same holds also in the opposite direction. In general, one or several values of S slightly above 1 from time to time should not be reason for an alarm■

4.3 Gross errors detectability

It is now time to ask the question:

How powerful the GE detection is?

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

Let's recall Eq. (3-1) defining a random error and let's modify it to the form

$$x^{+} = x + e + d \quad , \qquad (4\text{-}5)$$

where d is a gross/systematic error (which is a constant).

One has to begin with testing the GE presence hypothesis [3] (you can find more about hypotheses testing also in the Appendix 1 of the present report). **The hypothesis H_0 is:**

***There is no GE present* (this means d = 0)**

If the inequality (4-2) or (4-4) is NOT fulfilled, the hypothesis H_0 is **rejected**. As every statistical test, also the χ^2 test has its power characteristic:

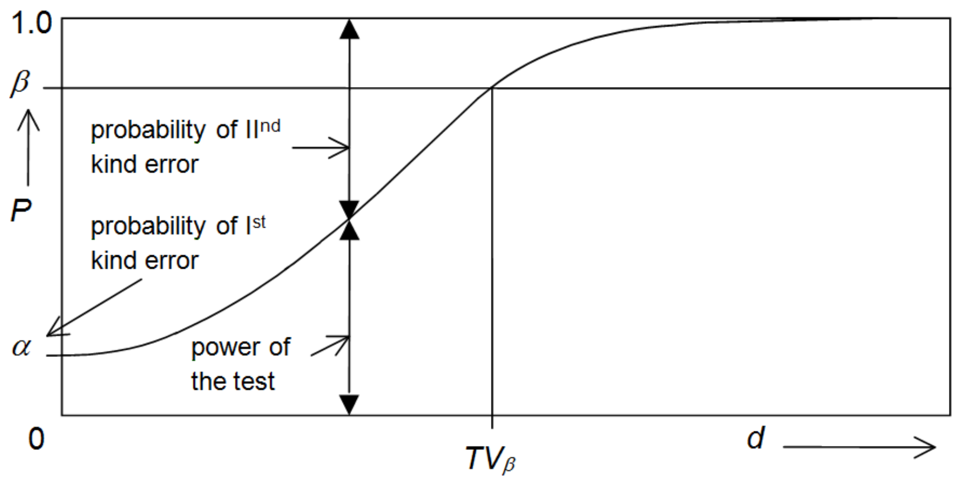


Fig. 4.1: The *power characteristic* of the χ^2 test

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 a redundant 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 ($d\rightarrow\infty$). The value (1 – power of the test) is called *probability of **the IInd kind error*** (gross error is present but it is not detected).

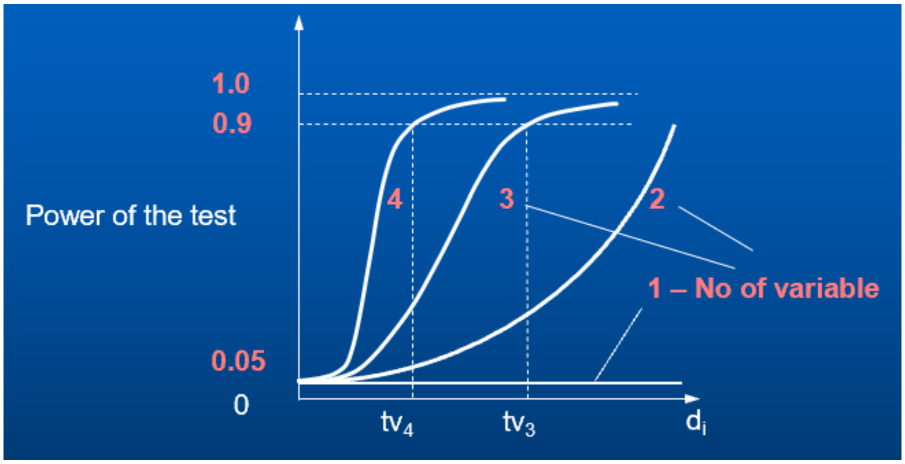


Fig. 4.2: Typical *power characteristics* of the χ^2 test

In this figure are some typical power characteristics of the GED test. The best situation is in the case of variable No. 4. The ability of the test is decreased for variables No. 3 and 2. Variable No. 1 (the flat line) is the nonredundant variable. Its GE will not be detected at all, even if it is very large.

The power characteristic represents though complete, but still too complicated information for the application in practice (imagine hundreds of such curves in a real size problem). Simpler is the characterization 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 a gross error that will be detected with probability β (we'll further assume $\beta = 0.9$). TV_β is the characteristic value for any measured redundant variable. The smaller TV_β , the better. TV_β is called the *Threshold Value*.

The threshold value can be computed from Equations (4-7) and (4-6)

$$q_i = \delta_\beta(\nu,\alpha)/[a_i(2-a_i)]^{1/2} \qquad (4\text{-}6)$$

where q_i is dimensionless threshold value TV_i/σ_i

$$q_i = TV_i/\sigma_i \qquad \text{or} \qquad TV_i = q_i\,\sigma_i \qquad (4\text{-}7)$$

and $\delta_\beta(\nu,\alpha)$ is the statistical 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 [3], p. 179 or [15]. See also Appendix 1.

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 [15].

Let us notice that for a measured variable, the threshold value is a simple function of its *adjustability* defined by Eq. (3-18); see also the following figure.

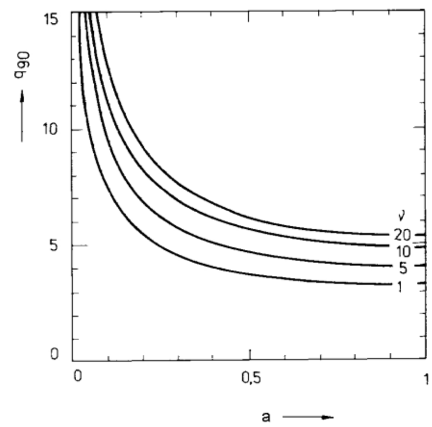


Fig. 4.3: Dimensionless threshold value *q* ($q = TV/\sigma$) as function of the degree of redundancy ν and adjustability *a* (for α=0.05 and β=0.9)

Note 4.3: I must admit that reading this Section is not easy. But, calculating threshold values is quite simple:

1. Find the dimensionless TV *q* from Fig. 4.3 or calculate it from Eq (4-6)
2. Calculate TV from Eq. (4-7)

That's all (of course, you must know the adjustability *a*). There are two other methods for calculating *q*, they are described in the Appendix 1■

From Fig. 4.3 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 TV)
- for adjustability smaller than 0.05, 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 uncertainty is taken as 1.96 times the standard deviation, the minimum threshold value results as 1.65 times the uncertainty. 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.
- increasing DoR (ν) does not mean automatically that smaller GEs will be detected.

Some DVR authors does not acknowledge unmeasured process variables (flowrates, temperatures, etc.), for example the German VDI standard [12]. Instead of this such variables are supposed to be "pseudomeasured" which means that some "engineering estimates" with large uncertainties are used as measured values. Such solution increase DoR significantly with the adverse effect on GE detectability (see Fig. 4.3) and can't be recommended.

4.4 Gross errors identification

The number of measured variables in typical real models is usually considerable and can be in the order of hundreds or thousands. Physical screening of all instruments is thus, in most cases, not realistic. Fortunately, there exist relatively effective methods of looking for measurements charged with gross errors. Although one usually does not succeed in finding directly only one source of gross error, it is sufficient to find a small group of suspected variables upon which one can concentrate the attention.

4.4.1 Estimation of GE magnitude

First of all, it is useful to estimate roughly the size of the GE. This estimation is based on the size of Q_{min} . It is clear that Q_{min} depends on the GE size (we further suppose that there is only one GE). The GE size d_i for the i-th variable equals approximately according to [3] p. 190:

$$d_i = (Q_{min} - \nu)^{1/2}\ \sigma_i^2/\sigma_{vi} \qquad (4\text{-}8)$$

The second method for estimating GE size is based on putting the variable among unmeasured ones and running DR. The difference between the calculated value and the original measured value is the estimated GE magnitude.

4.4.2 Analyzing nodes' imbalances

In the case of linear (mass balance) models the simplest method is to **examine the balance differences (imbalances)** around individual nodes. The balance difference Δ for mass or energy balance is defined as

$$\Delta \;=\; \text{sum of inputs} - \text{sum of outputs}. \qquad (4\text{-}9)$$

If the absolute value of Δ exceeds the threshold value, it is advisable to examine the values connected with this node. The threshold value depends on the uncertainties of individual measurements. If for example in two neighboring nodes one shows a surplus of mass and the second the shortage of mass, the incident stream between these nodes is suspect.

The weak point of this method consists in the frequent occurrence of unmeasured streams among inputs and outputs. Better solution is to analyze imbalances around so-called *macronodes* which has all input and output streams measured (this method is called the *reduction of the balance scheme*). The nodal method is sometimes recommendable for its simplicity and easy comprehensibility. It makes possible to find not only gross measurement errors, but also mistakes in the phase of the model creation and also to detect possible leaks.

This method is suitable mostly for simple mass balances which are typical for Yield Accounting. For more complex nonlinear models like multicomponent balancing or energy balancing this method is not suitable.

Example 4.1: Detection and identification of a gross error in the mass balance task

Recall the mass balance task presented in Appendix 2. We have introduced the gross error +10 kg/s to the stream S1 flowrate (value 100.1 kg/s was increased to 110.1 kg/s). There is the macronode consisting of nodes N1, N2 and N3 which comprises all unmeasured streams. Here is the report:

```
Task: MASBALLS1GE (Single-component balance)

 ERRORS / WARNINGS
 ================

 S U S P E C T   M A S S   I M B A L A N C E S

 MACRONODE:
  [ N1, N2, N3 ]

 INPUTS:
 Stream    From node To node       Value        Error
 --------------------------------------------------------------------
 S1        ENVIRON   N1            110.1        2.202 KG/S
                     ----------------
           Sum of inputs:          110.1

 OUTPUTS:
 Stream    From node  To node      Value        Error
 --------------------------------------------------------------------
 S6        N2        ENVIRON       19.8         0.792 KG/S
 S3        N3        N4            79           1.58 KG/S
                     ----------------
        Sum of outputs:            98.8

        Imbalance:                 11.3 (10.8%)
        Test (should be < 1.96):   7.844
```

■

4.4.3 Normalized adjustments

Considerably more general and sophisticated is the method of **normalized adjustments**. It starts from the idea that adjustments are in fact estimates of errors. Any adjustment v_i is a random variable, which has its standard deviation σ_{Vi}. The *normalized adjustment* u_i is defined by the relation

$$u_i = v_i / \sigma_{vi} \qquad (4\text{-}10)$$

The normalized adjustment has the normalized (standard) normal distribution. It is well known that if the variable is corrupted by a gross error, its normalized adjustment belongs to the largest in absolute value. It is thus sufficient to compute the normalized adjustments for all adjustable variables and to range them according to the increasing absolute value. At the end of the sequence, there is then a group with highest values of the normalized adjustments and thus a group with most suspected variables.

Having the group of suspect variables, we can go further. A suspect variable can be scrutinized by putting it among the unmeasured ones and carry out the data reconciliation again. If then no gross error is detected, this variable could be the source of the gross error. By putting the variable charged with a gross error among the unmeasured ones, we in fact have carried out the *gross error elimination*.

If it happens that the elimination of any one of the suspected variables does not suffice (a gross error is still detected), it is possible that more gross errors are present. In the course of the successive elimination we then trace the decrease of the variable Q_{min}. Suspected are those variables where the decrease is largest. See more in the Subsection 4.4.5

The method of normalized adjustments with the elimination of variables is quite effective, although not universal. It is suitable for gross errors of measurement, not for model errors. One feature of this method is that frequently happens that two or more variables have the same (or very near) absolute values of normalized adjustments. Some measured variables can have the same (or almost the same) influence on the Least squares function. This is called the **GE Equivalency** [8]. This is the main reason why frequently it is not possible to decide definitely which variable is corrupted by GE and further information independent of DVR is needed.

Example 4.2: Detection and identification of a gross error in the mass balance task

This example is the continuation of the Example 3.2. The gross error +10 kg/s was added to the original value of the stream S1. After execution DR the following values of the Status and related variables were found:

Qmin	6,4542E+01
Qcrit	5,9900E+00
Status (Qmin/Qcrit)	1,077E+01

As the Status is > 1, GE was detected.
In the following table are Adjustabilities and Threshold values:

Type	Variable	Adjustability	Threshold value Beta: 90%	Beta: 95%	Beta: 99%	Unit
MF	S1	0,350775	4,781	5,282	6,217	KG/S
MF	S3	0,216093	4,622	5,106	6,011	KG/S
MF	S4	0,172182	9,908	10,945	12,883	KG/S
MF	S5	0,392338	9,908	10,945	12,883	KG/S
MF	S6	0,046343	4,781	5,282	6,217	KG/S

We can see here that GE in S1 (+10 kg/s) is much greater than the Threshold value 4.781 kg/s (for Beta = 90 %). The GE detection is thus understandable.
For GE identification we will start with the method of normalized adjustments. The report follows:

REPORT ON GROSS ERRORS
======================

S U S P E C T   M E A S U R E M E N T S

Type	Variable	Norm.adjust	GE(abs)	Meter	Meas.	Calc.	Diff.	Unit	Description
MF	S6	8,020	11,3		19,800	20,721	0,921	KG/S	
MF	S1	-8,020	11,3		110,100	102,981	-7,119	KG/S	
MF	S3	6,811	10,7		79,000	82,260	3,260	KG/S	

One can see that three suspected streams of the ordered sequence have remained. The greatest absolute value belongs to streams S1 and S6, however stream S3 remains only slightly back. Let's discuss this table in details:

In the column *Norm.ajust*. are the normalized adjustments. Absolute values of two of them are exactly the same. This means that streams S1 and S6 has the same influence on the Least Squares function. This is the case of the GE equivalence discussed earlier. Such variables can't be distinguished by DVR methods.

The column *GE (abs)* contains the estimates of GE size calculated according to Eq. (4-8). The calculated value 11.3 is not far from the real GE value 10.

Remaining columns contain measured and reconciled values.

Let us further continue according to the method of **suspected variables elimination**. Successively, individual streams are put among unmeasured and reconciliation is carried out. Note that after the elimination DoR is lowered by 1. Results are given in the following table.

Results of elimination

Type	Variable	Meas.	Calc.	Diff.	Qmin	Status
MF	S6	19.8	31.2	-11.4	2.120E-01	5.520E-02
MF	S1	110	98.7	11.4	2.120E-01	5.520E-02
MF	S3	79.0	88.2	-9.25	1.815E+01	4.726E+00

This table contains the following information:

Columns *Meas.* and *Calc.* contain measured and reconciled values after the elimination of individual variables.

Column *Diff.* contains differences *Calc* – *Meas.* They can be interpreted also as estimates of GE.

The last two columns contain the Least Squares function Q_{min} and the Status of data quality.

It can be seen that the stream S3 can be eliminated from the list of suspects (after the elimination the Status is still > 1), but by the DVR method it is not possible to arbitrate between S1 and S6. Note also that while the difference between normalized adjustments is nor great, changes of the Status after the elimination of individual variables are significant.

There thus remain two suspected streams S1 and S6. The possible steps in this case will be discussed in the Subsection 4.4.5■

Note 4.4: It can be useful (from time to time) to look at normalized adjustments. Absolute values above 1.96 can be found frequently and this does not mean automatically the presence of some GE. Anyway, **systematically large normalized** adjustments of some variables signals that their measurement is biased or their uncertainty entering DVR is underestimated. Some DVR authors think that this information can be used for some compensation (correction) of measured values as adjustments are estimates of real measurement errors■

Note 4.5: The main problem connected with the method of normalized adjustments (NA) is the robustness of calculating NA for variables with low adjustability (approaching zero). In such cases, NA is the ratio of two very small numbers and numerical problems can occur. Luckily, this issue is not serious as will be discussed in the Subsection 4.4.5■

4.4.4 Decrease of the least squares function

This method is very similar to the previous one. After putting the suspect measurement among unmeasured variables, there should be the decrease of the least squares function Q_{min}. The variable with the largest decrease is the suspect.

For linear models the methods described in sub-subsections 4.4.3 and 4.4.4 should give the same result. For nonlinear models results can slightly differ. As the method described in this sub-subsection requires more activities, in practice is not used. Anyway, it can be the basis of simultaneous GEs described below.

4.4.5 The "common sense" methods

In Example 4.1 were two final candidates for the gross error – streams S1 and S6. The "normalized adjustment" method of localization fails as their normalized adjustments are equal. What remains is to apply for example:

- physical revision (calibration) of the individual flowmeters on the site
- tracing the trends of variables at the time when the problem has arisen.

In this Subsection further three "common sense" methods will be mentioned.

Measurement credibility

One possibility is to use the concept of *measurement* (*instrument*) *credibility.* Some methods of measurement are simpler and more reliable (or foolproof) than others. For example, the measurement based on levels in tanks can be sometimes more reliable than classical flowmeters. Some other possibilities were already discussed in the Section 4.1.

Using Threshold Values

The reasoning in this **simple but powerful method** is: Some variables have so high TV that detecting their GE is not probable. In other words, they can be corrupted by some GE but such GE will be never detected.

This method can be applied in advance, before the data evaluation step is done. The measured variables with very low adjustabilities (for example 0.01 or less) can be therefore ignored, even if their normalized adjustments are high.

Estimating GE magnitude

This is another simple but powerful method. See the Subsection 4.4.1. It is possible to estimate the magnitude of GE which could cause the Q_{min} or the Status values. Such GE may look unrealistic. This variable can be then deleted from the list of suspects.

Further on is described in details the method based on GE Threshold Values. We will start with the following Example:

Example 4.3: Adjustabilities and Threshold Values in the Heat exchanger model

Let's recall the Heat exchanger model described in the Appendix 3. The classification report follows:

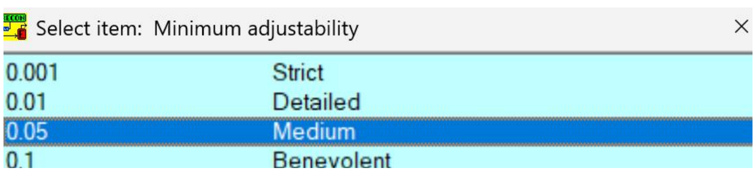
```
REPORT ON CLASSIFICATION OF VARIABLES
 =====================================

 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%
 -------------------------------------------------------------------------
  HF  QLOSS            0,000913      636,695      708,059      841,980  KJ/S
  MF  COLDIN           0,022799        7,766        8,637       10,270  KG/S
  MF  HOTIN            0,023460        3,829        4,258        5,063  KG/S
  P   atm              0,000000  1138639,037  1266262,752  1505761,128  KPA
  T   TCINP            0,201993        2,736        3,043        3,618  C
  T   TCOUT            0,201294        2,740        3,047        3,624  C
  T   THINP            0,046734        5,458        6,069        7,217  C
  T   THOUT            0,046160        5,491        6,106        7,261  C

 Legend:
  Adjustability   = relative cut of error due to reconciliation
  Threshold value = gross error that will be detected with probability Beta■
```

We can see that as concerns adjustabilities, the special position has the water pressure and also the heat loss stream QLOSS. Their adjustabilities are very low and threshold values are very high. TV Beta 90 % value for QLOSS is about 10 times greater than its value. TV for measured pressure is ca 1138 MPa (we can hardly meet such pressure in industrial processes). It is clear that their GEs could not trigger the GE alarm even if their measured values were not OK.

The explanation? The adjustability depends on the influence of some variable on other measured variables in the model. This can be the case of very accurate measurements. The stream QLOSS looks to be not very precise (uncertainty is 30 % of the estimated value 55 kW, this means 16.5 kW). But, the heat flow of the exchanger is about 4614 kW. So, the QLOSS is quite negligible in the overall heat balance.

In the second case, the influence of pressure on water specific enthalpy is very low (at low pressures). In other words, even a significant measurement error of pressure can't cause detection of GE. **This shows the very important role of adjustabilities and Threshold Values in GE localization.**

In practice this rule can be incorporated directly in the DVR software. In the search for the GE source, the user can try several minimum values of adjustabilities and thus limit the number of suspects. Below is the selection box in RECON [16].



4.4.6 The case of more gross errors

The theme of multiple gross errors is quite frequent in published papers about DVR. Researchers from academic strata see it as the significant challenge and important prerequisite of DVR application in industrial practice.

We can meet multiple gross errors mainly in the phase of model building (both measurement and model errors). In this case they can be identified and eliminated by a step-by- step building of the model with continuous checking and solving possible problems. Once an OLM system is properly tuned, model errors should not be present and possible malfunction of many instruments simultaneously in practice is quite rare. So, the occurrence of multiple measurement GEs in daily running is not probable but is rarely possible.

The methods of multiple GEs identification are based on two ideas:

1. *Serial elimination of GEs.* This idea is based on selecting measured variables with the highest absolute value of normalized adjustment and putting this variable among unmeasured ones. If some GE is still detected, this step is repeated until no GE is detected.

2. *Elimination of groups of measured variables.* In the first step are selected groups of two suspect variables and the individual couples are put as unmeasured and GE detection is evaluated. The group with lowest value of Q_{min} is the candidate for putting the couple of variables among unmeasured. If GE is still detected, the same procedure can be repeated with groups of 3 variables.

The second method looks very sophisticated but the number of possible selections of variables can be prohibitive. It is clear that the first method is significantly simpler than the second one.

4.4.7 Systematic measurement errors

If we have one set of data collected during one balancing time interval, the measurement data are corrupted by one set of overall errors. These errors are sums of systematic and random measurement errors. If you repeat balancing in time, we can observe the influence of constant systematic errors. In the next figure is typical graph of data obtained in practice. It is well known that small systematic errors of instruments prevail and they cause systematic differences between measured and reconciled data. Small fluctuations of variables are caused mainly by improper control of the process (probably mainly due to the unsteady state).

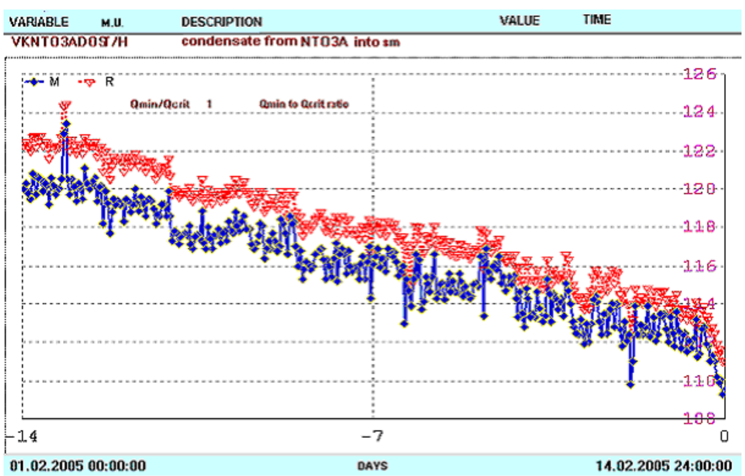


Fig. 4.4: Measured (red triangles) and reconciled (blue rectangles) values of one variable.

The average difference between reconciled and measured value is the estimate of the systematic error. If we observe such behavior for a long time, we can have a temptation to make software compensation of the measured value to be in tune with the model. Such solution is possible but at the first we should analyze the measurement process itself and to check whether a wrong compensation could not be the cause of the issue (to reveal possible defects in compensations).

4.5 Gross errors elimination

If a GE was detected and at least partially identified, the natural question is:

What to do?

The DVR methods help to find one or more candidates of GE. In the case that the decision is clear, the solution looks simple:

- to put the measured variable among unmeasured
- find the reason of the problem (for example repair the instrument)
- test the case after the repair
- put the variable back among the measured ones.

This process is generally not easy. It was shown earlier that mostly only a set of suspect sources of GEs can be found automatically due to GEs equivalence. The decision can be supported by several methods discussed in previous subsections. Anyway, as was shown above, the final decision requires man's judgement. Our opinion is that systems of automatic elimination of GEs offered by some DVR software providers can cause more harm than good. Perhaps AI will help in the future.

The easy inclusion of man's decision into the issue of gross errors requires a good human – machine interface. This problem will be discussed in Chapter 6.

4.6 Protection of Key results against GEs

I admit that this section is not easy to read. If you like, you can skip it and go to the Section 4.7. Anyway, the tireless reader will be rewarded by new DVR insights at the end of this section.

Quite often, the extensive industrial monitoring system with hundreds of measured variables is operated for obtaining few KPIs (overall plant efficiency, Heat Rate, Thermal power of a nuclear reactor, and the like). We speak about *key* (*target*) variables. Let's ask the question:

How to protect Key results (targets) of the whole measurement against GEs?

This task was for the first time solved in [3] (p.181) via the Monte Carlo Method. The solution below is based on using parametric sensitivities and threshold values and is based on the paper [15]. The basic idea is simple:

A gross measurement error should be detected earlier than it causes a significant error of the Key variable(s)"

Key variables and their protection against gross errors

There can be two cases:

A: We are successful, if "A gross error is present and eliminated while maintaining an accurate value for the target variable."

B: We are unsuccessful, if "A gross error is present but not identified and an inaccurate value for the target variable is calculated."

In analogy with statistics (power of statistical tests) we can define the probability of an event **A** as the Power of the Monitoring System Self-Protection (MSSP).

Let's further suppose that for a target variable h, we require the *maximum acceptable error* (uncertainty) e_{hmax}. This total uncertainty can be consumed by:

1. A random error e_{hr} of h caused by random errors of all measured variables (further we suppose Gaussian errors with Normal distribution). As the random errors are not known, we will substitute e_{hr} by e_{hrmax} which represents the *uncertainty* of h caused by random errors. This information is provided by the DR Engine.
2. A constant gross error e_{hg} caused by a gross error of one measured variable d in the sense of Eq. (4-5)

We require that

$$e_{hmax} > e_{hrmax} + e_{hg} \quad . \tag{4-11}$$

Inequality (4-11) sets the upper limit on the error e_{hg} caused by the gross error, further denoted as e_{hgmax}

$$e_{hgmax} = e_{hmax} - e_{hrmax} \tag{4-12}$$

This means that both errors' uncertainties add to form the overall uncertainty. The situation is illustrated in the next Fig. 4.5.

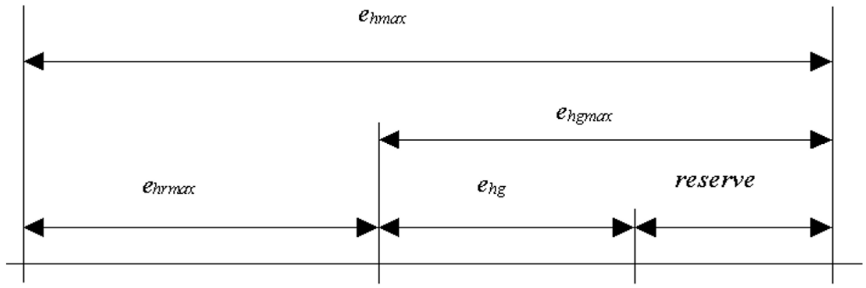


Fig. 4.5. The overall uncertainty e_{hmax} consumed by random and systematic errors

It is clear that the *reserve* should be non-negative to satisfy our MSSP request (4-11). The MSSP analysis will be based on a combination of two methods:

- gross error detection power described in Section 4.3
- the parametric sensitivity of the target variable with respect to the individual measured variables.

Let's suppose that the target variable *h* is a function of measured variables in the sense of Eq. (3-11)

$$h = h(\boldsymbol{x}^+) \tag{4-13}$$

A *parametric sensitivity* ζ_i of *h()* with respect to a measured variable x_i is defined as the partial derivative

$$\zeta_i = \partial h(\boldsymbol{x}^+)/\partial x_i^+ \tag{4-14}$$

The process consists of two steps, which are applied to all measured adjustable variables:

1. Determination of the threshold value for the *i*-th measured variable
2. Evaluation of the parametric sensitivity of the target variable with respect to the *i*-th measured variable.

The process is illustrated in the next Fig. 4.6, which is a modification of Fig. 4-1. On the right hand side *y* axis are errors of the target variable caused by a gross error of the *i*-th adjustable measured variable.

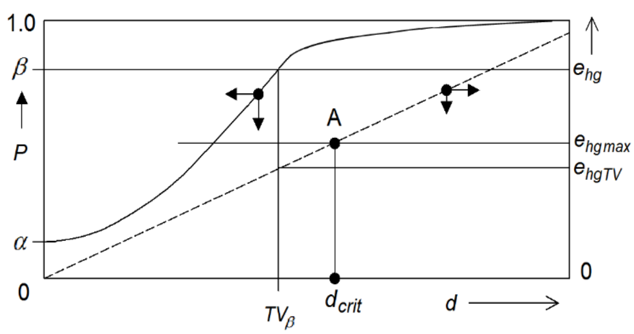


Fig 4.6: Power characteristics (full curve) and the parametric sensitivity (dashed straight line) for the *i*-th measured variable (the index i is omitted here for brevity)

It is supposed that the function (4-13) can be linearised and that a gross error of the *i*-th measured variable transforms to the error of the target variable according to Eq. (4-14)

$$e_{hg} = \zeta_i d_i \tag{4-15}$$

This equation is represented by the dashed straight line in Fig. 4.6. There are two important points on the x axis:

1. The threshold value TV_β which informs that gross error was detected (with probability β)
2. The critical value of the gross error d_{crit} . At this point e_{hg} reaches the maximum value e_{hgmax} and exhausts all uncertainty available (point **A** in the Fig. 4.6).

$$e_{hgmax} = |\zeta_i| \, d_{crit,i} \tag{4-16}$$

or

$$d_{crit,i} = e_{hgmax} / |\zeta_i| \tag{4-17}$$

Now, it is the time to compare the power characteristic curve with the parametric sensitivity straight line. The most important is the relation between $d_{crit,i}$ and $TV_{\beta,i}$. If there holds the inequality

$$d_{crit,i} > TV_{\beta,i} \quad , \tag{4-18}$$

the gross error will be detected before causing unacceptable error in the target variable and the system is well protected against a gross error of the respective measured variable (this case is depicted in Fig. 4.6). In the opposite case an undetected gross error can devalue the target value significantly before it is detected. The inequality (4-18) can be expressed also in the alternative way by substitution of $d_{crit,i}$ from (4-17) to (4-18):

$$e_{hgmax} > |\zeta_i| \, TV_{\beta,i} \tag{4-19}$$

saying that

The product of the parametric sensitivity and the threshold value should be less than the uncertainty belonging to the maximum gross error set a priori for the target variable.

The inequality (4-19) thus represents the sole criterion for assessing whether the target variable is self-protected by DVR (and the following data analysis steps) against gross error) in the *i*-th measured variable. The inequality (4-19) must be checked for all measured variables. For measured variables don’t fulfilling (4-19) the other methods of protection must be used.

This method will be illustrated by the example of monitoring nuclear reactor thermal power in Chapter 7.

4.7 Gross Errors - conclusions

It is now possible to summarize basic findings from this Chapter:

1. The introductory screening described in the Section 4.1 is the "Low cost", but very efficient solution.
2. The chi-square test for GE detection in 4.2 can be recommended. The alternative "Maximum normalized adjustment test" is not so advantageous.
3. Statistical DVR methods for GE detection described in 4.2 are not omnipotent. The understanding of the power of methods used is essential (Section 4.3). You should reconcile yourself with the fact that not all GE will be detected at all.
4. Several GE identification methods described in 4.4 are valuable and can help but the GE selectivity is frequently not sufficient to find one cause of the GE issue.
5. Important is the question: How well are protected key results of DVR against GE? This issue was solved in Section 4.6.

Let's finally state that although the methods described above represent a valuable aid in searching for gross errors, they alone quite often do not lead to finding the unique and true sources of gross errors. One says that these methods are not sufficiently ***selective*** in the GE identification. **It is thus indispensable to complete these methods by verification of suspected meters directly in site. It is also necessary to utilize practical knowledge about the measured system**. For these reasons it also cannot be recommended to apply the methods of automatic elimination of gross errors without the intervention of man (methods sometimes offered by some vendors of DVR software).

5 STEPS BEYOND DVR

The benefits of DVR are quite clear:

1. Consistent data (data which are in agreement with laws of nature) are obtained.
2. Reconciled validated data are more accurate than the original measured data (have smaller uncertainty).
3. DVR provides information needed for detection, localization and elimination of possible Gross Measurement Errors, instrumentation malfunction, etc.
4. DVR provides information about uncertainty of reconciled values and also about uncertainty of directly unmeasured calculated variables and model parameters (rates of chemical reactions, heat transfer coefficients, turbine segments efficiencies, KPIs, etc.).
5. DVR provides information about propagation of measurement errors in the chain of further data processing. In this way it can help to optimize the whole measurement process. Typical tasks here are the analysis of replacement of existing instruments by more accurate and precise ones, the optimization of instrument placement, etc.

To summarize, validated and reconciled data provide better information about the plant performance.

The natural question is:

"How to recast this knowledge (gained after hard work and some money spent) into material benefits (improved yields, improved heat rate, electricity production, higher profit, etc.)?"

Let's discuss some possibilities of better use of process plants data gained with the aid of DVR.

5.1 Process data driven simulation

DVR models can be classed among hybrid models. The core of the mathematical model contains equations based on the First Principles and thermodynamic calculations like phase equilibria, etc. This part of the model is usually created in the Graphical User Interface of some software. Such model can be complemented by user defined equations describing some special features of the plant. Important parts of the overall model can be empirical knowledge gained from equipment vendors or by the analysis of the long-term historical process data.

The time needed for DVR evaluation can be in the order of seconds or minutes. After the DVR step is completed, there are available values of model parameters (heat transfer coefficients, turbine segments efficiencies, etc.). Now, it is the time to use the model in the simulation mode. Model parameters are now inputs for calculations and outputs are values of process variables. In other words, we have created so called Digital Twin of the process. In this way it is possible to answer What if? Queries, for example:

- what will happen if the ambient temperature will rise by 5 K?
- what will happen if the cooling water flowrate will increase by 10 %?

All this can be available for operators for their decision support.

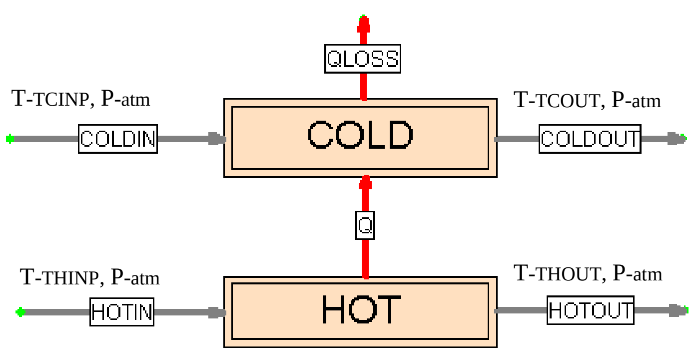
In other words, the simulation mode of calculation is based on changing data flow. The input of DVR models (field variables like temperatures or flowrates) are now the outputs and model parameters obtained in DVR mode are inputs for the calculation. The important difference between DVR and the simulation mode is in degrees of freedom of the model. While DVR mode is usually the system with redundancy (DoR > 0), the simulation model must be just determined (DoR = 0). This can be done in two steps by:

1. Removing the redundancy by changing some measured variables to unmeasured ones, to create the just determined system, and
2. Exchanging some previously measured variables for model parameters.

Let's show the simulation of the heat exchanger model from the Appendix 3.

Example 5.1: Simulation of the heat exchanger

Let's recall the simple example of the heat exchanger presented in in the Example 3.1 in the Section 3.1.



DoR in this case is 1. It is therefore sufficient to put one measured variable among the unmeasured ones. We can select the temperature TCOUT (the cold water exit temperature).

In the next step we can exchange the second exit temperature THOUT (measured) for the unmeasured Heat transfer coefficient HTC, so far unmeasured. Such model can be used for the simulation. The inputs for the simulation can be the input flowrates, input water temperatures and the heat transfer area of the heat exchanger■

Some results of this example will be presented in the Example 5.3 where they will be compared with results obtained by the method of parametric sensitivities.

Example 5.2: Wat if? Queries (WiQ)

This example is the continuation of Example 5.1. WiQ are using the simulation model variant for predicting behavior of a plant under changed conditions. Process data are imported from a process data historian and the DVR calculation is executed. The operator can change values of one or more input variables and run the simulation.

Task 'HEX-SIMULRED /#1': What-If Queries

Set date and time

Input (influencing) variables

EDIT Restore

Type	Variable	Description	Value	Unit	NewVal	Dif
HF	QLOSS	heat loss to the environment	55,357	[KJ/S]	55,357	0
MF	COLDIN	cold stream in	55,604	[KG/S]	60,604	5
MF	HOTIN	hot stream IN	27,745	[KG/S]	27,745	0
T	TCINP	water cold input	19,695	[C]	17,695	-2
T	THINP	water hot input	89,847	[C]	89,847	0
V	A	Heat transfer area	200	[m2]	200	0
V	HTC	Heat transfer coefficient	581,75	[W/m2/K]	581,75	0

In this example the flowrate of the cold water (column Value) was increased (in the New Val column) by 5 kg/s (column Dif). Similarly the input cold water temperature was decreased by 2 K. Results of the simulation are shown in the next panel:

Output (influenced) variables

CHANGE

Type	Variable	Description	Value	Unit	NewVal	Dif
HF	Q	heat exchanged	4613,9	[KJ/S]	4790,9	177
T	TCOUT	water cold output	39,305	[C]	36,385	-2,92
T	THOUT	water cold output	50,152	[C]	48,622	-1,53

For example, the heat exchanged *Q* has increased by 177 kJ/s■

5.2 Parametric sensitivities

The table of parametric sensitivities (PS) should be available automatically after the DVR calculation is completed. The use of PS can be twofold:

1. In the phase of DVR (the redundant system) we can study the influence of measured data on DVR results. This application of PS was already described and discussed in Section 3.5.

2. In the phase of simulation (the just determined system with DoR = 0). This will be the theme of this Section.

PS are useful for optimization studies (prediction of the behavior of an industrial system under changed conditions). The existence of a model makes possible to calculate parametric sensitivities of calculated variables and important KPIs on model input variables. For example, how will be changed the heat rate of a steam cycle if the heat transfer area of some heater will be increased by 10 %? Or how will be changed the heat rate of a power plant if the temperature of the cooling water will be raised by 2 K?

In the previous section was described the simulation method. The simulation requires one calculation of a model to find one change of input variable(s) on some target variables (for example KPIs). The table of parametric sensitivities provides more complex information which can be used for example for the optimization of the plant performance. The cost for this advantage is that PS are based on linearization of the model with influence on results´ fidelity. Let’s see the difference via one simple example.

Example 5.3: Parametric sensitivities versus simulation

This example is the continuation of Example 5.1. The table of parametric sensitivities for the heat exchanged (Variable Q) follows:

```
REPORT ON PARAMETRIC SENSITIVITY
 ================================
Type Variable    Description
 ------------------------------
  HF  Q           heat exchanged

 GIVEN VARIABLE IS SENSITIVE TO:

 Type Measured variable    Sensitivity  Unit
 ------------------------------------------
Type Measured variable    Sensitivity  Unit
 ------------------------------------------
  HF  QLOSS                      0,120  [KJ/S] / [KJ/S]    heat loss to the environment
  MF  COLDIN                     9,932  [KJ/S] / [KG/S]    cold stream in
  MF  HOTIN                     55,629  [KJ/S] / [KG/S]    hot stream IN
  P   atm                   -1,1807E-3  [KJ/S] / [KPA]     atm. press.
  T   TCINP                    -67,104  [KJ/S] / [C]       water cold input
  T   THINP                     66,299  [KJ/S] / [C]       water hot input
  V   A                         12,772  [KJ/S] / [1]       Heat transfer area [m2]
  V   HTC                        4,301  [KJ/S] / [1]       Heat transfer coefficient [W/m2/K]
```

For example, the PS of the heat exchanged on the heat exchanger heat transfer area A is 12.772 kJ/s per one square meter. In the next table are results of the simulation and PS methods. The base case is the heat transfer area A = 200 m^2. This value was increased by 10 and 50 %. Results are compared in the next table:

Case	A [m^2]	Q – simul. [kJ/s]	Q – Param.sens. [kJ/s].	Difference [kJ/s].	Difference [%]
0	200	4658.16	4658.16	0	0

1	220	4901.88	4913.60	11.72	0.24
2	300	5683.56	5935.36	251.80	4.43

It can be seen that differences between these two methods (related to the value obtained by simulation) exist and depend on the distance from the base case (A = 200 m^2). For example, the relative difference between the simulation and the PS method in the Case 1 is 0.24 % of the Q calculated by the simulation. (very small difference).
Anyway, it should be stressed that all calculations are based on the assumption that the heat transfer coefficient (HTC) remains constant. Some corrections of HTC may be required (HTC depends on physical properties of fluids inside). This problem will be discussed in the next Chapter in Section 6.4■

5.3 Instrumentation systems optimization

DVR methods play important role in the design, upgrade and optimization of instrumentation systems in the process industries. Typical tasks in this area are:

- design of new instrumentation systems
- classification of variables (observability, redundancy, …)
- applications of the errors' propagation theory
- instrumentation cost optimization
- design and upgrade of nonredundant and redundant systems.

These problems are only briefly mentioned in [3], Chapter 5. The comprehensive treatment of this subject can be found in [7].

5.4 Plant performance analysis and optimization

Performance analysis is based on comparison of plant operation with some reference state. But, it is not possible to compare real process data and reference data directly. There is no sense in **direct** comparing of two states of a plant which differ for example in the load, fuel quality, cooling water temperature, etc.

One solution of this problem is the **Correction** of Operating State (OS) to the Reference State (RS) condition**.**

The other possibility is the comparison of plant's OS with the RS by **Expectation** (Prediction) of the plant behavior based on extrapolation from the RS to operating conditions [21].

The situation is depicted in the next table. Recall that the state of an operating plant has 2 dimensions:

- dimension of the state of Equipment
- dimension of values of process variables

Table 5.1: The Equipment state – Process Variables Matrix

Values of **Process variables**

		Reference State	**Operational State**
Equipment state	**Reference**	Reference (rated) performance (1)	Expected performance (2)
	Operational	Corrected operational performance (3)	Operational performance (4)

There are four states which are characterized by **two states of equipment** and by **two states of process variables**.

Both, Correction and Expectation can be done in practice for example by correction curves provided by equipment vendors. It can be done also (more efficiently) with the aid of DVR modeling by:

- process data driven simulation
- parametric sensitivities.

Example: 5.4: Corrections and expectations

This example concerns a classical power plant. Let’s suppose that the only difference between RS and OS is the cooling water temperature CW_T and the only KPI is the Heat Rate (HR). HR is the ratio of energy of fuel and the electricity produced. The Correction and Expectation steps are shown in the next figure:

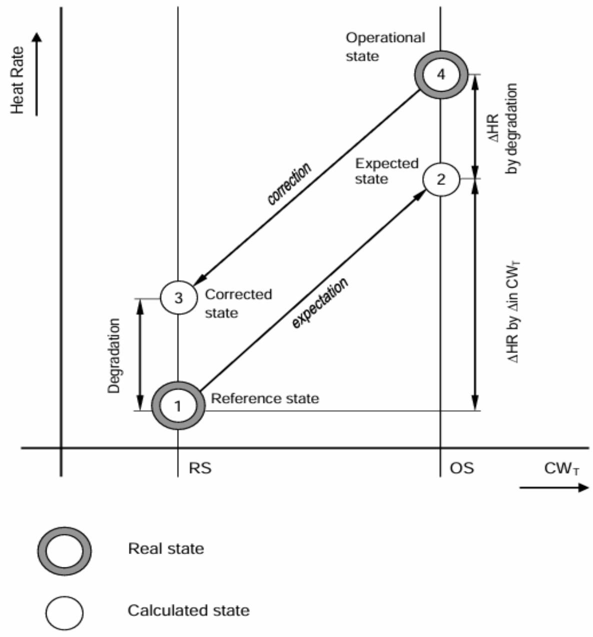


Fig. 5.1: Correction and Expectation (compare with Table 5.1)

In this case the real cooling water temperature is higher than in the RS which causes increase of HR from the state (1) to the state (2). This is our Expectation of plant's functioning under real operating conditions. Similarly, in the case of the Correction – the state (4) is corrected to the state (3). Notice that the difference (3) – (1) is one number while the difference (4) – (2) is a dynamic value depending on the real cooling water temperature**.**

The comparison of states (3) and (1) is the basis of **Monitoring Equipment Degradation.** The Degradation is characterized by one number independent on the power plant operation state. The comparison of states (4) - (2) and (2) – (1) is the basis of **Monitoring Plant Performance■**

Corrections and expectations shown in Fig. 5.1 can be calculated either by simulation or via parametric sensitivities.

Note 5.1: Instead of "Expectation" we can meet sometimes the term "Prediction". These two terms are synonymous■

The modeling of plants via process data driven simulation is the starting point for the classification of individual factors which influence plant economy. These factors belong to 3 categories:

- In the first category are External factors like ambient conditions, plant load. These factors can't be influenced by operators.
- In the second category are Internal factors which can be influenced by operators. This is the problem of the proper plant control.
- In the third category are losses caused by the Equipment degradation. This can be influenced by operators only partially, in some cases for example by cleaning of heat transfer areas. Some equipment degradation can require deeper maintenance.

A good Performance Analysis system should be able to separate the influence of three categories above and to quantify the amount of money which is lost by them.

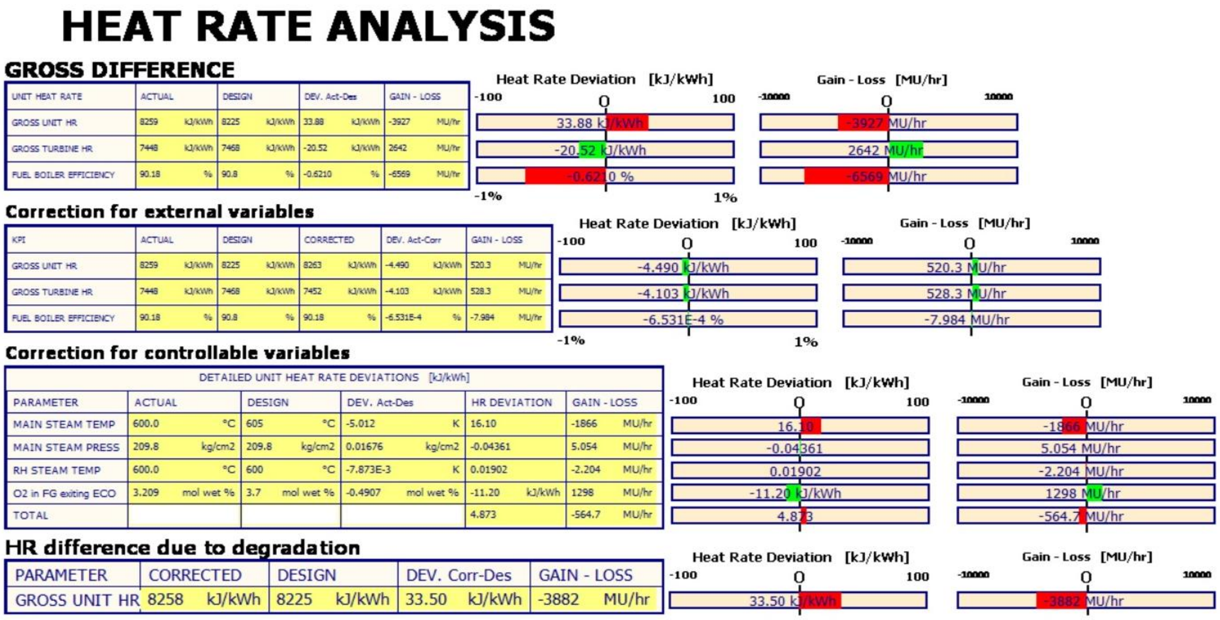


Fig.5.2: Example of a performance analysis dashboard – separation of factors influencing the plant economy [20]

The rest of this Section has no ambition to address advanced optimization methods like Real Time Optimization or optimization by a plant retrofit. For the off line optimization of a running plant we need to know:

The first: What is the real state of the system? This means complete and reliable mass and energy balance.

The second: To know the influence of control variables on the optimized KPIs. This can be solved by Data Driven Simulation or by Parametric Sensitivities

The third: How far are individual control variables from the optimum and what is their significance. This should be available for operators via active Performance Dashboards. More about this topic is in Section 6.3.

6 DATA AND MODEL MANAGEMENT

6.1 Working with process plant data

Process data measured by instrumentation are usually collected by a DCS system at high frequency (few seconds). For most of measured variables the process data historian then makes data compression by which real data are approximated by a system of connected straight lines. The maximum difference between real data and this approximation is set by the historian administrator. This maximum difference should be significantly less than measuring errors of individual measured variables.

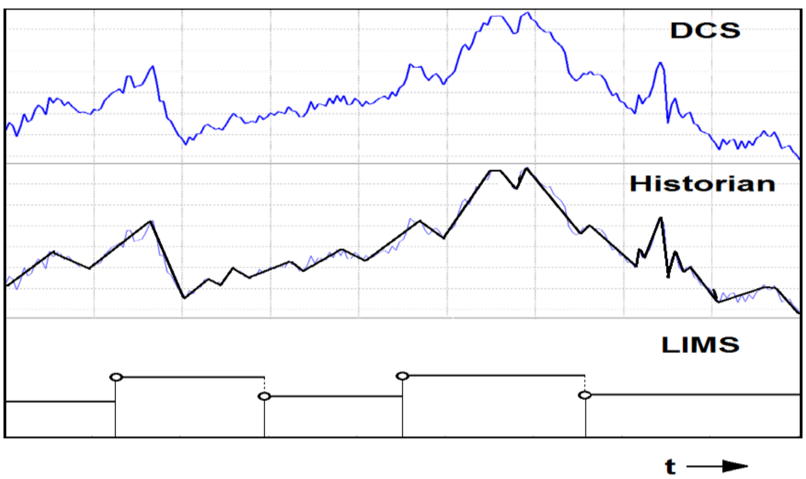


Fig. 6.1: Process data

In the case of laboratory data the situation is different. The frequency of lab data nowadays is not so high. Typical is one analysis per day or shift. It is supposed that the analyzed value is constant until the new analysis is available. Important is that there is some delay between the sampling and the knowledge of results. The issue of these laboratory data insufficiencies can be partially solved by so called *Quality estimators* (empirical regression models calculating the composition of streams on the basis of continuously measured variable like temperatures or flowrates, sometimes called *Soft sensors*).

Note 6.1: Nowadays, the main data sources for OLM are raw process data stored in historians. They are not based on classical relational databases – values of every variable are stored at different times. This is the main difference from results of regular balances which are calculated and stored at regular time intervals for all variables (average flows, temperatures, inventories, etc.)■

Note 6.2: Special kind of process data are so called *binary signals* which can have values 0 or 1 (ON/OFF). They serve for information whether some equipment or a plant part is in operation. Such signals need during OLM special treatment■

There are several possibilities of storing DVR inputs and results:

- to use some corporate database (historian or other database)
- to create the special database for DVR data
- variant 2 with sending selected results to the corporate database.

Every variant has its pros and cons (see Section 6.5). Variant 3 seems to be the best.

6.2 Balancing of nonstationary processes

Variability of industrial processes can be of three kinds [3]:

- minor fluctuations in the vicinity of nominal values (called *stationary* or *steady* state)
- long term trends with minor fluctuations – so called *quasi-stationary* state
- abrupt changes caused by change-over of plant to a new process regime, plant shut down, etc.

While the first two kinds of variability are manageable by present DVR methods, the third kind can cause problems. The issue of balancing dynamic processes solves theory of stochastic processes, for details see [3], sections 2.4 and A3.6.

The basis of models are balances of mass and energy. For any node, one can write the balance equations for mass, energy and further balanced variables. This equation reads generally

sum of inputs + source = sum of outputs + consumption + accumulation (6-1)

The accumulation means the increase of the balanced variable in the node (it can also be negative) – let us consider for example the change of mass in a reservoir.

The balance of a general unsteady state process is described by Eq. (6-1). The accumulation term can be defined as

accumulation = closing inventory - opening inventory (6-2)

or

$$a = \omega - \alpha \qquad (6\text{-}3)$$

where *a* stands for accumulation

α opening inventory

ω closing inventory.

This replacement of the classical accumulation by 2 new terms is not only a formal one. Both inventories are better for describing dynamic balances as they can be directly measured and their values can be used in models better than the accumulation (as will be shown later).

The steady state balance equations (6-1) and (6-3) can now be generalized for unsteady processes. Before this will be done, let's discuss one important difference between steady state and unsteady state balancing.

For steady state balancing of continuous processes the inputs and outputs of a node are usually regarded as rates (e.g. kg/s). For the unsteady state (dynamic) balancing is typical that it can be defined rigorously only for a specified time interval, say from time t_1 to time t_2. This defines the length of the time interval $\tau = t_2 - t_1$. Equations (6-1) and (6-2) can be now re-written as

$$\sum_i \varepsilon_i \int_{t1}^{t2} m_i \,\mathrm{d}t + \alpha - \omega = 0 \qquad (6\text{-}4)$$

where ε_i = +1 for inputs, ε_i = – 1 for outputs, summation over all streams incident with the node.

Such balance can be written also for components and energy.

For the original flow rates were substituted integrals of variable flow rates over the balancing interval. These integrals can be also expressed in the following form:

$$\int_{t1}^{t2} m_i \,\mathrm{d}t = \overline{m}_i \tau \qquad (6\text{-}5)$$

where $\overline{m}_i$ are the mean integral values of flow rates over the balancing time interval τ. The introduction of these mean integral values is important when dealing with balances based on real plant data. Mean values can be usually gathered from plant information systems (process historians) by standard queries.

It is worth mentioning the problem of determining the opening and closing inventories. In *chemical engineering* is distinguished between systems with *lumped* and *distributed parameters*.

A node with lumped parameters has constant values of all state variables in its volume. For example, a prototype of a node with lumped parameters is well known CSTR (Continuous Stirred Tank Reactor). This model can be accepted for situations where state variables are relatively homogenous in the node. Such assumption is typical of balancing of tank farms.

However, there are many situations where this model is not acceptable. The antonym to CSTR is a tubular reactor, a typical unit operation with distributed parameters. Further, let’s imagine a distillation column with significant concentration profile inside, or a pipeline transporting from time to time different liquid products or a gas of different density. Such systems can be sometimes “lumped”, which means separation of a system with distributed parameters into more subsystems with almost constant values of state variables.

From this discussion is clear that the opening and closing inventories of mass, heat or individual components can be gathered for nodes with lumped parameters. This assertion does not hold for nodes with distributed parameters.

There are 2 basic rules for continuous balancing with changing inventories:

1. The closing inventory of the previous time interval is the opening inventory for the next time interval
2. The balance of the $n+1^{th}$ time interval must not influence results of the balance of the n-th time interval (which has been already closed).

The rule 1 represents the continuity of the balancing process. The rule 2 guarantees the consistency of the whole balancing process: If every time interval is balanced (in the sense that inputs equal outputs), also the overall balance from t_0 to t_n is balanced. Important is that aside of balancing without inventories, we need not only data about the running time interval but also the reconciled inventory prom the previous interval (which must be read from some historical database and fixed).

It is usually supposed that dynamic balancing (balancing of a system at unsteady state) is much more difficult than balancing at steady state. However, this assertion does not hold generally. Let’s conclude this subsection by several comments:

- the most important attribute of a balanced system is the presence of inventories (holdups). If there are no significant inventories (and no possible accumulation), the only difference between steady state and dynamic balancing lies in using mean values instead of constant flow rates.
- the only significant difference between steady state and unsteady state balancing is in the case of balancing with significantly changed inventories.

6.3 Model variants

In practice several variants of models can be required, for example:

- variant for DVR
- variants for different production regimes
- variant for simulation.

Creation of different variants as **stand-alone models** is not efficient from the point of view of maintenance and further model development. The efficient solution is the so-called Base Case model with individual model variants stored in a database which contains only information about differences between the Base Case model and the individual variants. Typical differences are:

- classification of variables (measured/unmeasured)
- user defined equations (active/inactive)

In this way the whole model can be maintained and developed on the basis of the Base Case model without a need for modifying all variants. The simulation variant of the heat exchanger Base Case model presented in Appendix 3 was created in Example 5.1

6.4 Model maintenance

Another important part of models' maintenance is the problem of gross errors. After some GE is detected and localized, the natural solution of this issue is to put the problematic measured variable among unmeasured ones, until the instrument is repaired.

There are three important points of such activity:

1. **Who will do it?** There is a broad spectrum of people in contact with DVR (DVR providers and developers, process technologists, shift engineers, operators, …). The direct work with the DVR software requires quite deep knowledge of the user about the software. But a problem with GEs is revealed at the lowest level of users in the control room of a plant. So, there should be some interface between the DVR software and the final users of it. Such interface should be able to allow making changes of limited extent to prevent some damage of the model (typically to change Measured variables to Unmeasured and vice versa).
2. **How to do it?** The solution can be based on active dashboards available for shift engineers and operators. In this way users is not in a direct contact with the model itself. Process engineers (technologists) or shift engineers are probably the right persons to do it
3. **Do not forget it!** Important is to have some database of all activities in this area (the time of the change, the reason for it and who has done the change). This is important for **restoring** the DVR

system after the GE issue was resolved. If this is not respected, after some time the DVR system will collapse.

Active dashboards should allow detection, identification and elimination of GEs without a **direct** contact of users with the DR Engine. Typical solution should consist of:

- warning about the existence of GE (Status of data quality > 1).
- warning about other problems like not converging calculations, etc.
- seeing warning messages with suspect variables
- trying to put selected variables among unmeasured and see results
- selection of final eliminated variable(s) and the elimination proper. This activity is automatically recorded in a database.
- cancelation of elimination after the issue was resolved.

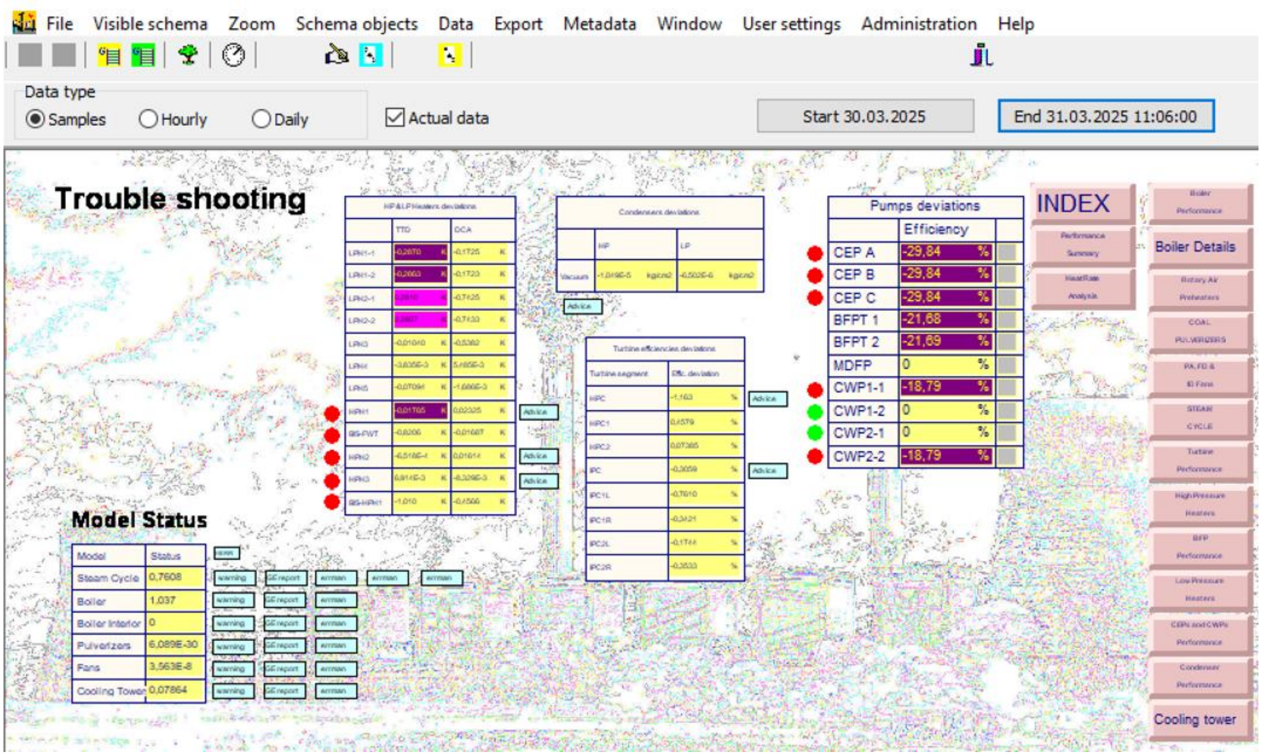


Fig. 6.2: Example of the interactive dashboard in PDIS [20]

6.5 Presenting DVR results

DVR systems have usually only limited possibilities of presentation of results to final users (in comparison with modern data historians and SCADA systems with their advanced graphics). There are usually several possibilities:

- to integrate the DVR system with some corporate system of data storing and presentation like PI, PHD, InfoPlus.21 or similar systems. The advantage of such solution is in easy access of the DVR results by all users. The main problem can be the cooperation of administrators of DVR and corporate systems (for example adding new results to the corporate system database). This can be crucial in the phase of DVR development.
- to use a special database and presentation system for reconciled data. As was already pointed out in Section 6.1, the structure of data in historians differs from data obtained from balancing systems. This fact can be utilized for effective data storing of DVR data in the special database. Also, the special DVR presentation may require functions that are not available in the corporate information system.
- the hybrid system where aside of the complete database of DVR data, a part of results is sent to the corporate database to be available for some corporate users. The complete separate DVR database can be used by plant operators and a smaller group of users for detailed data analysis, data mining and similar activities (process and shift engineers).

6.6 Process Data History – Data Mining

Historical process data is the invaluable source of information for plant's function analysis, equipment state diagnostics and generally for production optimization. Aside of using historical data for these purposes, it can be used also for improving models in the stage of model validation and improving. In practice, with the modern hardware and software there is no problem to store selected plant data for tens of years.

This short Section is about improving the simulation model by using the empirical model based on historical data. Such possibility is relevant mostly for improving simulation models. Using models based on constant parameters obtained on one set of data is limited to some near vicinity of the plant's operating point. The next example illustrates this topic:

Example 6.1: Modeling heat transfer in a power plant condenser

The functioning of the steam condenser influences the power plant efficiency significantly. In the basic simulation solution is supposed that the Heat Transfer Coefficient (HTC) is constant.

From the heat transfer theory of condensers follows that the heat transfer is influenced manly by the cooling water temperature and flowrate. The correlation and regression analysis of one year historical data revealed very good empirical model between HTC and two cooling water parameters.

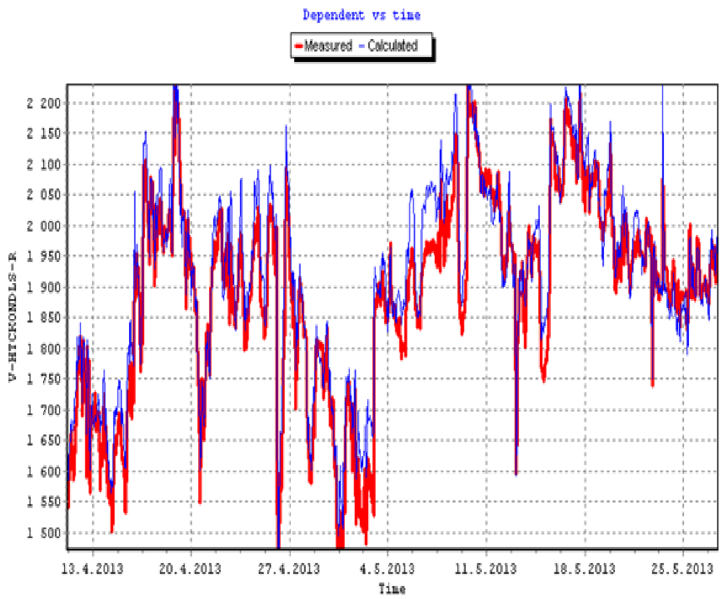


Fig. 6.3: Prediction of HTC in time (one year data)

The empirical model between HTC and the cooling water temperature and flowrate can thus improve the fidelity of the overall model significantly■

Note 6.3: Empirical models are different from the First Law models. They are only approximate and they **can't be used for increasing redundancy**. Their role is in necessary removing of unobservable variables or in improving quality of prediction of plant's behavior under changed conditions (simulation)■

The process of using empirical models for improving fidelity of models is simple:

1. Change the type of the model parameter (obtained in the DVR process) from Measured to Unmeasured
2. Add the model equation relating the parameter of the model to other process variables. In this way the degree of redundancy will remain the same as before the empirical model was introduced into the model.

Note 6.4: Nowadays are available many commercial software packages for data correlation and regression analysis. Some problem can be in the integration of such packages with process databases and historians. Data mining should be highly automatic and fast, without a need of manual data imports, exports, etc. In other words, many empirical model variants should be evaluated and tested by several key strokes on a PC (like in present gold mines where many tons of the ore must be crushed and processes for isolating one kilogram of gold)■

7 CASE STUDY: DETERMINATION OF NUCLEAR REACTOR THERMAL POWER (NRTP)

Steam generators (SG) in the nuclear power plant (NPP) convert a hot water into steam from heat generated in the NR core. As there exists no method of a direct measurement of the NR thermal power, the thermal power assessment is based on the detailed mass and energy balance of the SG system. The assessment of NRTP is very important for NPP economy. For NPPs are characteristic high CAPEX and low OPEX expenditures. It is therefore required to run NPPs at the highest possible NRTP. On the other hand side NRTP is strictly licensed by authorities. Decreasing the NRTP uncertainty thus makes possible to run NR closer to limits with significant profit increase.

This Chapter uses some special abbreviations. Their list is at the end of this Chapter.

7.1 Nuclear Steam Supply System (NSSS)

NSSS for a PWR consists of the reactor and the reactor coolant pumps, steam generators and further equipment in the containment with associated piping. A detailed description of such system can be found for example in the IAEA document [13]. There exists also the ASME PTC [14] which is the Performance Test Code targeted at procedures for conducting tests to determine the thermal performance of a NSSS including assessment of the Nuclear Reactor Thermal Power (NRTP). Even if this document is no longer an American National Standard or an ASME approved document, it can serve as a good starting point for a NSSS analysis.

In words, the NRTP can be expressed as:

NRTP = SG power – Electric Energy inputs + Loss , (7-1)

The simplest is the case of the overall balance of the NR containment, which contains a NR and steam generator. The balance envelope is in the next Fig. 7.1:

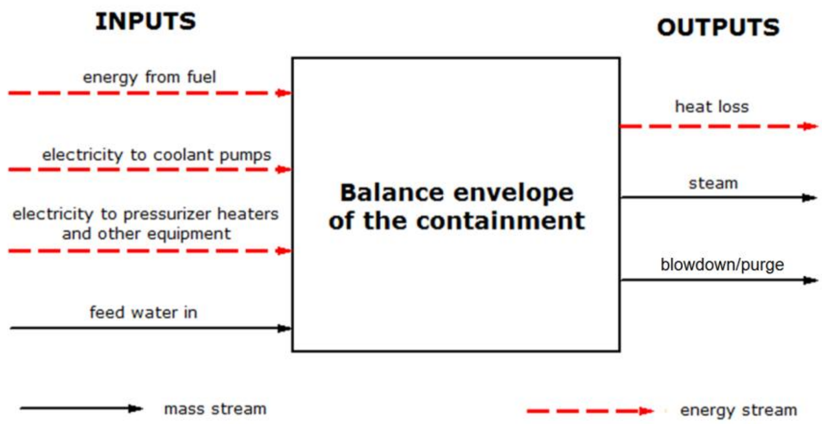


Fig. 7.1: Balance envelope of the containment

The NR thermal power (NRTP) is denoted here as “energy from fuel”. The mass and heat balance around this envelope generates 2 equations (one mass and one energy balance). In [14] the steam flow is supposed to be unmeasured and is calculated from the mass balance (the measurement of a wet steam is problematic). So, the remaining energy balance equation can be used for calculating the directly unmeasurable energy flux from the fuel, which is the NR power.

Inside this balance envelope there can be some measurements on a steam generator.

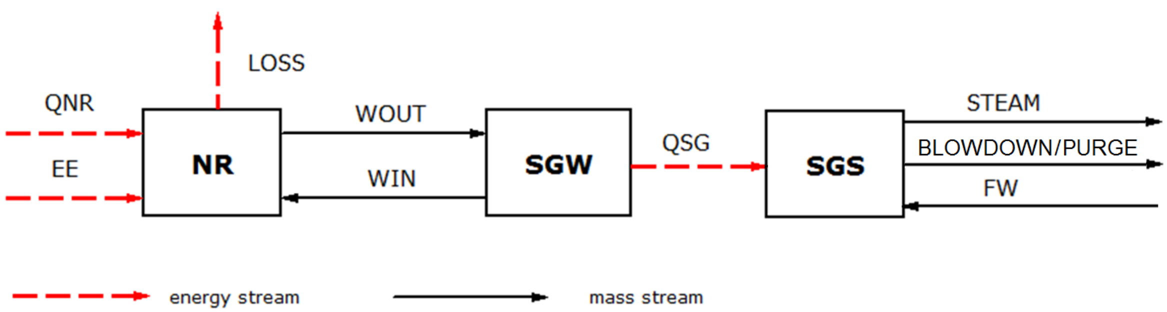


Fig. 7.2: Detailed balance flowsheet of the containment

The SGW means the hot water side of the SG, SGS means the steam side of the SG and QNR means the NRTP. QSG means heat transferred to the steam cycle. EE means mainly electric energy input into the containment.

In practice, there are usually 3 – 6 steam generators serving for one nuclear reactor. An example of the NSSS, the flowsheet is shown in the next figure:

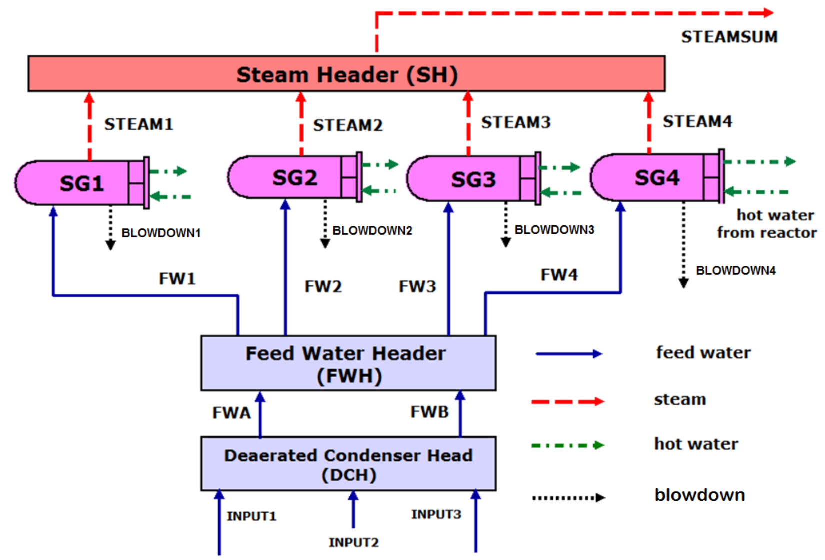


Fig. 7.3: The NSSS flowsheet

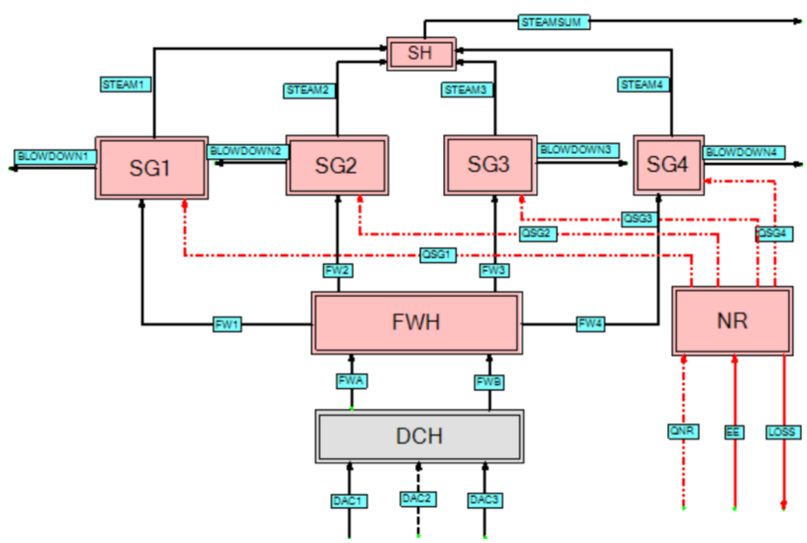


Fig. 7.4: Flowsheet for DVR [16]

Redundancy in such system stems from mass and enthalpy balances around 7 nodes and also from the temperature – pressure equilibria in steam generators (streams of steam). Let's suppose that there exist flowmeters on all streams of deaerated condensate (DAC), feed water at two levels (FW), steam from SGs and blowdown (purge) streams. Further, pressure and temperature are measured on all streams. In this example the following measurement uncertainties were supposed:

Table 7.1: Measurement uncertainties

Type	**Stream**	**Uncertainty**
Temperature	All	1 ^{o}C
Flow	STEAM	1.8 %
Flow	BLOWDOWN	3 %
Flow	FW	1.2 %
Pressure	All	0.5 %
Electricity input	EE	1%
Heat loss	LOSS	20 %
Wetness	STEAM	0.05 %

7.2 Influence of redundancy on the NRTP uncertainty

As can be seen from Fig. 7.3, there are several redundant measurements on the feed water streams between the deaerated condensate header and individual steam generators. This redundancy can be used for improving the reactor heat power accuracy (lowering its uncertainty). There exist also other benefits stemming from data reconciliation which will be studied later.

In the next table are uncertainties U of NRTP calculated for several variants of data redundancy (DoR), starting from the nonredundant system described in [14] to the system with maximal redundancy.

Table 7.2: Uncertainties U of NRTP for different variants of redundancy

No	Variant	DoR	U [%]
1	Balanced SGs, steam flows unmeasured	0	0.623
2	Variant 1 + measured SG steam flows + balance around SH	6	0.504
3	Variant 2 + balance around FWH	8	0.438
4	Variant 3 + balance around DCH	9	0.396
5	Variant 4 + water-steam equilibrium in SGs	14	0.396

It can be seen that the influence of redundancy on NRTP uncertainty is not negligible. Variant 1 recommended in [14] with zero redundancy has uncertainty 0.623 % while Variant 4 with DoR = 9 has uncertainty 0.396 %. The difference 0.227 % represents 2.27 MWe in the case of 1000 MWe nuclear block.

7.3 Parametric sensitivities and propagation of measurement errors

According to (3-12), calculated unmeasured variables are approximately linear functions of measured variables. It is therefore possible to estimate their sensitivities to measurement errors. Parametric sensitivities (PS) of NRTP to values of selected measured variables for Variant 4 are presented in the next table:

```
Table 7.3: REPORT ON PARAMETRIC SENSITIVITY
=================================
Variable Heat Flow  NRTP         NR heat power

 GIVEN VARIABLE IS SENSITIVE TO:
Type Measured variable    Sensitivity  Unit
 -------------------------------------------
  HF  EE                       -1,000  [MJ/S] / [MJ/S]    electric energy input
  HF  LOSS                      1,000  [MJ/S] / [MJ/S]    heat loss
  MF  BLOWDOWN1                -1,262  [MJ/S] / [KG/S]    blowdown (purge) from SG 1
  MF  DAC1                      0,359  [MJ/S] / [KG/S]    deaerated condensate 1
  MF  FW1                       0,721  [MJ/S] / [KG/S]    feed water 1
  MF  FWA                       0,366  [MJ/S] / [KG/S]    feed water A
  MF  STEAM1                    0,321  [MJ/S] / [KG/S]    steam from SG 1
  MF  STEAMSUM                  0,079  [MJ/S] / [KG/S]    steam to the turbine
  T   FWA                      -1,193  [MJ/S] / [C]       feed water A
  T   FWSG1                    -1,131  [MJ/S] / [C]       feed water to SG1
  T   SG1                      -0,141  [MJ/S] / [C]       steam generator 1
  T   steamsum                 -0,140  [MJ/S] / [C]       steam header
  X   SGsteam                 -25,541  [MJ/S] / [%]       wet steam

 Legend:

  HF  Heat flow
  MF  Mass flow
  T   Temperature
  X   Steam wetness
```

As can be seen from Fig. 7.3, the flowsheet is symmetrical according the vertical axis. The parallel streams have very similar values of parameters. For the reason of brevity their characteristics are represented by one value, for example FW1 represents FW1.

For example, the PS = 0.721 means that the increase of FW1 measured flow (for example caused by a measurement error) by 1 kg/s will cause the change of calculated NRTP by 0.721 MW.

Equation (3-12) is also basis for calculating propagation of measurement errors during calculation of final results. The variance (σ^2) of a resulting value is the sum of contributions of individual measured variables (so called shares of measured variables). The information of shares for NRTP is shown in the next table:

Table 7.4: THE VECTOR OF SHARES

REPORT ABOUT PROPAGATION OF ERRORS

==================================

Heat flow NRTP NR Thermal Power

THE VARIANCE OF GIVEN VARIABLE IS CAUSED MAINLY BY:

Type	Measured variable	Share	
MF	DAC1	9 %	deaerated condensate 1
MF	DAC3	9 %	deaerated condensate 3
MF	FW1	8 %	feed water 1
MF	FW2	9 %	feed water 2
MF	FW3	9 %	feed water 3
MF	FW4	9 %	feed water 4
MF	FWA	9 %	feed water A
MF	FWB	9 %	feed water B
MF	STEAM1	4 %	steam from SG 1
MF	STEAM2	4 %	steam from SG 2
MF	STEAM3	4 %	steam from SG 3
MF	STEAM4	4 %	steam from SG 4
MF	STEAMSUM	4 %	steam to the turbine
Sum		92 %	

There are 32 measured variables in the NSSS model. 92 % of the NRTP variance is caused by 13 measured variables in Table 7.4. The total contribution of remaining 19 measured variables is 8 % only. It is clear, that for lowering the overall variance (NRTP uncertainty) is important to cut down uncertainty of flowmeters in Table 7.4, especially flowmeters of feed water (better maintenance, calibration, installation of more precise ones). The opportunities of other measured variables are from this point of view negligible.

7.4 Gross errors detectability

In Section 4.3 was solved the following problem: What is the probability that a gross measurement error will be detected at all? Every redundant measured variable has its own Threshold Value (TV). A gross error greater than TV_β will be detected with probability greater than β. It was shown that TV_β depends on the adjustability of the variable, uncertainty of the measurement proper and also on Degree of Redundancy of the model. In the next table are some selected redundant variables with their TVs.

Table 7.5: REPORT ON GE DETACTABILITY

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 Beta: 90%	Beta: 95%	Unit Beta: 99%	
MF	BLOWDOWN1	0,000099	17,325	18,897	21,798	KG/S
MF	DAC1	0,217824	33,303	36,324	41,901	KG/S
MF	FW1	0,210281	16,385	17,871	20,615	KG/S
MF	FWA	0,228623	33,099	36,102	41,646	KG/S
MF	STEAM1	0,473045	17,728	19,336	22,305	KG/S
MF	STEAMSUM	0,793117	65,213	71,129	82,051	KG/S
T	FWA	0,184960	3,923	4,279	4,936	C
T	FWSG1	0,038540	8,268	9,018	10,402	C
T	SG1	0,024335	10,367	11,307	13,044	C
T	steamsum	0,552660	2,542	2,772	3,198	C

Legend:
Adjustability = relative cut of error due to reconciliation
Threshold value = gross error that will be detected with probability Beta

MF Mass flow
T Temperature

For example, for beta = 90 % (used throughout this case study) the flowrate FWA has TV_{β} = 33.099 kg/s. The flowrate of FWA equals 368.5 kg/s. TV is therefore ca 9.0 % of the measured value.

The flowrate STEAM1 has TV_{β} = 17.728 kg/s. The flowrate of STEAM1 equals 368.2 kg/s. TV is therefore ca 4.8 % of measured value.

The knowledge of GE detectability plays role in protection of target results (e.g. NRTP) against gross errors. Two factors should be taken into account:

1. TV of the redundant variable (the lower TV the better)
2. Parametric sensitivity of the target result on the measured variable (the smaller the better).

The complete solution of this important problem can be found in [15].

7.5 Protection of NR Thermal Power monitoring against Gross Errors

In the Section 4.6 was solved the question:

How well are protected Key results (targets) of the whole measurement against GEs?

Please recall theory presented in Section 4.6. Data needed for the solution are taken over from [15].

The problem statement:

It is required that the overall error of *NRTP* should not exceed 1.2 % of the nominal value, which is 3000 MW, i.e. 36 MW

The actual calculated reconciled NR thermal power

$NRTP = 2820.7 \pm 10.8$ MW,

therefore the uncertainty of *NRTP* belonging to random errors e_{hrmax} equals 10.8 MW (0.38% of the calculated value).

As the maximum allowed uncertainty is 36 MW, the undetected gross error should not cause greater error in *NRTP* than 36 – 10.8 = 25.2 MW (according to Eq. 4-11).

Data needed for the GE protection analysis is given in the next table:

Table 7.6: Analysis of MSSP for the Example

Type	Stream	Adjusta-bility	Threshold value *TV*	Parametric sensitivity ζ	$TV\,\lvert\zeta\rvert$ (the critical value is 26.8)
Flow	DAC1-3	0.23	55.4	0.367	20.3
Flow	FW1-4	0.17	31.7	0.729	23.1
Flow	FWA,B	0.23	55.3	0.368	20.4
Flow	PURGE	0.00	34.6	-1.29	44.6
Flow	STEAM	0.57	38.0	0.192	7.3
Flow	STEAMSUM	0.68	73.4	0.186	13.7
T	FWA,B	0.18	3.9	-0.176	0.7
T	FW1-4	0.04	7.9	-0.176	1.4
T	STEAM	0.03	10.2	0.014	0.1
T	STEAMSUM	0.55	2.5	0.014	0.0
Flow	***PURGE****	***0.02***	***1.8***	***-1.016***	***1.8***

* values after installation of the measurement of sum of purges (blowdowns)

Values in the last column are now compared with the critical value, which is 26.8 MW according to the Inequality (4-19). From the Table 7.6 follows that the target variable *NRTP* is quite well protected against gross errors for most of measured variables as they pass the inequality (4-19). The only exceptions are the PURGE streams.

Really, any of the purge streams has relatively high threshold value and at the same time also high parametric sensitivity. The value from the last column of Table 7.6 is 44.6 MW which is almost twice the allowed uncertainty for NRTP (26.8 MW). This means that the system is not protected against gross errors in purge flow measurements.

Let's try to raise the redundancy of the instrumentation system. The redundancy of the purge streams is very low (they are checked only by the balance of steam generators, while feed waters and steam have its own redundant balancing sub-flowsheets). By adding the measurement of the sum of all purge streams (uncertainty 5 % of the measured value), the problem is completely solved. After this step the threshold values of all purge streams fell from 44.6 to 1.8 kg/s. The result is presented in the last row of Tab. 7.6.

Interpretation of results

Results of the Example can be interpreted in the following way. For the whole system we can conclude that it is (after installing the new measurement of the sum of purges) well self-protected against gross errors as concerns the target variable *NRTP* and its required uncertainty. Especially:

The probability that any undetected gross error will impair the required uncertainty of NRTP (36 MW) is less than 10 %.

Similar, but sharper assertions, can be stated about individual measured variables. For example, for the measurement of steam flows from the individual steam generators holds (see Table 7.6), that

The probability that any undetected gross error in a steam flow will increase the error of NRTP more than 7.3 MW is less than 10 %.

Such interpretation can help in deciding which measured variables are self-protected by DR and which need independent checking, calibration or additional redundancy.

This example shows that DVR can help with improving precision of KPIs and also with detection of gross errors.

7.6 List of abbreviations for Chapter 7

DCH	Deaerated Condenser Header
FWH	Feed Water Header
NPP	Nuclear Power Plant
NR	Nuclear Reactor
NRTP	Nuclear Reactor Thermal Power
NSSS	Nuclear Steam Supply System
SG	Steam Generator
SH	Steam Header

8 DISCUSSION AND CONCLUSIONS

DVR is nowadays the matured method for analysis of operating process plants. Methods presented so far are based on more than 60 years of DVR development in the world-wide DVR community. Basic recommendations following from this Report are:

1. The basis of a plant model can be the combination of First Laws (mass, energy and momentum balancing), thermodynamic calculations and empirical relations based on data from equipment vendors and historical plant data (the hybrid model) (Chapter 3).
2. The Data Reconciliation proper requires efficient software capable of:
   - classification of variables
   - data reconciliation proper by a nonlinear optimization technique
   - all available information should be extracted (values of directly unmeasured variables and model parameters), uncertainties of results, parametric sensitivities, info about propagation of errors)
   - detailed discussion is in Section 3.6.
3. Important part of DVR is data analysis as concerns gross measurement errors (Chapter 4). GEs should be detected, identified and eliminated to prevent errors in other variables:
   - all available methods for Detection, Identification and Elimination of Gross Measurement Errors should be applied
   - important is the knowledge of the Power of GE detection test. Not all GEs can be detected
   - even if the presence of some GE is detected, frequently we get only a list of suspect measurement. The final decision usually requires further analysis
   - detailed discussion is in Section 4.7
4. Good validated data can be used for:
   - plant performance analysis and equipment diagnostics
   - instrumentation system maintenance and optimization
   - process data driven simulation
   - plant optimization.
5. Important is a good management of models and of validated data (Chapter 6). This concerns:
   - storing and presentation of validated data (important is the continuous analysis of historical process data and Data mining)
   - maintenance of models (model variants, changes in the process of GE elimination)
   - presentation of results.

To summarize, DVR with the aid of physical modeling is the basis for plant performance analysis, equipment diagnostics, process data driven simulation (Digital Twins) and for overall plant optimization.

LITERATURE CITED

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

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

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

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

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

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

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

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

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

[10] ASME PTC Code 19.1-218: Test Uncertainty

[11] ISO Guide 98-3-1995: Guide to the Expression of Uncertainty in Measurement

[12] VDI 2048 Blatt 1-2017: Control and quality improvement of process data and their uncertainties by means of correction calculation for operation and acceptance tests. VEREIN DEUTSCHER INGENIEURE 2017

[13] IAEA-TECDOC-1971 Thermal Performance Monitoring and Optimization in Nuclear Power Plants

[14] ASME PTC 32.1-1969(R1992): Nuclear Steam Supply Systems

[15] 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

[16] [RECON - Mass, heat and momentum balancing software with data reconciliation. RECON | ChemPlant Technology - process data information systems, mass and energy balancing software

[17] Abur, A., Exposito, A.G., 2004,: Power System State Estimation. Theory and Implementation. Marcel Dekker, New York

[18] Report CPT- 450-22. Process Data Reconciliation and Monte Carlo Method. Usti nad Labem 2022

[19] Report CPT- 445-21. Processing of Multiple Instruments Data. Usti nad Labem 2022

[20] PDIS – Process data information and data mining system PDIS | ChemPlant Technology - process data information systems, mass and energy balancing software

[21] Gay, R.R., Palmer, C.A. and Erbes, M.R.: *Power Plant Performance Monitoring*. Techniz Books International, New Delhi 2006

APPENDIX 1: MORE STATISTICS

A1.1: Testing statistical hypotheses – the χ^2 test

This Appendix is the continuation of Section 4.3. Its purpose is the deeper explanation of the χ^2 test.

χ^2 probability distribution

Let us have *v* random variables U_1, U_2, …, U_ν, mutually uncorrelated and having each the distribution $N(0,1)$. The random variable χ^2 is defined as the sum of squares of these random variables

$$\chi^2 = U_1^2 + U_2^2 + \ldots + U_\nu^2 \qquad \text{(A1-1)}$$

has the *chi-square distribution* with ν degrees of freedom, denoted as $\chi^2(\nu)$. Diagrams of probability densities of the χ^2-distribution are presented, for certain degrees of freedom, in Fig. A1.1. Critical values of χ^2-distribution can be found in statistical tables.

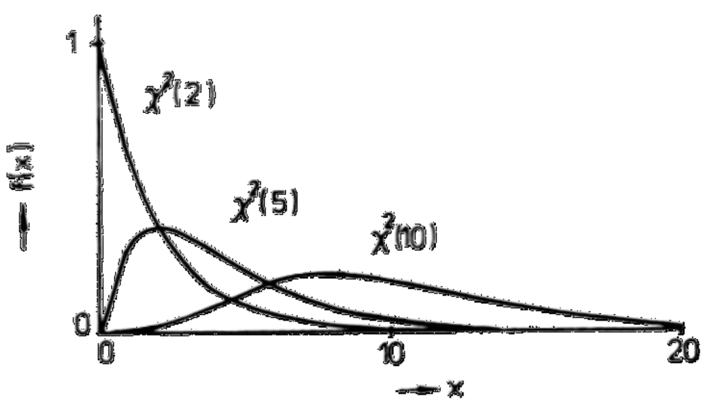


Fig. A1.1: Probability density functions of the χ^2-distribution

For the mean and variance (dispersion) holds

$$E[\chi^2(\nu)] = \nu \qquad \text{(A1-2)}$$

$$D[\chi^2(\nu)] = 2\nu \qquad \text{(A1-3)}$$

Noncentral χ'^2 distribution

Let us have v random variables U_1, U_2, …, U_ν, mutually uncorrelated and having each the distribution $N(\mu_i,1)$. The random variable χ^2 defined as the sum of squares of these random variables

$$\chi'^2 = U_1^2 + U_2^2 + \ldots + U_\nu^2 \qquad \text{(A1-4)}$$

has the *noncentral chi-square* distribution with ν degrees of freedom and with noncentrality parameter δ; it is denoted $\chi'^2(\nu, \delta)$. The noncentrality parameter δ is defined by

$$\delta = [\Sigma\mu_i^2]^{1/2} \qquad \text{(A1-5)}$$

for $i = 1,\ldots,I$.

If the noncentrality parameter equals zero, one obtains the common (central) χ^2 distribution. If a χ^2 distribution is given without more precise denotation, the central χ^2 distribution is meant. For the mean and the variance holds that

$$E[\chi'^2(v, \delta)] = \nu + \delta^2. \qquad \text{(A1-6)}$$

$$D[\chi'^2(v, \delta)] = 2\nu + 4\delta^2. \qquad \text{(A1-7)}$$

It is clear that with growing δ the probability density function of the χ'^2 moves to the right and becomes more flattened (see also the illustrative Fig. A1.1 where individual curves represent $\chi'^2(\nu, \delta)$. In the next figure are shown examples of probability density functions of the noncentral χ^2 distribution.

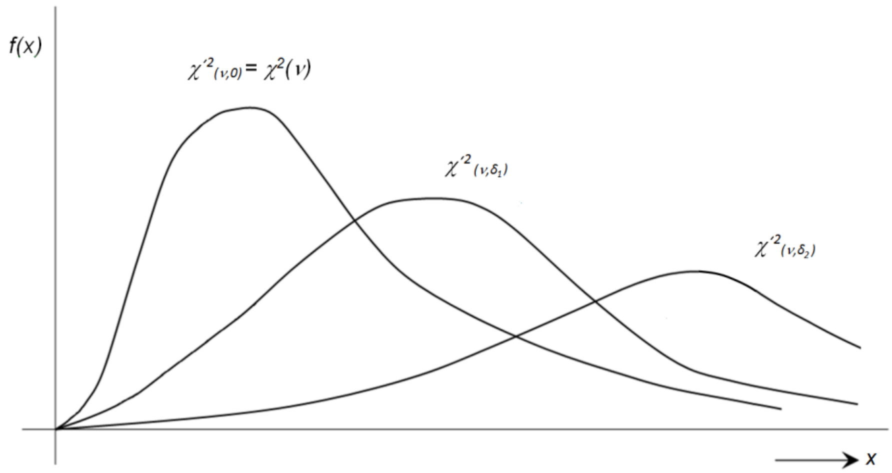


Fig A1.2: Examples of $\chi'^2(\nu, \delta)$ for one ν and $\delta_1 < \delta_2$

Now, let's return back to the problem of GE detection solved in Section 4.2. Let's suppose that there is one measured variable corrupted by a GE d_i. It is substantial that between δ and parameters of this variable holds ([3], p. 178)

$$\delta = d_i \sigma_{Vi} / \sigma_i^2 = q_i \sigma_{Vi} / \sigma_i \qquad \text{(A1-8)}$$

This is sufficient for construction of the power of the power characteristc curve. Now we will open the matter of *testing statistical hypotheses*. I am the witness that this theme is not very popular among engineers. Only very briefly (in [3] there are about 3 pages about this subject, starting p. 292):

One postulates a null hypothesis H_0 about data, in this case the hypothesis is (see Section 4.3):

There is no GE present (this means $d = 0$)

In testing hypotheses we can commit basically two kinds of errors. The **error of the I*st* *kind*** consists in rejecting the hypothesis while the hypothesis, in fact, holds true. Like in a court, the innocent man is hanged. The probability of this error is the significance level parameter α.

If the hypothesis H_0 does not hold but it is not rejected, one speaks about an **error of the IInd kind**. Like in a court, the criminal is not punished. The probability of an error of IInd kind is denoted β, and the value $(1 - \beta)$ is called the ***Power of the test***.

While the probability of an error of the Ist kind is a single number (equal to the significance level α), the power of the test depends on how much the null hypothesis differs from the reality. The dependence of β on the deviation (the magnitude of a GE) is called the ***operating characteristic* of the test**. The testing of hypotheses is illustrated by the next figure.

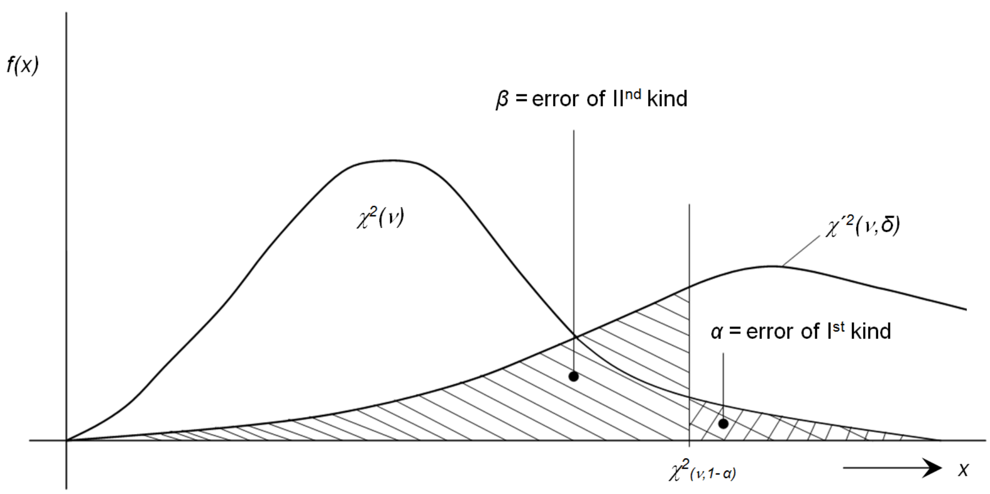


Fig A1.3: Testing the hypothesis about a GE by the χ^2 test

We can see two curves: one represents the central χ^2 distribution (probability density function) and the second is the noncentral $\chi'^2(\nu, \delta)$ where the parameter δ is the function of the GE d_i (see Eq. (A1-8)). On the x axis is the critical value $\chi^2_{(1-\alpha)}(\nu)$. The cross-hatched area α represents the error of Ist kind, the hatched area β is the error of IInd kind. This picture also illustrates the fact that decreasing α ($\chi^2_{(1-\alpha)}(\nu)$ moves to the right) means increasing β and vice versa.

The Appendix A1.1 presents a lot of statistical theory about GE detectability. We can ask: How efficient is it in practice? To answer this questions, probably the best is the Monte Carlo Method (MCM). In the report [18] is the Section 5 3 about GE detectability. It proves that the overall concept (χ^2 test, Threshold Values, etc.) works fine.

A1.2: More about Adjustability and Threshold Values

A lot of was written about redundancy. There is the **global redundancy** expressed by the Degree of Redundancy (DoR). As was already mentioned earlier, this notion is not so important (see Note 3.8). Much more important is the **redundancy of individual variables** expressed via their **adjustability**. There exists relation between adjustabilities and GE Threshold Values. The purpose of this Appendix is to explain this relation in details. What follows is the continuation of Section 4.3.

Recall the method of TV calculation by Equations (4-6) and (4-7):

$$q_i = \delta_\beta(\nu,\alpha)/[a_i(2-a_i)]^{1/2} \tag{4-6}$$

$$q_i = TV_i/\sigma_i \qquad \text{or} \qquad TV_i = q_i\,\sigma_i \tag{4-7}$$

These Equations use for TV calculation two parameters - the adjustability *a* and the standard deviation of the measurement error σ_i . In what follows will be shown two other methods of TV calculation.

Among 4 variables (adjustability a_i and standard deviations of measurement errors σ_i, reconciled values $\sigma_{x'i}$ and adjustments σ_{vi}) hold 2 equations:

$$\sigma_i^2 = \sigma_{vi}^2 + \sigma_{x'i}^2 \tag{3-15}$$

$$a_i = 1 - \sigma_{x'i} / \sigma_i \tag{3-18}$$

Only 2 of them are therefore independent. It is possible to use for calculating TV two different couples of variables.

In [3], page 179, is the calculation of q_i based on σ_i and σ_{vi} :

$$q_i = \delta_\beta(\nu,\alpha)\ \sigma_i/\sigma_{vi} \tag{A1-9}$$

In the next page 280 is another possibility - q_i as the function of $\sigma_{x'i}$ and σ_i . It is possible to substitute σ_i/σ_{vi} in (A1-9) by $\sigma_{x'i}$ from Eq. (3-15):

$$q_i = \delta_\beta(\nu,\alpha)/(1 - \sigma_{x'i}^2/\sigma_i^2)^{1/2} \tag{A1-10}$$

All variants, (4-6), (A1-9) and (A1-10) are equivalent and give the same results. The variant (4-6) is preferred as the dimensionless TV q_i is function of only one parameter – the adjustability a_i. This help in interpretation of GE detection and identification results.

Note A.1: In reality, the adjustability itself contains two parameters by its definition – see Eq. (3-18). Equations (4-6) and (A1-10) differs only in their divisors. The derivation of Eq. (4-6) starts by elimination of $\sigma_{x'i}$ from (A1-10) by insertion of $\sigma_{x'i}$ calculated from (3-18)

$$\sigma_{x'i} = \sigma_i (1 - a) \tag{A1-11}$$

After substituting $\sigma_{x'i}$ into (A1-10) and a simple equation rearrangement we get Eq. (4-6)■

APPENDIX 2: MASS BALANCE DVR MODEL

Throughout this document was used very simple model of the mass balance.

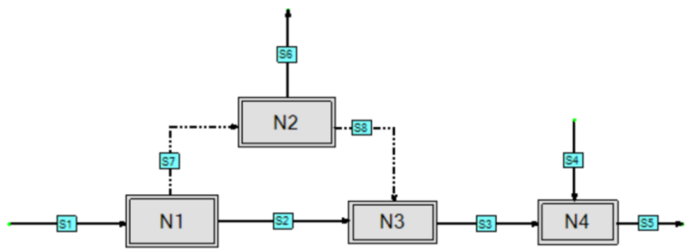


Fig. A2.1: Mass balance scheme

The model has 4 nodes and 8 streams. The full arrows (6) are measured streams and the dot-dashed arrows (2) are unmeasured streams. The mass balance model generates 4 balance equations.

The values and uncertainties of variables follow:

Input data

```
Task: MASBALL (Single-component balance)

 G L O B A L   D A T A

 Number of nodes                                    4
 Number of streams                                  8
 Number of components                               1

M A T E R I A L   S T R E A M S

 ID         Type           Value   Max.error
 -------------------------------------------
 S1         M           100,1000      2,0000%  KG/S
 S2         M            41,1000      4,0000%  KG/S
 S3         M            79,0000      2,0000%  KG/S
 S4         M            30,6000     10,0000%  KG/S
 S5         M           108,3000      4,0000%  KG/S
 S6         M            19,8000      4,0000%  KG/S
 S7         N            10,0000               KG/S
 S8         N            10,0000               KG/S
```

Results of data reconciliation

```
Task: MASSBALL (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

 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


 G L O B A L   D A T A

 Number of nodes                                       4
 Number of streams                                     8
 Number of components                                  1

 Number of measured variables                          6
 Number of adjusted variables                          5
 Number of non-measured variables                      2
 Number of observed variables                          2
 Number of non-observed variables                      0
 Number of free variables                              0
 Number of equations                                   4
 Number of independent equations                       4
 Number of user-defined equations                      0

 Degree of redundancy                                  2

 Mean residue of equations                             0
 Qmin                                         1,3081E+00
 Qcrit                                        5,9900E+00
 Status (Qmin/Qcrit)                             0,21837


 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


 End of results
```

APPENDIX 3: HEAT EXCHANGER DVR MODEL

Throughout this document was used very simple model of the countercurrent heat exchanger described in Section 3.1.

The heat exchanger in Fig. 2.1 serves for exchanging heat between the COLD and hot streams.

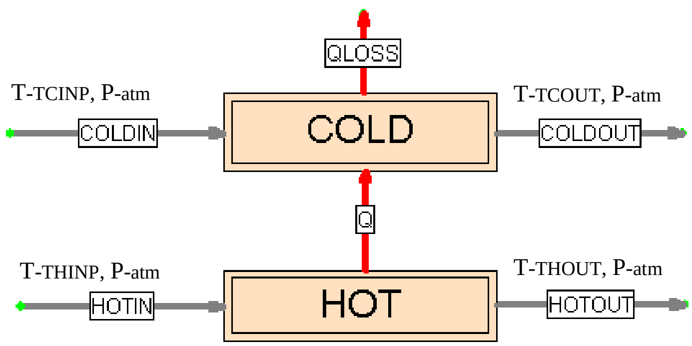


Fig. A3.1: Heat exchanger balance scheme

The model has 2 nodes (COLD and HOT sides of the exchanger, 4 mass streams and 2 heat fluxes (exchanger heat flux Q and heat loss QLOSS). The mass flows are measured at the inlets to the exchanger, measured are also all input and output temperatures. The pressure is atmospheric. We suppose that the specific enthalpy of water does not depend on the pressure. The heat loss from the shell to the environment is approximately known (estimated).

The model generates altogether 4 balance equations – 2 mass balances and 2 energy balances around both of the nodes. The model has now three unknowns – the heat flux Q through the exchanger (red energy stream) and two unknown flowrates at the outlets from both nodes. The equations of the model are:

(1) $F_{HOTIN} - F_{HOTOUT} = 0$

(2) $F_{COLDIN} - F_{COLDOUT} = 0$

(3) $F_{HOTIN}*ENT(T_{HINP}) - Q - F_{HOTOUT}*ENT(T_{HOUT}) = 0$

(4) $F_{COLDIN}*ENT(T_{CINP}) + Q - F_{COLDOUT}*ENT(T_{COUT}) - QLOSS = 0$

The fifth and sixth equation of the model can be definitions of the Logarithmic Mean Temperature Difference (LMTD) for the countercurrent heat exchanger and the Heat Transfer Coefficient HTC

(5) $LMTD - [(THOTIN - TCOLDOUT) - (THOTOUT - TCOLDIN)]/LN([(THOTIN - TCOLDOUT) / (THOTOUT - TCOLDIN)])$

(6) $Q - HTC*A*LMTD(T_{HOTIN}, T_{HOTOUT}, T_{CINP}, T_{COUT}) = 0$

where	F* are	flowrates
	T*	temperatures
	ENT(T*)	specific enthalpy function
	HTC	Heat Transfer Coefficient
	A	heat transfer area

LMTD Logarithmic Mean Temperature Difference

There are the following equations and variables in the model:

- 6 model equations
- 6 measured variables (2 flowrates, 4 temperatures, heat transfer area, heat loss flux)
- 5 unmeasured variables (2 flowrates, heat flux of the exchanger Q, HTC and LMTD)

The values and uncertainties of variables follow:

Input data

```
Task: One heat exchanger

G L O B A L   D A T A

 Heat balance calculations                             No

 Number of nodes                                        2
 Number of heat nodes                                   2
 Number of streams                                      6
 Number of energy streams                               2
 Number of components                                   1
 Number of temperatures                                 4
 Number of pressures                                    1
 Number of auxiliaries                                  3

 N O D E S

 ID         Description                   Remark
 -------------------------------------------------------------------
 ENVIRON    Environment                   unbalanced
 COLD       cold side of a heat exchanger
 HOT        hot side of a heat exchanger

M A T E R I A L   S T R E A M S

 ID         Type           Value   Max.error
 -------------------------------------------
 COLDIN     M            55,5000      1,0000   KG/S
 COLDOUT    N            55,5556               KG/S
 HOTIN      M            27,8000      0,5000   KG/S
 HOTOUT     N            27,7778               KG/S

 E N E R G Y   S T R E A M S    [KW]

 ID         Type           Value   Max.error
 -------------------------------------------
 Q          N          4000,0000
 QLOSS      M            55,0000      30,0000%


 T E M P E R A T U R E S    [C]

 ID         Type           Value   Max.error
 -------------------------------------------
 TCINP      M            20,0000      1,0000
 TCOUT      M            39,0000      1,0000
 THINP      M            90,0000      1,0000
 THOUT      M            50,0000      1,0000
```

```
P R E S S U R E S    [KPA]

ID         Type           Value   Max.error
-------------------------------------------
atm        M           100,0000      1,0000
A U X I L I A R I E S

ID         Type           Value   Max.error
-------------------------------------------
A          M           200,0000      1,0000   m2
HTC        N           500,0000               W/m2/K
LMTD       N            30,0000               K


U S E R   E Q U A T I O N S

ID         Description                                                Remark
           Programmatic code
---------------------------------------------------------------------------------
HTC        Exchanger HOT -> COLD: Heat-transfer coefficient           Model
           [S<Q>]-[V<HTC>]*[V<A>]*[LMTD2<THINP:THOUT:TCINP:TCOUT:0>]

LMTD                                                                  Model
           [V<LMTD>]-[LMTD2<THINP:THOUT:TCINP:TCOUT:0>]
```

Results of data reconciliation

```
Task: One heat exchanger

I T E R A T I O N S

Iter            Qeq            Qx              Qy             Qmin
-----------------------------------------------------------------
START     1,3593E+05
   1      1,8876E+02    4,4703E+01      1,2277E+05      9,8514E-01
   2      2,8487E-04    3,2059E-02      1,7760E+01      9,8373E-01
   3      3,9549E-09    1,6734E-08      3,6858E-06      9,8373E-01

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

G L O B A L   D A T A

Number of nodes                                    2
Number of heat nodes                               2
Number of streams                                  6
Number of energy streams                           2
Number of components                               1
Number of temperatures                             4
Number of pressures                                1
Number of auxiliaries                              3

Number of measured variables                       9
Number of adjusted variables                       8
Number of non-measured variables                   4
Number of observed variables                       4
```

```
Number of non-observed variables                  0
Number of equations (incl. UDE)                   5
Number of independent equations                   5
Number of user-defined equations (UDE)            2

Degree of redundancy                              1

Mean residue of equations                3,9549E-09
Qmin                                     9,8373E-01
Qcrit                                    3,8400E+00
Status (Qmin/Qcrit)                         0,25618

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

Name             Type      Inp.value      Rec.value      Abs.error
-------------------------------------------------------------------
COLDIN           MC           55,500         55,604          0,977  KG/S
COLDOUT          NO           55,556         55,604          0,977  KG/S
HOTIN            MC           27,800         27,745          0,488  KG/S
HOTOUT           NO           27,778         27,745          0,488  KG/S

E N E R G Y   S T R E A M S

Name             Type      Inp.value      Rec.value      Abs.error
-------------------------------------------------------------------
Q                NO         4000,000       4613,920        161,961  KW
QLOSS            MC           55,000         55,357         16,485  KW

T E M P E R A T U R E S

Name               Type      Inp.value      Rec.value      Abs.error
---------------------------------------------------------------------
TCINP              MC           20,000         19,695          0,798  C
TCOUT              MC           39,000         39,305          0,799  C
THINP              MC           90,000         89,847          0,953  C
THOUT              MC           50,000         50,152          0,954  C

P R E S S U R E S

Name               Type      Inp.value      Rec.value      Abs.error
---------------------------------------------------------------------
atm                MC          100,000        100,000          1,000  KPA

A U X I L I A R I E S

Name               Type      Inp.value      Rec.value      Abs.error
---------------------------------------------------------------------
A                  MN          200,000        200,000          1,000  m2
HTC                NO          500,000        581,751         25,949  W/m2/K

End of results
```