

Report CPT-229-07

Date of issue: 30. 1. 2007

This report has 133 pages.



Process data validation in practice

Applications from chemical, oil, mineral and power
industries

*Authors: František Madron
Vladimír Veverka
Miloslav Hošťálek*

Ústí nad Labem

January 2007

This report is the integral part of program RECON documentation. No part of this document can be reproduced or distributed in any form without permission in writing from ChemPlant Technology, s.r.o.

CONTENT

1	INTRODUCTION	1
2	MATHEMATICAL MODELS	3
2.1	Basic notions	3
2.2	Balance equations	4
2.3	Mass and chemical component balances	5
2.4	Balance of energy	8
2.5	Phase equilibria	10
2.6	Enthalpies of streams	11
2.7	Momentum balance	12
2.8	User defined variables and equations	13
2.9	Model of an industrial system	14
2.10	Model errors	16
3	STATISTICAL TREATMENT OF MEASURED DATA	19
3.1	Measurement errors	19
3.2	Classification of variables	20
3.3	Degree of redundancy	22
3.4	Data reconciliation	22
3.5	Benefits of the reconciled data	25
3.6	Detection of gross errors	27
3.7	Identification and elimination of gross errors	29
3.8	Efficiency of gross measurement errors detection	32
3.9	Propagation of errors at data processing and measurement system optimization	34
3.10	Parametric sensitivity	36
3.11	Data validation in industrial processes monitoring	37
3.12	Integration of DR into process information systems	42
4	EXAMPLES OF MASS AND COMPONENT BALANCES	44

4.1	Mass balance at steady state without redundancy	45
4.2	Mass balance at steady state with redundancy	49
4.3	Mass balance in steady state with unobservable variables	51
4.4	Mass balance at steady state – an unsolvable system	53
4.5	Multicomponent balance of the LPG separation	55
4.6	Component balance of the chlorination of methane	58
4.7	Ammonia synthesis reactor – master and dependent streams	62
4.8	Multicomponent balance of a flotation circuit	65
5	MODELS OF THERMAL UNIT OPERATIONS	67
5.1	Mixer of thermal streams	69
5.2	Simple heat exchanger	71
5.3	General heat exchanger	73
5.4	Steam generator	76
5.5	Condensation of steam	79
5.6	Expander	82
5.7	Throttling of steam	84
5.8	Wetness separator	87
5.9	Input / output of energy (pump)	89
5.10	Phase equilibrium water – water vapor	95
5.11	User defined variables and equations	98
5.12	Splitter of streams	100
	CASE STUDY 1: IDENTIFICATION OF STEAM GENERATOR THERMAL POWER	101
	Propagation of errors in data processing	102
	Strategy of measurement	103
	CASE STUDY 2: IDENTIFICATION OF NUCLEAR REACTOR THERMAL POWER	105
	Main results	106
	Further information	107
	Detection and identification of gross errors	108

LITERATURE	110
APPENDIX 1: DISTRIBUTION OF RANDOM VARIABLES	111
D1.1 Normal (Gaussian) distribution	111
D1.2 Distribution χ^2	112
APPENDIX 2: MEASUREMENT ERRORS	114
D2.1 Basic concepts and classification of errors	114
D2.2 Correlated random errors	116
D2.3 Systematic errors	119
D2.4 Gross errors	120
D2.5 Measurement precision and accuracy	120
APPENDIX 3: TESTING OF STATISTICAL HYPOTHESES	122
APPENDIX 4: EFFECTIVENESS OF GROSS ERROR DETECTION	125
APPENDIX 5: MOMENTUM BALANCE IN A PIPELINE	127
APPENDIX 6: CLASSIFICATION OF VARIABLES	129
APPENDIX 7: ABOUT PROGRAM RECON	132

1 Introduction

Information about industrial systems is gained via measuring physical variables (flow-rates, temperatures, pressures and the like). In practice, the measurement itself is frequently subject to measurement errors, usually classified as random, systematic and gross. These errors often unfavorably affect the results of the measurement, as well as other derived variables.

The data **validation** then means increasing their quality, attained in particular by elimination of gross and compensation of systematic errors, and minimizing the influence of random errors. The most frequent approach consists in the idea that the data must obey the exactly valid mathematical models (physical laws). The typical representatives are balance models for mass and energy (the laws of their conservation) that hold true under very general conditions.

A fundamental method applied to the validation of data is their **reconciliation** making them obey the physical laws. The common expression is just *Data Reconciliation*, briefly *DR*.

The methodology of measured data reconciliation based on exactly valid models is not new. It was developed more than 200 years ago and used first in exact sciences, such as astronomy or geodesy. The first industrial applications began in the sixties of the 20-th century (crude oil processing) and have spread rapidly into further industrial domains (general chemicals, petrochemicals, minerals processing and power generation, including nuclear power plants).

From the first applications, *DR* successively became a **standard method** in a number of areas that have direct influence on economy and process safety, such as for example:

- efficiency monitoring of technological processes (specific consumption of energy, Key Performance Indicators, etc.)
- integration of industrial measurements into plant accounting
- instrumentation maintenance
- Advanced Process Control

The aim of the handbook presented is to gather the most important methods to be applied to data validation in the chemical and power industries. In addition to mass and energy balances that form the substantial part of the mathematical models, briefly mentioned will also be momentum balances that can be applied to flow of fluids in pipeline systems, and phase equilibrium.

The data validation is based on (and falls with) three necessary assumptions:

- the existence of: an exactly valid mathematical model involving measured variables as variables, most frequently in the form of mass and energy balances
- the measurement errors model, commonly in the form of uncertainties (maximum errors) in the measured values
- data proper, which must be in a certain sense redundant (they can be mutually confronted as to their compatibility on the base of the mathematical model).

The handbook has four parts. The first part is devoted to the problems of mathematical models and oriented towards models occurring in the chemical and power industries. The second part deals with statistical treatment of measured data,

which is always the base of the validation proper. The third, pragmatically oriented, contains examples of unit operations typical of the chemical and power industries and from which the models of more complex systems are composed. In the fourth part, there are then several more extensive case studies taken from practice.

The theoretical part of the handbook presumes certain minimal knowledge about mathematical modeling and statistical treatment of measured data. So as to make it easier, it contains several supplements with the aim to help the reader in brushing up earlier acquired knowledge. However, the supplements are not intended for primary study.

Even in simple cases, the data validation in practice is not applicable without computer technique. A number of simple examples will be solved using the balancing program RECON. The reader will have the possibility to configure himself/herself the examples described in the handbook and try to solve them. The handbook is also supplemented by files with solved examples.

The present document considers its goal to show the reader

- how easy and rapid it is to create the balancing flowsheets based on exact physical models
- what it is the data validation (giving more precision, protection of the data against gross errors)
- how to compute the directly unmeasured variables
- the possibilities of data reconciliation in monitoring industrial processes
- employing data reconciliation in measurement systems maintenance
- how to optimize the existing measurement system
- how to automate the whole validation process and create thus storage of validated data ready for further application
- how to create a system that, based on current measurements, enables one acquiring on-line information, which had earlier resulted from laborious and expensive process measurements and tests.

2 Mathematical models

Mathematical models can be classified from different points of view. One of them is how truly the model describes the reality. We then distinguish *approximate* and *exact* models. In the latter case, we also speak of the *laws of nature*. If there is a contradiction between the exactly valid model and the measured data then it is necessary to examine the data or the conditions under which the exact models have been applied.

The basis for the data validation is a confrontation of the data with exactly valid mathematical models. In most cases, these are the balances of mass, energy and possibly also other variables. In some situations, also models of phase equilibrium may be used.

2.1 Basic notions

The technological scheme is divided into units (called **nodes**) connected by **streams** of material and energy. Typical nodes are apparatuses or their parts. For the needs of balancing we frequently also introduce as nodes the connections or disjoints of pipes.

Most models are composed of one-node balances illustrated in the following figure.

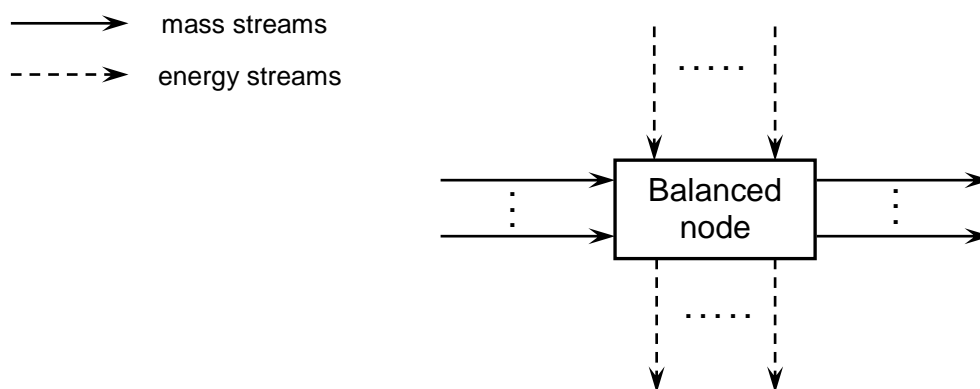


Fig. 2.1-1: Streams incident with one node

Material streams clearly need no explanation. Note only that generally, the streams are mixtures of several components (chemical individua), usually in one state of aggregation (most frequently gas or liquid). In the special case of power plants, typical representatives are for example feed water or steam. Prevaingly, the chemical individuum is then H₂O in different states of aggregation (liquid, vapor or also both: a two-phase mixture).

Energy streams represent net transport of energy without material carrying medium. Representatives are in particular

- heat losses by radiation and heat conduction into the environment
- heat flow in heat exchangers (between the tube and shell sides)

- work on the shaft exerted in turbines
- work supplied to pumps by motors
- electric power

While in mass balancing the flow-rates can be directly measured by flow-meters, such a case is rather an exception for energy flows. There is thus a difference between the flow of energy associated with a real mass stream and a net energy stream.

We'll be further interested in the one-node balance. The combination of several nodes creating a balancing scheme will be dealt with in Section 2.9.

2.2 Balance equations

For any node, one can write the balance equations for mass, energy and further balanced variables. This equation reads generally

$$\text{sum of inputs} + \text{source} = \text{sum of outputs} + \text{consumption} + \text{accumulation} \quad (2.2-1)$$

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

The source or the increase of the balanced variable inside the system can be caused for example by the generation or vanishing of chemical components as a consequence of chemical reactions. Observe that according to this conception, we do not consider the loss of a balanced variable into the environment; the latter should be represented by a stream directed from the node into the environment.

For the variables subject to the laws of their conservation (mainly mass and energy), the source and consumption equal zero.

For the steady-state (stationary) processes, we assume zero (or negligible) accumulation. The balance equation for mass or energy is then, for a stationary process, of the simple form

$$\text{sum of inputs} = \text{sum of outputs} \quad (2.2-2)$$

While in the mass balance the summation concerns the material streams only, the balance of energy includes the energies of all the streams.

So long as we consider (at least on a rough time scale) the process at *steady state*, there is no accumulation of the balanced variables in the node. The balances can then be expressed in variables per unit time, i.e. in flowrates (e.g. kg/s for mass, J/s = W for energy). These flow-rates are then, in fact, mean values of instantaneous flow-rates in the chosen interval (for instance 1 h). Under the assumption that no relevant fluctuations occur, also the parameters of the streams (such as temperatures and pressures) can be taken as averages.

2.3 Mass and chemical component balances

The mass balance is clearly the most familiar type of balance. It is a special type of so-called *single-component* balance ignoring the composition of the streams; we are then only interested in the total mass of the stream. If two or more chemical individuals (components) in the stream are distinguished, they can be balanced separately and one then speaks of a *multi-component* balance.

2.3.1 Mass balance

The balance of mass for a given node in steady state is straightforward:

$$\sum_{i, \text{in}} m_i = \sum_{i, \text{out}} m_i \quad (2.3-1)$$

thus the sum of input flow-rates (i, in) equals that of the output ones (i, out); m_i is the value of the flow-rate in stream i .

2.3.2 Component balance

Let us first consider the case where individual components are conserved in the balancing node (in practice, there are no chemical reactions among components). If there are K chemical components in the system then for each of the nodes we have, in the steady state, K balance equations

$$\sum_{i, \text{in}} m_{ik} = \sum_{i, \text{out}} m_{ik} \quad , \quad k = 1, 2, \dots, K \quad (2.3-2)$$

where m_{ik} is the mass flow-rate of the k -th component (chemical species) in the i -th stream incident to the node.

In practice, the material flow-rates m_{ik} are usually expressed in terms of total flow-rate m_i and mass fraction x_{ik} of the k -th species in the i -th stream

$$m_{ik} = m_i x_{ik} \quad (2.3-3)$$

The equation (2.3-2) is written for every component, which is incident with the node (the component is present at least in one stream connected with the node).

If all species present in the mixture are being balanced then the mass fractions in each of the streams must obey the so-called *normalization equation*

$$\sum_k x_{ik} = 1 \quad . \quad i = 1, 2, \dots, I \quad (2.3-4)$$

It is worth mentioning the problem of so-called *splitters*. A *simple splitter* is a node with one input stream and two or more output streams of the same composition (which, of course, is the same as the composition of the input stream) – imagine a pipe branched into two or more pipelines. Around any simple splitter only one balance equation is written (mass balance). The component balances are dependent on the mass balance equation, as the reader can easily verify. A *general splitter* can have more input streams of different composition and the same holds for output streams except the condition that at least two output streams are of the same composition. It can be shown that any general splitter can be transformed to a system of one standard node and one or more simple splitters.

This complication is solved in RECON by so-called *dependent* and *master* streams. The composition of a dependent stream is the same as that of the master stream (in essence, there is no concentration of a dependent stream in computer memory, only a pointer to the concentration of the master stream). This is very easy way how to tell RECON about this quite complicated situation. Of course, there is one limitation: The master stream cannot depend on any other stream.

2.3.3 Component balance with chemical reactions

In the case where chemical reactions are present, the conservation of individual species does no longer hold. The multi-component balance of a node at steady state then reads

$$\text{sum of inputs} + \text{source} = \text{sum of outputs} + \text{consumption} \quad (2.3-5)$$

Admitting the convention that the consumption is virtually a negative source, Eq. (2.3-5) is further simplified

$$\text{sum of inputs} + \text{source} = \text{sum of outputs}. \quad (2.3-6)$$

Let us now assume that there are R chemical reactions running in the balancing node (reactor). The stoichiometry of the reactions can be expressed in form of the matrix of stoichiometric coefficients \mathbf{B} whose elements B_{rk} represent the stoichiometric coefficients of the k -th species in the r -th equation (by convention, the initial reaction species have positive, the products negative stoichiometric coefficients).

$$\sum_k B_{rk} S_k = 0 \quad , \quad r = 1, \dots, R \quad (2.3-7)$$

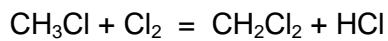
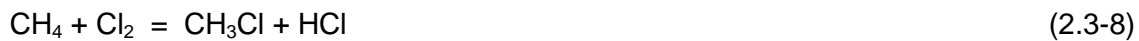
where S_k is the k -th species.

The values of the stoichiometric coefficients depend on the units used in the balancing. In chemistry, the balancing of chemical species is traditionally in molar units, such as kilomoles (kmol). In technological practice, more common is the

balancing of mass variables, for example kilograms. In this case, the stoichiometric coefficients are given as products of the species (molar) coefficient and mole mass of the respective species (kg/kmol).

Example 2.3-1: Chlorination of methane

In a reactor, the chlorination of methane is described by the following stoichiometric equations:



There are 5 species in the system: CH₄(1), Cl₂(2), CH₃Cl(3), CH₂Cl₂(4) and HCl(5). The set of stoichiometric equations (2.3-8) then can be expressed by the matrix of stoichiometric coefficients

No. of species ⇒	1	2	3	4	5	No. of reaction
						↓
B =	1	1	-1	0	-1	1
	0	1	1	-1	-1	2

In terms of mass the matrix of stoichiometric coefficients reads (rounded-off)

B =	16.0	70.9	-50.5	0	-36.5	1	
	0	70.9	50.5	-84.9	-36.5	2	♦

The source of the *k*-th species in Eq. (2.3-6) can then be expressed by the relation

$$z_k = - \sum_r B_{rk} x_r \quad , \quad k = 1, 2, \dots, K \quad (2.3-9)$$

where x_r is the so-called *extent* of the r -th chemical reaction giving how many times the reaction has run in the reactor in the given time interval (possibly unit of time).

Example 2.3-2: Chlorination of methane – Example 2.3-1 continued

E.g. for CH_4 (species No. 1) and CH_3Cl (species No. 3) the equations (2.3-9) in molar units are of the form

$$Z_1 = -X_1 \tag{2.3-10}$$

$$Z_3 = X_1 - X_2, \tag{2.3-11}$$

because CH_4 is present only in the first equation with coefficient 1 and CH_3Cl is in the first equation with coefficient 1 and in the second equation with coefficient -1 . So if the first equation runs once (with extent 1) then 1 unit of methane (e.g. kilomole) is consumed, while if both reactions run with extent one then the variable of CH_3Cl remains unchanged (the generation by the first reaction equals the consumption in the second one) ♦

2.4 Balance of energy

Before all, let us recall that we are distinguishing material streams and streams of energy. The energy associated with material streams consists of several items. Frequently, we can cope with the following kinds of energy in a material stream.

$$E = U + PV + E_{kin} + E_{pot} \tag{2.4-1}$$

where U is *internal energy*

PV 'pressure energy' (say; more precisely interpreted by thermodynamics)

E_{kin} kinetic energy

E_{pot} potential energy (in gravitational field).

The sum of the internal and pressure energies is another thermodynamic state variable: *enthalpy* denoted further as H

$$H = U + PV, \tag{2.4-2}$$

which usually replaces the first two terms of the right-hand side in Eq. (2.4-1).

In material streams, it is convenient to introduce *specific* values of energy related to the unit of mass of the stream. Thus for example the overall enthalpy of the stream i can be expressed as

$$H_i = m_i h_i \tag{2.4-3}$$

where h_i is the *specific* enthalpy per unit mass, m_i the flowrate.

In the same manner, the kinetic energy can be written as

$$E_{kin} = m \frac{1}{2} v^2 \quad (2.4-4)$$

Further kinds of energy (for instance the surface energy having some importance for tiny droplets of water present in wet steam) will further be regarded as negligible. The importance of individual items in the balance has to be judged according to the case, taking into account the given conditions (pressure, temperature, velocity of the stream and the size of the system).

Note: In the literature dealing with engineering thermodynamics is given that kinetic energy can be neglected for velocities up to 40 m/s. For this velocity the kinetic energy per unit mass equals 800 J/kg, which represents for example ca. 0.035 % of the evaporation heat of water at normal pressure. Anyway, it is quite easy to introduce kinetic energy into the balance in the RECON program. The user inputs only the equivalent diameter of the cross section through which water or steam flows. ♦

Besides material streams, the balance takes also into account the net energy streams mentioned above. Heat flows are currently denoted by the letter Q , mechanical work (power) by W .

The balance of energy for a given node in stationary state then generally reads:

$$\sum_{i,in} E_i + \sum_{j,in} Q_j + \sum_{k,in} W_k = \sum_{i,out} E_i + \sum_{j,out} Q_j + \sum_{k,out} W_k \quad (2.4-5)$$

Example 2.4-1: Energy balance of a turbine segment

Let us consider a turbine segment depicted in the following figure.

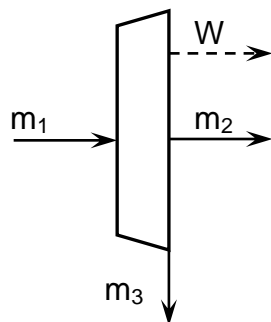


Fig. 2.4-1: Turbine segment

Steam enters as stream No.1, exerts work (stream W) and further advances (stream No.2) into the next segment, having been partially withdrawn as stream No.3. The mass balance at steady state reads

$$m_1 = m_2 + m_3 \quad (2.4-6)$$

where m_i are flowrates.

So long as we, for example, neglect kinetic and potential energy and heat losses into the environment, the energy balance is of the form

$$m_1 h_1 = m_2 h_2 + m_3 h_3 + W \quad (2.4-7)$$

where h_i are specific enthalpies of the streams

W is power transferred to the turbine shaft ♦

Setting-up the balance models 'manually' is considerably laborious and the user is then likely to commit a series of errors. For the application in practice, balancing programs have been formed that make this activity easier. One of them is RECON, which makes possible setting-up balancing models of quite complex systems relatively easily in the graphical editor.

2.5 Phase equilibria

Another type of exactly valid models that can be used in the data validation is that of phase equilibria. In practice, it is in particular the equilibrium between liquid water and steam at the saturation curve. This equilibrium can be expressed as the function representing the dependency of equilibrium vapor pressure P^* on temperature.

$$P^* = P(T) \quad (2.5-1)$$

There are a number of situations where the equilibrium can be assumed with a good approximation, such as

- production of steam in a steam generator
- transport of wet steam in pipeline
- release of overheated liquid to lower pressure (expanders)
- steam saturators.

In such cases, if pressure and temperature are measured simultaneously in one place then these measurements can be reconciled on the basis of the phase equilibrium.

2.6 Enthalpies of streams

For the models in the power industry, of basic importance is to know the enthalpy of the streams, which plays the main role in the energy balances. Because one deals here with a variable that cannot be directly measured in practice, one can speak of its modeling, which represents a relevant part of engineering thermodynamics. We must rely upon the results of works performed by generations before us.

We must naturally also rely upon the program RECON where these models are intrinsically programmed. RECON contains, in essence, three models for computing the enthalpy.

1. the well-known database for water and steam properties IAPWS IF-97, which nowadays represents the worldwide standard
2. Berghoff's model for computing the enthalpies of crude oil and its fractions
3. general polynomial model for enthalpy

The last two models must be configured by the user (inserting the proper values of the parameters).

For the calculations in power industry, the first variant is essential. In computing the enthalpies, RECON makes use of altogether 5 functions. Besides the thermodynamic state variables T (temperature) and P (pressure) there also occurs (in systems where water and steam are in equilibrium) the so-called *wetness* X , i.e. the mass fraction (%) of liquid water in the mixture with steam. The value $X = 0$ means dry steam and $X = 100$ means liquid water.

1. $H_2O(P,X)$ – enthalpy of equilibrium steam-water mixture at pressure P and wetness X
2. $H_2O(T,P)$ – enthalpy of water or steam defined by temperature and pressure. The program itself finds, if one deals with liquid or steam
3. $H_2O(T,X)$ – enthalpy of equilibrium steam-water mixture at temperature T and wetness X
4. $H_2OL(T,P)$ – enthalpy of liquid water defined by temperature and pressure. If temperature and pressure don't correspond to the liquid phase, the nearest values of the variables corresponding to liquid phase are chosen
5. $H_2OV(T,P)$ – enthalpy of water vapor defined by temperature and pressure. If temperature and pressure don't correspond to the vapor phase, the nearest values of the variables corresponding to vapor phase are chosen.

The last two variants are chosen when we are sure that one deals with the given phase and the state is defined by temperature and pressure. It can, indeed, happen that due to temperature or pressure measurement errors, possibly during the computation, this computation 'slides' into a state to which corresponds the wrong phase. This would then cause a great jump change in the enthalpy and the computation proper would meet with serious problems. In such cases, the functions No. 4 and 5 replace the function No. 2.

Concluding this section, let us in addition mention one anomaly in the system water-steam. In Fig. 2.6-1, we have the so-called $p-i$ diagram for this system, which is the dependence of water vapor pressure on its enthalpy.

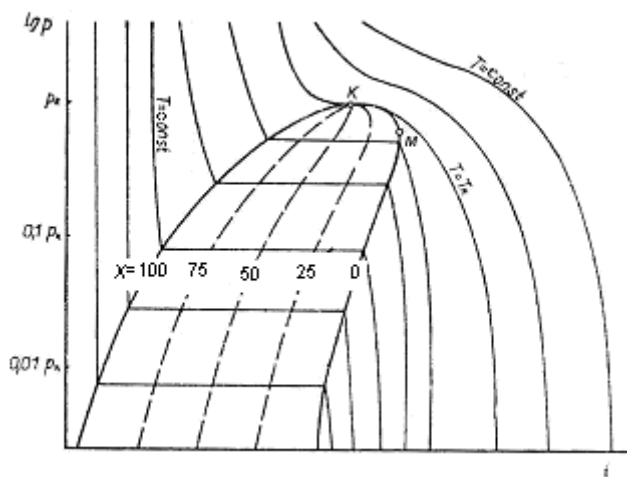


Fig. 2.6-1: p - i diagram

The details about this diagram can be found in textbooks on engineering thermodynamics, (a similar diagram could also be set up for temperature). Here, also drawn are curves of constant temperature T and steam wetness X ($X = 0$ means dry steam, $X = 100$ means water at boiling point).

In the diagram, two relevant points are marked. K represents the well-known critical point. M represents the state where the vapor phase at equilibrium has, as dependent on pressure (or also temperature) the maximum specific enthalpy. This point is characterized by the pressure ca. 3.0 MPa, resp. temperature 234°C. The consequence is that in the neighborhood of this point, the enthalpy of saturated vapor is not a unique function of pressure and the dependency of enthalpy on the state variables is quite flat (as the reader can easily verify in tables of steam properties). This fact is important in the balance calculations, as will be shown later.

2.7 Momentum balance

In physics, the momentum is defined as product of mass and its velocity vector. In problems of fluid flow in a pipeline, the momentum balance means the mathematical model of how the kinetic, potential and pressure energies are interrelated and how they are dissipated due to friction. A brief description of this model can be found in Appendix 5, possibly also in specialized monographs.

In practice, the case of interest is for us conveying the fluid through a pipeline of constant cross-section between two points characterized by pressure and height. Although the model of momentum conservation described in Appendix 5 holds true, under the given assumptions exactly, it contains an empirical parameter – the so-called *roughness of the pipeline*. The roughness must be in practice identified on the basis of the real behavior of the system considered.

The model of momentum for pipeline flow generates one equation, which is just sufficient for the identification of the roughness. This then offers the possibility of using this roughness as a constant of the model. The question is, whether the roughness can be considered constant (consider corrosion, fouling of the pipeline etc.). In the latter case the model of momentum conservation has no longer absolute validity, but it can be used for the detection of changes that occur in the course of time in any system. The justification of applying the momentum equation to the data validation is disputable and requires always a detailed analysis.

2.8 User defined variables and equations

When working with the RECON program, in most cases we can cope with the problem definition in the graphical medium. Besides the graphically featured nodes and streams, one can here also define further variables:

- temperatures
- pressures
- wetnesses
- pipeline roughnesses

All these variables can occur as measured (given), unmeasured (with the possibility of being computed from the model) or fixed (constants).

For all these variables, RECON also automatically creates the set of equations of the mathematical model. In practice however, there arises the need to enrich these possibilities by *user defined variables and equations*.

The (in RECON called:) *auxiliary variables* are most frequently for example

- specific consumptions
- process efficiencies
- phase equilibria
- measured variables compensated for state conditions
- dividing ratios of the streams
- parameters of more complex mathematical models.

Having in view their heterogeneity, it would not be productive to define these variables in the standard manner, i.e. as further groups of variables. So there exists here a general group of variables, which can again be of arbitrary types (measured, unmeasured or fixed). Substantial for these variables is the fact that they are not present in the automatically generated equations. They can thus play a role in user defined equations only.

The user defined equations give thus the user a chance to accommodate the model according to his needs.

In creating the user defined equations there must always be guaranteed that the new equations introduced by the user are 'sound' and will not adversely affect the solution. Indeed, each equation in DR forms a bond between the variables and if the equation is 'bad', it can unfavorably influence the quality of the results.

The motivation to introducing user defined equations can be different. Justified are the cases where the graphical editor in a special situation does not allow us to create the model, even if exactly valid. An example is the model of phase equilibrium dealt with in the preceding section; the latter can only be created in the editor of user defined equations.

A further justified case is introducing the equations that define *auxiliary unmeasured variables*. For example if we have several steam generators in the system, it is useful to create a new variable that is sum of thermal performances of all the steam generators: And in order to define this variable, just one equation is needed. From the standpoint of the degree of redundancy (see Section 3.3), nothing has changed because with one new unknown, one new equation was added.

Further examples need not be, however, so clear. Generally, it can be recommended to avoid the presence of unobservable variables (see Section 3.2). This requirement can be solved just by completing the model by user defined equations. Of course these equations can be of different quality, from the quite justified ones to those which can be ranged into the category of the author's 'pious desires'. If, on creating a new equation the value of Q_{min} (see Section 3.6) has not changed, the new equation has only contributed to enhancing the observability. If, however, the value of Q_{min} has increased then the new equation has already entered the process of reconciliation and must be scrutinized in detail as to its justification.

While, in creating the model in the graphical editor, RECON takes care of whether no *dependent equations* arise, it can happen that the user still creates an equation dependent on the already existing ones. The consequences are not fatal – the program simply discards this equation. The user is informed about it in the summary of results: the whole number of equations differs from that of the independent ones.. To be safe, it is recommended to repeat the computation after introducing each new user defined equation and examine possible consequences.

Note: When creating user defined equations in the RECON program, the following fact is to be taken into consideration. The values of the state variables (flowrates, temperatures, pressures etc.) present in the equation are inserted by RECON in units of the SI system. Let us for example consider an auxiliary variable defined by the user's equation as the sum of powers of several (net) energy streams. The result for this variable will be in the system SI, i.e. in W, even if *individual* powers for the respective streams are, in the user's system, e.g. in MW (and as such also further presented for him). If we require to have also this *new* power in MW, the necessary re-calculation must be prescribed already with the user defined equation ♦

2.9 Model of an industrial system

So far, we have only dealt with balancing models of one node, contingently with hydraulic calculations of one stream. The model of a production system is formed by nodes interconnected by streams. A simple example is the balance system of a steam generator created in the graphical editor of the RECON program.

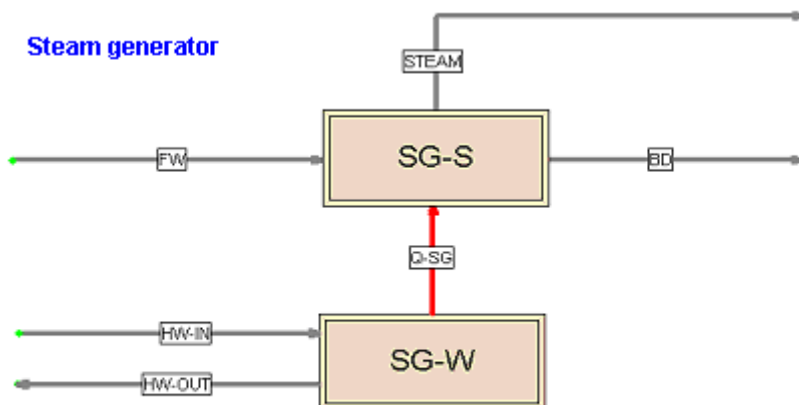


Fig. 2.9-1: Balance scheme of a steam generator

The balance scheme of steam generator in a nuclear power plant consists of two nodes. Node SG-W represents the tube space of the generator, where hot water (HW) flows. The other node is the steam space SG-S where steam (STEAM) is

generated from the feed water (FW). Further stream leaving the system is BD (Blow Down). Both nodes (isolated as for the mass flow) are connected by (net) energy stream Q-SG representing heat flow from hot water into the steam space.

As we'll see in the chapter dealing with typical operations in the power industry, around each node the program RECON can generate two equations – the mass and energy balances. The result of the reconciliation is reconciled (consistent) balances of both the nodes for mass as well as energy.

The information on how individual nodes are connected by streams is generally expressed by so-called *incidence matrix A*, the elements of which have values 0, 1 or -1 with the following interpretation:

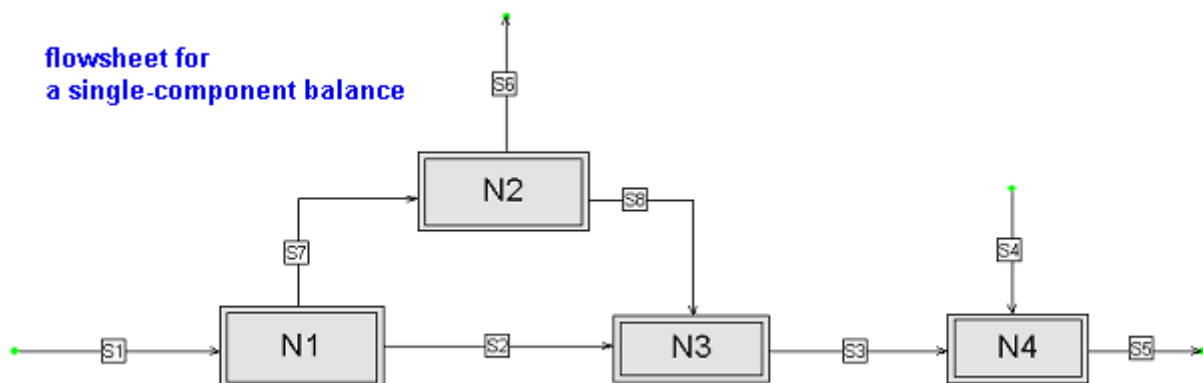
$A_{ij} = 1$ if the j -th stream enters the i -th node

$A_{ij} = -1$ if the j -th stream leaves the i -th node

$A_{ij} = 0$ if the j -th stream is not incident with the i -th node

Example 2.9-1: Incidence matrix of a process flowsheet

Let us consider the flowsheet consisting of 4 nodes and 8 streams



The incidence matrix of this scheme is

No. of stream	1	2	3	4	5	6	7	8	No. of node
A	1	-1	0	0	0	0	-1	0	1
=	0	0	0	0	0	-1	1	-1	2
	0	1	-1	0	0	0	0	1	3
	0	0	1	1	-1	0	0	0	4

E.g. in the second column of the matrix (which belongs to stream No. 2), we can see that stream No. 2 leaves node No. 1 and enters node No. 3. In a corresponding manner, in the

first row of the matrix (belonging to node No. 1) we can see that stream No. 1 enters and streams No. 2 and 7 leave it. ♦

In an analogous way (even if more compressed), the information about the structure of process schemes is stored in the computer memory.

If \mathbf{m} represents the column vector of the corresponding mass flowrates m_j ($j = 1, 2, \dots, J$), the mathematical model of mass balance can be briefly written as the vector-matrix equation

$$\mathbf{A} \mathbf{m} = \mathbf{0} \quad . \quad (2.9-1)$$

or in form of vector components for the i -th node (summation over j)

$$\sum A_{ij} m_j = 0 \quad \text{for } i = 1, 2, \dots, I \quad . \quad (2.9-2)$$

Specifically in Example 2.9-1 the balance of node No. 1 reads

$$m_1 - m_2 - m_7 = 0 \quad . \quad (2.9-3)$$

Compare with the one-node balance in Section 2.3 – Eq. (2.3-1).

In the majority of cases we require that the properties of any stream leaving one node and those of the same stream entering the other node be also the same. This concerns the total mass of the stream as well as its composition and energy. With this hypothesis, from the theory of balancing then follows that if all node balances are reconciled then also the balance of the whole system is reconciled (the difference inputs minus outputs equals zero).

The only two exception from this rule are the momentum balances and heat exchangers, where the input and output states may differ.

2.10 Model errors

The results of DR are sensitive to model errors. If the model does not truly describe the reality, the results can be misleading. This holds of course generally for arbitrary exploitation of the models, technological calculations etc.

In the domain of mass and energy balancing there is no doubt about the validity of the balances; the problem is rather in the neglect of certain particular phenomena.

In the balancing, mainly two problems can play a role.

- neglecting the accumulation (unjustified stationarity hypothesis)
- material or energy losses into the environment

If the deviations from the assumptions cannot be neglected, they must be built into the model.

In the nonstationary case this means creating the dynamic balancing model, in the simplest case at least with mass accumulation in certain nodes.

Example 2.10-1: Unsteady-state balance

Let us consider a tank, into which one mass stream of water m_1 continuously flows, while the second stream m_2 continuously flows out. The tank is endowed with a level meter. In the control system, the level is re-calculated to the total mass of water in the tank. The scheme of measurement is in Fig. 2.10-1 a.

The balance scheme proper is in Fig. 2.10-1b.

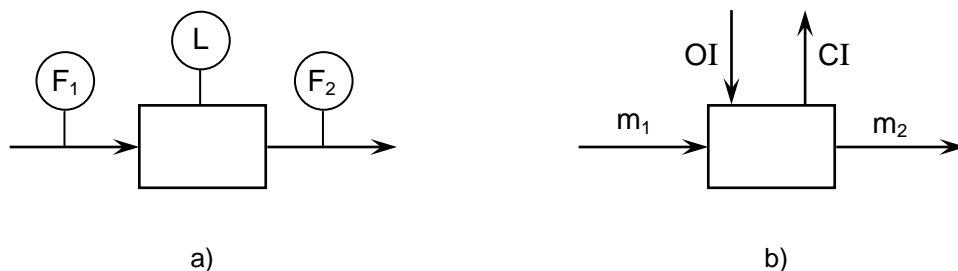


Fig. 2.10-1: Balance of a tank in unsteady state

There are here two new fictitious streams – OI (*Opening Inventory*) and CI (*Closing Inventory*).

The balance equation is then of the form

$$m_1 + OI = m_2 + CI \quad (2.10-1)$$

It is important that the balance equation holds for a defined balance period and the variables m_1 and m_2 are integrals of the flowrates having the physical dimension of mass (not that of mass flowrate). The values OI and CI are the states of mass (inventory) at the beginning and at the end of the balancing period. ♦

The losses into the environment are mainly heat losses. A relatively simple solution is introducing the loss streams into the model proper. Most often and with sufficient precision, the losses can be estimated or measured and then given as fixed values. Consider that if the losses are small (though not negligible), it is sufficient to estimate them with a precision of tens per cent (and even then, the absolute error will be small).

Let us further draw the reader's attention to the fact that in many cases, it is **not expedient** to compute the energy losses into the environment from the balance (inputs minus outputs). If the losses are small (several units per cent of the input energy), as a result of measurement errors one may easily come to the conclusion that energy is not lost but gained.

It holds generally that the mentioned effects can be neglected when they are substantially smaller than the error in the main balance streams.

Let us for example consider a tank ranged in the flow of condensate (according to Fig. 2.10-1). The balancing interval is 1 hour and the mean flowrate 1000 t/h. Let the measurement uncertainty be 3 % of the measured flowrate. The uncertainty in the throughput is thus 30 t. The level fluctuation in the tank thus can certainly be tolerated if it represents less than 3 t (which is 1/10 of the flowrate measurement error). In practice, one could be even considerably more benevolent. The uncertainty proper in the flowrate measurement error is certainly more than 10 %.

In using the phase equilibria, a possible problem is the existence of metastable regions (subcooled vapor or overheated liquid). This can be a menace mainly with rapid phenomena where the equilibrium need not be installed instantaneously (in fact for example during the condensation, due to the surface tension and curvature, at given pressure the equilibrium temperature near the surface of the arising small droplets [nuclei] is lower than the standard equilibrium value – this even is the metastability). Generally, the problem is quite complicated and cannot be discussed here in detail. The best way is starting from real data and verify if they are, in long term, compatible with the phase equilibrium model. If it is not the case, the best solution is to discard the contradictory parts of the model. Sometimes, it is possible to take advantage of the fact that certain subsystems occur in the scheme repeatedly (e.g. the existence of several practically identical steam generators in one power plant). If the model works properly in the majority of the identical (macro)units and in certain cases doesn't, the error is to be sought rather in concrete data.

3 Statistical treatment of measured data

The validation of data is based on the confrontation of the measured data with generally valid laws expressed in the form of mathematical models. In this context, mathematical statistics and the theory of measurement errors is of considerable importance.

3.1 Measurement errors

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

$$x^+ = x + e \quad (3.1-1)$$

where x^+ is the measured value

x is the true (unknown) value

e is the measurement error

Given one set of data, the errors proper are further divided into *random* and *gross*. The random errors are an unavoidable part of any measurement; they are characterized by standard deviation σ_1 . Gross errors are remarkably greater than random errors and if the whole measurement has not to be disvalued, they must be eliminated from the measured data set. If we have several sets of data (time series) and the error is deterministically repeated in time, we speak of a *systematic error*.

Information about possible errors of measurement is indispensable for the data reconciliation to be discussed later. Sometimes, one speaks of the *uncertainty* in the result of measurement, basically identical with so-called *maximum error* that is assumed for the result. With measurement by instruments, there is often given the so-called *tolerance*, which again has the character of maximum error assumed at the measurement (or guaranteed by the producer). The tolerance expressed in per cents of the instrument's range is called *class of accuracy*. For example class of accuracy 2 means maximum error $\pm 2\%$ of the instrument's range.

Example 3.1-1: The producer of the flowmeter declares that the maximum error in the flowrate will not be greater than 1% from the range of the instrument. One can thus suppose that the (unknown) true value x occurs in the interval

$$x \in \langle x^+ - e_{\max}; x^+ + e_{\max} \rangle \quad (3.1-2)$$

where e_{\max} is the maximum error.

In the statistical treatment of data, one sometimes requires the knowledge of so-called *standard deviation* σ of the measured value. In practice, we here take half the maximum error

$$\sigma = 1/2 e_{\max} \quad (3.1-3) \blacklozenge$$

More details about the measurement errors are given in Appendix 2.

3.2 Classification of variables

The physical variables occurring in the problem can be classified from different points of view. According to whether it is possible at all to measure the variable by some feasible method, the variables are divided into *directly measurable* and *directly unmeasurable*.

E.g. for the flowrate measurement of feed introduced into the distillation column, there is a number of possibilities. They differ by the easiness of installation, precision, price and further properties. So one deals with a directly measurable variable. On the other hand let us imagine a turbine segment with outlet steam that enters the following segment. We here hardly can imagine a method which would make possible to measure directly the steam flowrate. This variable is thus directly unmeasurable (but it can be found by computation; we then speak of an *indirectly measured* variable, thus identifiable by computation from other directly measured variables). If a directly unmeasurable variable has the character of a constant, it is called *parameter of the model*.

Another viewpoint is whether the measurement has been realized in the concrete case. The variables are then divided into *measured* and *unmeasured*. As to the measured variables, an important property of the measured value is also its *uncertainty*, which limits the magnitude of possible random error. In certain cases, the measurement is (perhaps by convention) regarded as errorless (with null error) – let us imagine for example a very precise custody – transfer flowmeter which serves as the basis of accounting between two companies.

Further classification of variables is simultaneously based on the division of the variables into measured and unmeasured, and on the properties of the mathematical model binding them together. It will be clarified with the aid of the following example.

Example 3.2-1: Classification of variables

In Fig. 3.2-1, we have a simple mass balance scheme. The full arrows denote measured streams, the dashed ones unmeasured streams.

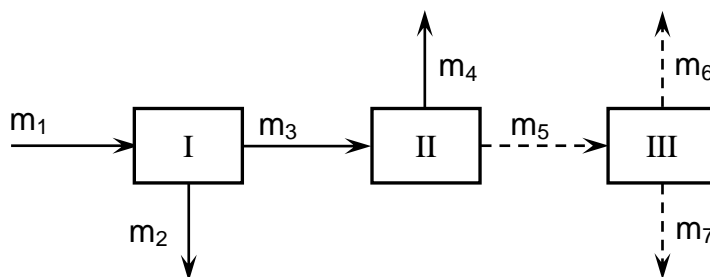


Fig. 3.2-1: Mass balance scheme

For this system, altogether 3 balance equations with 6 flowrates can be written.

$$m_1 = m_2 + m_3 \quad (3.2-1)$$

$$m_3 = m_4 + m_5 \quad (3.2-2)$$

$$m_5 = m_6 + m_7 \quad (3.2-3)$$

We can see here several characteristic situations.

All the variables around node I are measured. One has to do with so-called *redundant* variables. One of them could be computed from the latter two. Further consequence of this situation is the fact that due to measurement errors, with the measured values the equation will not be exactly satisfied. We are thus led to the necessity of *data reconciliation*. The measured flowrates around the node must further be adjusted, i.e. *reconciled*, in order that this balance equation be obeyed. This holds for redundant measurements generally.

From the second equation, we can compute the magnitude of stream No.5. This unmeasured variable is called *observable* (it can be derived from the measured ones).

In the third equation, we already know the variable m_5 ; however, it does not suffice for the computation of flowrates m_6 and m_7 . Indeed, we here have one equation for two unknowns. These unmeasured variables cannot be uniquely computed from the model and we call them *unobservable*♦

In addition to the variables measured with some error and the unmeasured ones, we sometimes introduce also a special kind of measured values that we call *errorless*. They are thus a priori known variables, often obtained by very precise measurement where the error can be neglected. 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 also called *fixed*. The whole situation is depicted in the following figure.

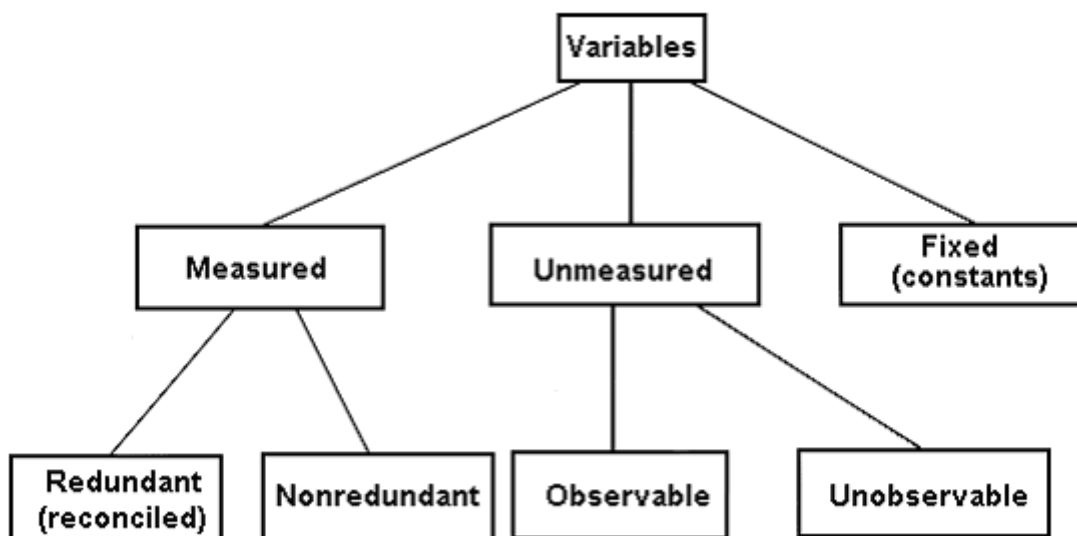


Fig. 3.2-2: Classification of variables

Further properties of the variables follow from the results of reconciliation. The redundant ones are then adjusted and they are thus also called *adjustable*. The remaining ones are called *nonadjustable*.

The example shows that 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 the measurement is absent and certain variables are unobservable.

Generally, the user is to be recommended to avoid, when possible, the models with unobservable variables. Otherwise, in the latter case one can meet with a number of theoretical problems which, alas, manifest themselves also in practice. Basically, there are two ways how to attain full observability of all the unmeasured variables.

- simplify (reduce) the system in places where the measurement is insufficient (in particular by merging several balancing nodes)
- complete the model by user defined equations.

In general, the classification of variables is not an easy task. In the Appendix 6 there is described the method based on the solvability analysis of a system of equations which is used in RECON.

3.3 Degree of redundancy

In the above example (Fig. 3.2.-1), there was one node around which all the flowrates were measured. If we ranged arbitrary one of these flowrates among the unmeasured ones (deleted the respective measurement), this flowrate could be computed from the balance equation. In this case, it holds that we here have just one degree of redundancy. For the mass balance of a system, assuming that all flowrates are observable, the degree of redundancy v is found according to the simple relation

$$v = M - N \quad (3.3-1)$$

where M is the number of equations

N is the number of unmeasured variables.

Note: If this formula were applied to the above example, one would obtain a wrong result due to the presence of unobservable variables. For more complex systems, it is not so easy to identify the degree of redundancy and it is necessary to rely upon the program RECON ♦

It holds true generally that the degree of redundancy is a measure of the overall excess in the measurement places of the whole system. Its high value indicates good chances for the data validation. It is also to be noted that **positive redundancy is a necessary condition for the data reconciliation and validation**. If the redundancy is null, the data validation is not possible.

3.4 Data reconciliation

Let us start from the mathematical model

$$F(x,y,c) = 0 \quad (3.4-1)$$

where $F()$ is the vector of model equations

x is the vector of directly measured variables

y is the vector of directly unmeasured variables

c is the vector of precisely known constants

Eq. (3.4-1) holds for the true (unknown) values of the variables. If we replace them by the measured values x^+ , the equations need not (and most likely will not) be exactly satisfied:

$$F(x^+,y,c) \neq 0 \quad (3.4-2)$$

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

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

$$x_i' = x_i^+ + v_i \quad , \quad (3.4-3)$$

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 (method of *maximum likelihood*) [1,2,3,4].

The adjustments must satisfy two fundamental conditions:

1) The reconciled values obey Eq. (3.4-1) – we say that they are consistent with the model

$$F(x',y',c) = 0 \quad (3.4-4)$$

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

$$\text{minimize} \quad \sum (v_i/\sigma_i)^2 = \sum ((x_i' - x_i^+)/\sigma_i)^2. \quad (3.4-5a)$$

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

The least squares function (3.4-5a) is used in the case of uncorrelated (statistically independent) errors (see Appendix 2). In the case of correlated errors a more general criterion is minimized:

$$\text{minimize} \quad \mathbf{v}^T \mathbf{F}^{-1} \mathbf{v} \quad (3.4-5b)$$

where \mathbf{v} is vector of adjustments and \mathbf{F} is covariance matrix.

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

The mathematics of the solution itself was, in the last decades many times described in the literature (e.g. [1,2,3,4]) and will not be mentioned in the sequel. See also the *Balancing and Data Reconciliation Minibook* [5], which is free of charge at hand to be downloaded from internet.

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

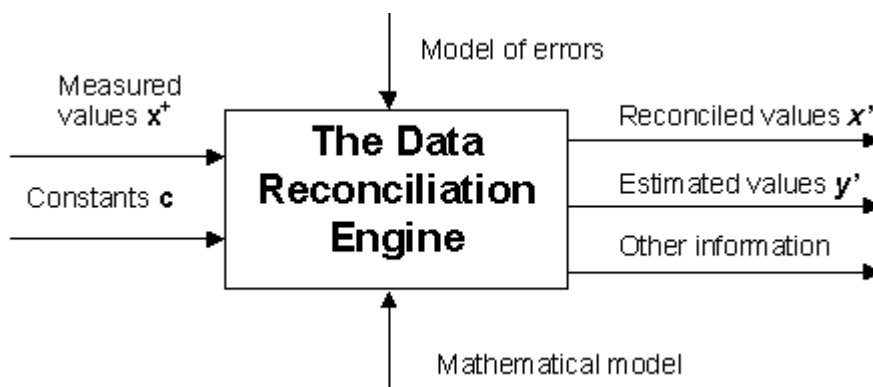


Fig. 3.4-1: The Data Reconciliation Engine

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

We can then write symbolically

$$\mathbf{x}' = \mathbf{h}_1(\mathbf{x}^+) \quad (3.4-6)$$

$$\mathbf{y}' = \mathbf{h}_2(\mathbf{x}^+) \quad (3.4-7)$$

Example 3.4-1: Reconciliation of mass balance

Let us consider the scheme according to the following figure.

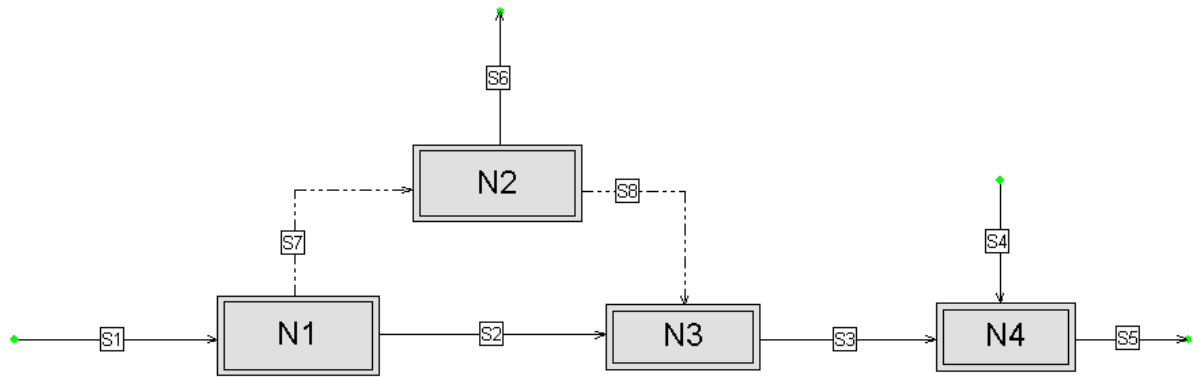


Fig. 3.4-2: Mass balance scheme created in RECON

The scheme consists of four nodes and eight streams. The model contains 4 linear equations for 8 streams. The input data and results of the reconciliation are given in the following table.

Table 3.4-1

Stream	Type	Measured value (kg/h)	Measurement uncertainty (%)	Reconciled value
S1	M	100.1	2	99.29
S2	M	41.1	4	41.10
S3	M	79.0	2	79.36
S4	M	30.6	10	30.05
S5	M	108.3	4	109.41
S6	M	19.8	4	19.93
S7	N	10.0	-	58.19
S8	N	10.0	-	38.26

At the type of measurement, M means masured, N unmeasured (nonmeasured) value. The measurement uncertainty (maximum error) is expressed in per cents of the measured value.

The 'measured value' for unmeasured variables represents so-called *first guess* required by the RECON program so as to initiate the computation. It cannot be told generally, how close to the truth it shall be. It ought not to be a nonsense, but rather a technical estimate that the practician improvise without long thinking. Theoretically, the estimated value should not influence the final value. However a good estimate facilitates and makes faster the computation, and minimizes the risk of possible nonconvergence. For reasons not to be analyzed here, RECON does not accept as first estimate the value 0 (zero). ♦

3.5 Benefits of the reconciled data

They are several.

1. Reconciled data are in agreement with laws of nature. In addition to higher credibility, they are also acceptable in accounting, where disproportions in data cause significant problems.
2. Reconciled data are more precise in the statistical sense, if compared with the measured ones. The enhanced precision can be quantified with the aid of the standard deviation of the reconciled value, which is always smaller (or, exceptionally, at least not greater) than the standard deviation of the measurement error.

$$\sigma_{x'} < \sigma_{x^+} \quad (3.5-1)$$

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

$$a = 1 - \sigma_{x'} / \sigma_{x^+} \quad (3.5-2)$$

The adjustability characterizes the reduction of the standard deviation thus also that of the uncertainty in the result, if compared with the primary measurement. If for example the adjustability of the reconciled value is 0.5, the uncertainty has been reduced by half. Adjustability 0.25 means reducing the uncertainty by a quarter, and so on. The greater the adjustability is, the greater is also the reduction of the uncertainty.

There is generally a relation between adjustability and redundancy. With increasing redundancy, also adjustability increases. Of null adjustability are *nonadjustable (just determined)* variables. As follows from their name, the latter variables remain unchanged by the reconciliation and are not influenced by other measured variables.

Example 3.5-1: Enhanced precision of the results

The following table summarizes the enhancement of the results' precision as a consequence of the reconciliation. The example is a continuation of the preceding example 3.4-1 – see Fig. 3.4-2.

Table 3.5-1

Stream	Measurement uncertainty (kg/h)	Uncertainty of reconciled values (kg/h)	Adjustability
S1	2.02	1.30	0.35
S2	1.64	1.64	0.00
S3	1.58	1.24	0.22
S4	3.06	2.53	0.17
S5	4.33	2.63	0.39
S6	0.79	0.76	0.05

One can see that the adjustabilities are quite different and range from zero value for stream S2 (nonadjustable variable) up to the highest value at stream S5 ♦

3. **Detection and elimination of gross errors.** A possible gross error among the measured variables would, as a consequence of the reconciliation, spread and corrupt also other measured and reconciled variables. So long as DR is applied, one must therefore find a systematic solution for safeguarding the data against gross errors. One of the results of the reconciliation is thus an indication that signals the (possible) presence of a gross error. In such a case the results cannot be exploited, until gross error(s) is (are) identified and eliminated. The systematic use of DR guarantees, in this manner, the protection of final results against gross errors.

Besides the methodology itself for gross errors presence detection, also effective methods for the localization of gross error sources have been elaborated in the frame of DR. Automatically, a small group of suspected variables is found to be further scrutinized.

4. **Uncertainties of results.** One of the side products of DR is also information on uncertainties as well in the reconciled as in the observable directly unmeasured values. This information is important not only when judging the overall quality of the results, but for example also when comparing the results of various process tests and the like.

5. **Propagation of errors and optimization of the systems of measurement.** In the frame of DR, one can also trace the propagation of errors in the primary variables during the data processing. One can thus find, which of the primary variables are the key ones from the viewpoint of the target values precision (the latter are those, whose identification is the goal of the whole measurement). It is then relatively easy to find the bottlenecks for the measurement system precision and to optimize the system.

The elimination of gross errors plays the key role in data validation. In essence, it consists of three steps.

- **detection**, i.e. their presence indication
- **identification**, i.e. finding their sources
- **elimination** of these errors.

Let us now successively go through the named steps.

3.6 Detection of gross errors

The process data treatment could be called 'struggle against gross errors of measurement'. As already mentioned above, a single gross error can, as a consequence of the reconciliation, propagate and corrupt the whole set of data. If these errors did not exist, working with data would be a delightful activity. Alas, this is not the case and therefore, from the very beginning of the industrial DR application, this problem has been paid much attention; successively, several effective methods were elaborated for the detection of gross error(s) presence.

Let us recall once more that gross measurement error means an error that is greater than the uncertainty in the given measurement. Let us start from the weighted sum of squares of adjustments defined in Eq. (3.4-5a); it will be denoted by Q_{min} .

$$Q_{min} = \sum (v_i/\sigma_i)^2 \quad (3.6-1)$$

Clearly, if there were no errors present then no adjustments would be necessary and Q_{min} would equal zero. If only random errors are present with Gauss (normal) distribution, one can show that Q_{min} is a random variable with so-called χ^2 -distribution and ν *degrees of freedom* (called so in statistics). This ν is here equal to the degree of redundancy introduced in Section 3.3 and denoted by the same letter. If only random errors are present in the frame of our hypotheses about the precision (uncertainty) of individual measurements, there must hold with probability $(1-\alpha)$:

$$Q_{min} < \chi^2_{(1-\alpha)}(\nu) \quad (3.6-2)$$

where $\chi^2_{(1-\alpha)}(\nu)$ is critical value of the χ^2 -distribution with ν degrees of freedom. This value will further be called Q_{crit} .

The value of α , so-called *level of significance*, is in technological practice usually taken as 0.05 i.e. 5 %. The Q_{crit} for different values of α and ν can be found in statistical tables, but its values are also at hand in the frame of the RECON program.

So far about the statistical properties of the variable Q_{min} in the presence of random errors only. If one or more gross errors occur in the measured data, it is likely that the adjustments must be greater and such must also be the value of Q_{min} . This is also the basis for testing the gross error presence hypothesis (in fact the gross error absence hypothesis is tested). In order to brush up the knowledge about testing statistical hypotheses, Appendix 3 is enclosed.

The essence consists in the fact that if the inequality (3.6-2) is not satisfied thus (omitting the equality) if

$$Q_{min} > Q_{crit} \quad , \quad (3.6-3)$$

the (possible) presence of a gross error has been **detected**. Further steps are its **identification** and **elimination**.

Example 3.6-1: Gross error detection

This example is continuation of the preceding Example 3.4-1 – see Fig. 3.4.-2.

Let us first summarize the information from the preceding example. The model is formed by four equations in eight variables. Altogether two variables are unmeasured. Both the unmeasured variables are observable thus there are two degrees of redundancy in the problem. The critical value of the chi-square distribution for the significance level 0.05 is 5.97. The Q_{min} -value found in the original example was 1.308. If put into the inequality (3.6-3) we have

$$1.308 < 5.97 \quad (3.6-4)$$

and no gross error is detected.

Let us now introduce a gross error into the data, say +10 kg/h at stream S1. The data processed by RECON are in the following table.

Table 3.6-1

Stream	Type	Measured value (kg/h)	Measurement uncertainty (%)	Reconciled value (kg/h)	Reconciled value without a gross error
S1	M	110.1	2	102.98	99.29
S2	M	41.1	4	41.1	41.10
S3	M	79.0	2	82.26	79.36
S4	M	30.6	10	29.08	30.05
S5	M	108.3	4	111.34	109.41
S6	M	19.8	4	20.72	19.93
S7	N	10.0	-	61.88	58.19
S8	N	10.0	-	41.16	38.26

Further information:

Degree of redundancy $\nu = 2$

$Q_{min} = 64.54$

$Q_{crit} = 5.97$

On substituting in Ineq. (3.6-3) we can see that gross error has been detected. For comparison, in the last column of the table are given the reconciled values for data without gross error, taken over from Example 3.4-1. One can see that a single gross error has been, indeed, 'diffused' among the other reconciled values. The only exception is the nonadjustable stream S2. ♦

3.7 Identification and elimination of gross errors

The number of measured variables in real balancing models is usually considerable and can be several hundreds. Screening all meters 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 the source of gross error, it is sufficient to find a small group of suspected variables upon which one can concentrate the attention.

The simplest method is to **examine the balance differences** around individual nodes. The balance difference Δ for mass or energy balance is defined as

$$\Delta = \text{sum of inputs} - \text{sum of outputs.} \quad (3.7-1)$$

It can also be expressed relatively in per cents from the (arithmetic) mean flowrate around the node

$$\Delta_{rel} = (2\Delta/(\text{sum of inputs} + \text{sum of outputs})) * 100 \quad (3.7-2)$$

If the absolute value of Δ_{rel} 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.

The weak point of this method consists in the frequent occurrence of unmeasured variables among the inputs and outputs. The method works even with estimates of the unmeasured variables, but its reliability is then smaller. Still, this method is recommendable because of its simplicity and easy comprehensibility. In any case, it is very convenient for the creation of models. Indeed, it makes possible to find not only gross measurement errors, but also mistakes at the model creation. The program RECON enables the user to easily create the survey of all balance differences.

Considerably more sophisticated is the method of **normalized adjustments**. It starts from the idea that the adjustments are in fact estimates of errors. Any adjustment v is a random variable, which has its standard deviation σ_v . The normalized adjustment u is defined by the relation

$$u = v/\sigma_v \quad (3.7-3)$$

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 suspected variables, we can go further. A suspected variable can be scrutinized on ranging it among the unmeasured ones and carry out the data reconciliation. If then no gross error is detected, this variable could be the source of the gross error. By the ranging of 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.

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.

Example 3.7-1: Identification of a gross error

This example is continuation of the preceding example 3.4-1 – see also Fig.3.4-2.

Let us begin with the method of **normalized adjustments**. In the following table, there are given the normalized adjustments for the measured streams. Only normalized adjustments greater than 1.96 are shown, where 1.96 is the critical value for the normal distribution.

Table 3.7-1

Stream	Type	Normalized adjustment
S1	M	-8.02
S3	M	6.81
S6	M	8.02

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 us further continue according to the method of **suspected variables elimination**. Successively, individual streams are ranged among unmeasured and reconciliation is carried out. The results are given in the following table.

Table 3.7-2

Stream	Type	Q_{min}	Q_{crit}
S1	N	0.21	3.84
S3	N	18.15	3.84
S6	N	0.21	3.84

The individual streams are now unmeasured. Let us first take note of the last column of the table. The critical value is now smaller than in the preceding case. Indeed, including one variable in the unmeasured ones the degree of redundancy has decreased by one. The important thing is that only at stream S3, Q_{min} remains greater than the critical value Q_{crit} and gross error is thus still detected. In other words, including S3 in the unmeasured set did not help so this stream can be immediately discarded from the list of the suspects.

There thus remain two suspected streams S1 and S6. Both methods of localization rate them equally and what remains is the classical procedure, for example

- revision of the individual flowmeters
- tracing the trends of the variables at the time when the problem has arisen
- judge the measured values also from other points of view than those from which they are treated by the model proper (physical meaning of the variables, further information not included in the model and data processed) ♦

Let us finally state that although the above described methods 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 gross errors 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 the vendors of data validation software).

3.8 Efficiency of gross measurement errors detection

The following section is addressed to those who are more deeply interested in the DR. At the first reading, the section can be skipped. We'll further confine ourselves to the problems of gross **measurement** errors, where a consistent theory is available. The cases of gross **model** errors are more complicated and must be solved according to the case.

One often meets with the idea that DR automatically protects the results against all possible gross errors. Although this is usually the main argument of the DR software vendors, these ideas have to be put right.

The possibility of gross errors detection is limited by the redundancy of the data. As shown above, in a real process scheme there can exist simultaneously places with high redundancy (with many measurements) and places with few measurements where the redundancy is null or even unobservable unmeasured variables are present. It is thus legitimate to suppose that the possibilities of gross error detection for individual variables will differ from case to case. The whole topic is elaborated in details in Appendices 3 and 4 where the reader can also find the explanation of certain statistical concepts made use of in the sequel.

The probability that a gross error will be detected is not a general property of the given model, but it must be stated for any measured variable separately. The probability depends on three factors.

1. On the redundancy of the given variable expressed via its adjustability (see Eq. (3.5-2)): The higher the adjustability is, the easier is the detection of the gross error for this variable. On the other hand the probability of gross error detection for a *nonadjustable* variable equals zero. This is given by the fact that the latter variables, in a certain manner stand aside from the other measured variables.
2. On the magnitude of the gross error: This is comprehensible as the higher the gross error, the easier it will be detected. Clearly, small gross errors (or slightly greater than the uncertainty of the measurement proper) can easily be masked by the presence of the other random errors.
3. On the degree of redundancy ν of the whole problem (see Eq. (3.3-1)): This influence is not too relevant for one given model and makes itself valid for all the variables nearly equally.

The detailed analysis of these influences is carried out in the above mentioned Appendix 4. It is essential that any redundant variable can be characterized, from the standpoint of gross error detection, by its so-called *gross error threshold value*.

Gross error threshold value tv_β is the value that will be detected with probability β

The threshold value is thus of probabilistic character. Whether a gross error will be really detected, that depends not only on its magnitude but also on the other random errors in the given data set, which may compensate the impact of the gross error on the testing criterion Q_{min} . The probability value β to be further applied is 0.9 (90 %). It is obvious that the smaller the value of tv is, the greater is the chance for the gross error detection. Nonadjustable (just determined) variables have an infinitely great threshold value.

Example 3.8-1: Gross error identification

This example is continuation of the preceding Example 3.4-1 – see also Fig. 3.4-2.

In the following table are given threshold values tv_{90} of the gross error detection for individual measured variables.

Table 3.8-1

Stream	Type	Measured value (kg/h)	Threshold value of gross error detection tv_{90} (kg/h)	Threshold value of gross error detection tv_{90} (%)
S1	M	100.1	4.78	4.8
S2	M	41.1	∞	∞
S3	M	79.0	4.62	5.8
S4	M	30.6	9.90	32
S5	M	108.3	9.90	9.1
S6	M	19.8	4.78	24

The last two columns of the table contain the threshold values (absolute in kg/h and relative in percents of the measured value). The interpretation is, e.g. for stream S1

Gross error 4.78 kg/h will be detected with probability 90 %

If the gross error is greater, of course also the probability of its detection will be greater. Let us stress the probabilistic character of this assertion. One can see that the threshold values for individual streams differ considerably as absolutely, as relatively with respect to the value of given variable. For nonadjustable variables (stream S2), an arbitrarily great gross error will never be detected. ♦

Concluding this section, let us still put the question how to distinguish the case of an actual (great) gross error from an error of the 1st kind (5 % of cases in the testing); see Appendix 3. It is then possible to start from the sum of squares of adjustments Q_{min} and its comparison with the *chi-square* distribution (see Appendix 1). From the course of the probability density for this distribution one can see that the probability of values greater than the critical one decreases rapidly with increasing value of the random variable (this holds in particular for the degrees of redundancy higher than 10). It is legitimate to suppose that by random errors, higher Q_{min} -values cannot be caused than for example twice the critical value.

3.9 Propagation of errors at data processing and measurement system optimization

The symbolical equations between measured and reconciled variables (3.4-6 and 7) comprise the information about how the measurement errors proper are propagated during the data processing. One can thus (in the statistical sense) trace back, which of the directly measured variables are the key ones from the point of view of the results uncertainty and upon which one must concentrate when enhancing the results precision. The problems to be solved will first be made clear using a very simple example.

Example 3.9-1: Propagation of measurement errors at measured data processing

Let us suppose that a directly unmeasured variable y is linear function of two directly measured variables x_1 and x_2

$$y = ax_1 + bx_2 \quad (3.9-1)$$

where a and b are constants $a = 1$ and $b = 2$. Let us further assume that the errors in x_1 and x_2 are uncorrelated and their standard deviations equal $\sigma_1 = 1$ and $\sigma_2 = 2$. For the dispersion (variance) of variable y we then have [2]

$$\sigma_y^2 = a^2\sigma_1^2 + b^2\sigma_2^2 \quad (3.9-2)$$

The terms $a^2\sigma_1^2$ and $b^2\sigma_2^2$ thus represent *contributions* of individual measured variables to the dispersion of the resulting variable y .

Evaluation yields

$$\sigma_y^2 = 1^2 \cdot 1^2 + 2^2 \cdot 2^2 = 1 + 16 = 17 \quad (3.9-3)$$

In this case we see that the dominant influence upon the dispersion of the result is that of variable x_2 , which contributes by ca. 94 % while variable x_1 only by 6 %. If requiring higher precision of the result, one has to concentrate on variable x_2 that represents the bottleneck of the whole process of measurement. If its dispersion is lowered to one half (e.g. by installing a more precise meter), the dispersion of the result decreases according to (3.9-2) from 17 to 5. On the contrary even an errorless measurement of variable x_1 ($\sigma_1 = 0$) lowers the result to 16 only. ♦

The above procedure can be applied to an arbitrary linear function of measured variables under the assumption that the measurement errors are uncorrelated. In practice, it can be applied with a certain imprecision even to nonlinear functions, so long as we linearize them by the Taylor series development. Eq. (3.9-2) then can be, for one variable y , approximated by the development at point \mathbf{x}^* in the symbolic form

$$y = h(\mathbf{x}) \cong h(\mathbf{x}^*) + \sum \partial h(\mathbf{x}) / \partial x_i \Delta x_i \quad (3.9-4)$$

and for the dispersion of variable y we then have

$$\sigma_y^2 \cong \sum (\partial h(\mathbf{x}) / \partial x_i)^2 \sigma_i^2 \quad (3.9-5)$$

The terms on the right hand side of Eq. (3.9-5) are always nonnegative and represent the *contributions* of individual measured variables to the dispersion of the result. We now form the vector of the relative contributions thus shares $\mathbf{s} = (s_1, s_2, \dots, s_i)^T$, the elements of which represent the percentual share of individual measured variables on the dispersion of the result:

$$s_i = 100 [(\partial h(\mathbf{x}) / \partial x_i)^2 \sigma_i^2] / \sigma_y^2 \quad (3.9-6)$$

The sum of the shares (vector elements of \mathbf{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 dispersion. Minimizing the dispersion 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 dispersion. The vector of shares itself is thus to be interpreted as the first information for further optimization steps supported by detailed calculations.

Example 3.9-2: Propagation of errors in mass balance

This example is continuation of the previous Example 3.4-1 – see also Fig.3.4-2.

The following table gives the share vectors for directly unmeasured streams S7 and S8.

Table 3.9-1

Stream	S1	S2	S3	S6	Sum
S7	16	62	16	5	99
S8	10	64	22	-	96

The interpretation of this table can be such that for stream S7, 16 % of the dispersion is caused by stream S1, 62 % by stream S2 etc. We see that for the uncertainty in computed (observable) streams S7 and S8, only measured streams S1, S2, S3 and S6 are of practical importance. The remaining two streams are of negligible influence and are also omitted here by the RECON program (they represent only 1 % for stream 7 and 4 % for stream 8).

The variable S2 is of dominant influence on the uncertainty (dispersion) in both streams; also non-negligible but smaller is the influence of streams S1 and S3. This means that if we want

to decrease the uncertainty in both the computed streams, it is reasonable to enhance the measurement precision of stream S2.

Let us now suppose that we have the possibility of installing, for this stream, a new flowmeter with half the original uncertainty (maximum error). The previous maximum error was 4 % of the measured value, so let us further assume 2 %.

Table 3.9-2

Stream	Type	Measured value (kg/h)	Measurement uncertainty (%)	Uncertainty of results (kg/h)	Measurement uncertainty (%)	Uncertainty of results (kg/h)
S1	M	100.1	2	1.30	2	1.30
S2	M	41.1	4	1.64	2	0.82
S3	M	79.0	2	1.24	2	1.24
S4	M	30.6	10	2.53	10	2.53
S5	M	108.3	4	2.63	4	2.63
S6	M	19.8	4	0.76	4	0.76
S7	N	10.0	-	2.10	-	1.54
S8	N	10.0	-	2.06	-	1.49

One can see that the uncertainties in the computed streams S7 and S8 have decreased to 73 % resp. 72 % of the original values. If the uncertainties in some of the other variables were lowered, the result would not be so pronounced.

Finally, let us look at the new vectors of shares.

Table 3.9-3

Stream	S1	S2	S3	S5	S6	Sum
S7	30	29	30	-	9	98
S8	20	31	43	3	3	99

If comparing this table with the preceding one we see that the situation has changed sensibly. The stream S2, from the viewpoint of its importance is no longer dominating and it is now better to concentrate on decreasing the uncertainty at stream S3. As already stated beforehand, in optimizing the measurement system one must proceed step by step. ♦

3.10 Parametric sensitivity

Parametric sensitivity is a well-known concept in the domain of mathematical modeling. With a considerable simplification, it can be formulated as the question:

What is the change in the result of computation with unit change in the parameter value?

Note: Consider for example the typical problem of nuclear power engineering. *How influenced is the nuclear reactor thermal power identification by the measurement error in a certain temperature? What error would be caused by the measurement of this temperature with error for example 2 deg C?* In the preceding section, this problem was solved in the statistical sense in terms of standard deviation of the result. Now, one deals with a deterministic problem under certain constellation of all the measured data. ♦

In the common mathematical modeling (e.g. simulation calculations), the situation is relatively straightforward. If the model is linear, the parametric sensitivity can be found easily and exactly. With nonlinear models, we find an approximate solution via the linearization of the model by Taylor series development. If the precise value of parametric sensitivity is to be identified, there is no other way than successive perturbation of all values of the variables, with so many calculations as is the number of the parameters.

In the case of DR, the situation is even more complicated. Let us again start from the symbolic formulation of the DR process by Eq. (3.4-7). In the previous section (Eq. (3.9-4)), we have already given the linearized form of the relation between measured variables and results. In contrast to classical mathematical modeling, the situation is now complicated by the fact that the reconciliation operator depends not only on the variables' values and the model form, but also on standard deviations of the measurement (which can again be defined as functions of the measured variables' values.) This is typical of the flowrates, where the measurement uncertainty is often expressed in per cents of the measured value. For this reason, one must take all results based on the linearization as orientative and when needed, scrutinize them by detailed calculations with perturbed variables.

Example 3.10-1: Parametric sensitivity in mass balance

This example is continuation of the previous example 3.4-1 – see also Fig. 3.4-2. The following table gives vectors of parametric sensitivity for streams S7 and S8.

Table 3.10-1

Stream	S1	S2	S3	S4	S5	S6
S7	0.42	-1.00	0.53	-0.05	0.05	0.58
S8	0.33	-1.00	0.62	-0.06	0.06	-0.33

This table can be interpreted in the manner that increasing the S1 flowrate value by unity increases the S7 flowrate value approximately by 0.42 kg/h, etc. It is to be noted that these conclusions hold only for the given set of measured values. They can thus be generally applied to processes running at steady state for a long time. ♦

3.11 Data validation in industrial processes monitoring

So far, we always solved the problem for one data set. Let us now consider that the data validation is part of a monitoring system evaluating the data automatically and regularly (e.g. every hour). Under these assumptions, we can trace the time series of measured and reconciled variables. Let us however consider first the data quality indicators. Recall that we here have

Q_{min} – weighted sum of squares of the adjustments due to reconciliation

Q_{crit} – critical value for Q_{min} , which should not be exceeded. This variable is constant for given number of degrees of freedom of the chi-square distribution (thus in our case the degree of redundancy)

It is in addition convenient to introduce so-called data *status* S defined by the relation

$$S = Q_{min} / Q_{crit} \tag{3.11-1}$$

Clearly, if it holds

$$S > 1 \tag{3.11-2}$$

the gross error presence has been detected. So the status summarizes in one number the information on data quality. If thus, during the monitoring, for a given data set the status is higher than 1 then the results of this set should not be accepted for further processing (until the gross error is found and eliminated).

In the following figure, we have the diagram of the S values evaluated on the basis of one hour averages for a period of 14 days (altogether 336 values). For completeness let us add that in this case, the degree of redundancy was 30, to which corresponds the critical value $Q_{crit} = 43.64$.

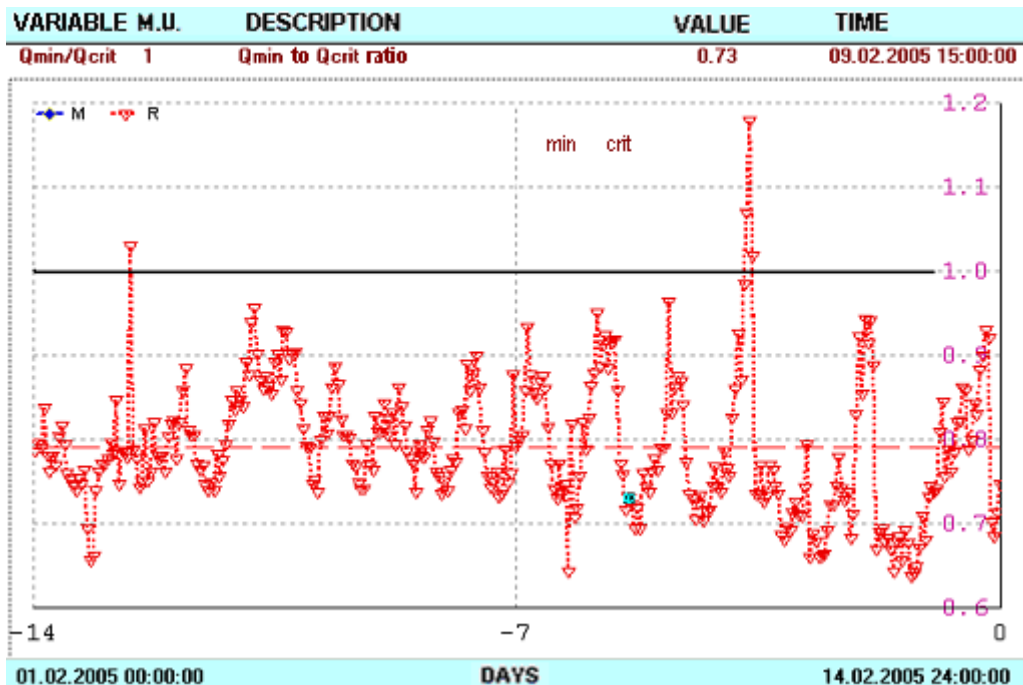


Fig. 3.11-1: Time series of variable S (status)

Altogether in 4 cases, a gross error was detected ($S > 1$) and the average value of the status (dashed line) is 0.79. Let us now put the question what this information

says about the data quality. Let us further examine certain properties of this time series.

MEAN STATUS VALUE

The starting point of further considerations is the fact that the sum of squares of the adjustments Q_{min} is a random variable with *chi-square* distribution having 30 degrees of freedom. For this random variable it holds that its mean value equals the number of degrees of freedom. So the mean value of Q_{min} should be, in the given case about 30. With the critical value 43.64, the assumed status results as $30/43.64 = 0.69$ (such status should result as the average after a long-term tracing, provided that all of our assumptions hold true). This value is fairly near to (though somewhat smaller than) the actual value 0.79.

FREQUENCY OF GROSS ERROR DETECTION

Another viewpoint is the frequency of exceeding the critical value for Q_{min} thus Q_{crit} , i.e. the gross error presence detection. The testing is carried out at the significance level 0.05. This means that the probability of an error of the 1st kind equals 0.05. In 5 % of cases we thus can expect that a gross error will be detected, although absent in reality. With the number of data 336 as is our case, we arrive at the number $336 \times 0.05 = 17$. This value is substantially higher than the above number 4 in our example.

TIME CORRELATEDNESS OF THE STATUS

If only random errors made themselves valid, the status values would oscillate around the mean value. The actual value would be independent of the previous ones. In Fig. 3.11-1 one can see that this is not the case. One can see here irregular 'waves' where the values go up and down again. It is a consequence of the fact that besides the random errors, also systematic errors are present. Moreover, the systematic errors may work for an arbitrarily long time – let us consider for example effects with daily, seasonal or one year's frequency, or defects of instruments causing a jump change until the instrument is repaired. Further causes can be for example calibration of meters or their swapping during the run of the production process. It is well known that with the measurement by instruments, systematic errors are mostly of greater importance than random errors.

Even better than using the status, the importance of systematic errors can be seen in the diagram of measured and reconciled variables. See for example the following figure.

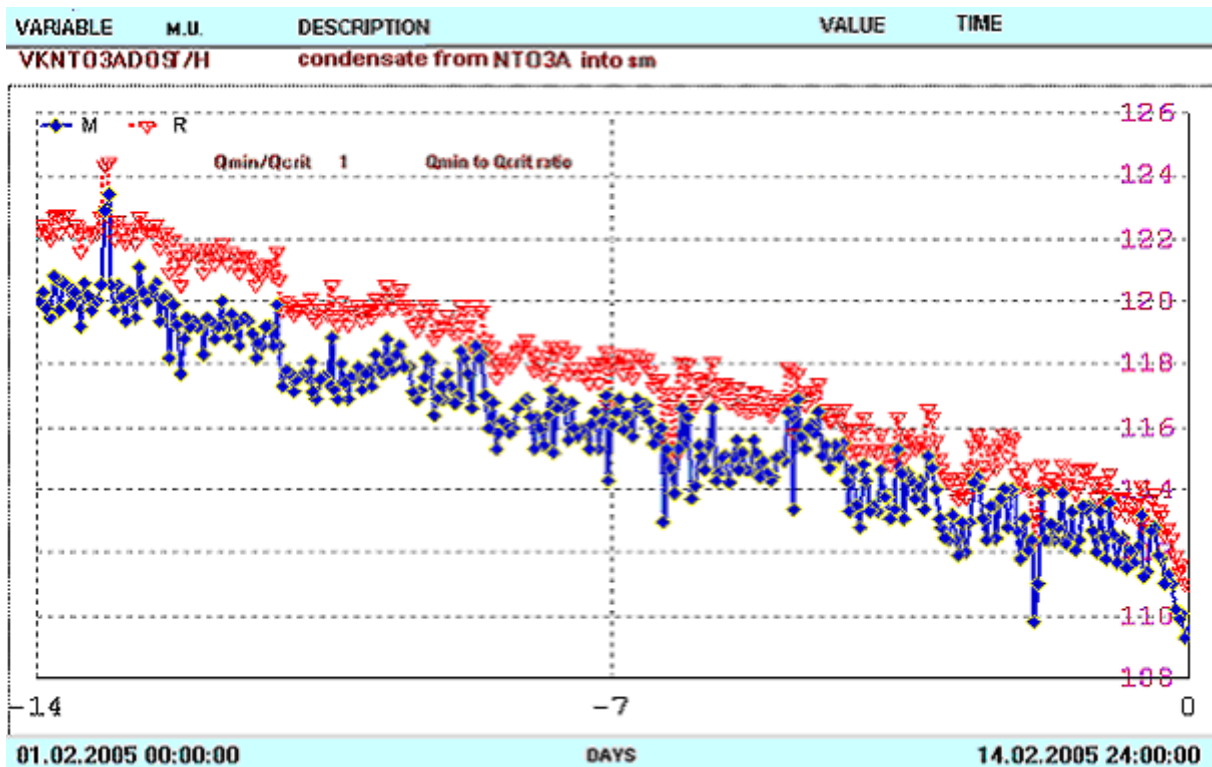


Fig. 3.11-2: Example of measured and reconciled values

Points marked by triangles represent reconciled values (the upper diagram), squares are measured values. If only random errors were present, the two diagrams should 'interpenetrate'. In addition the adjustments (differences between reconciled and measured values) should oscillate around zero. The figure shows clearly, that this is not the case. The measured values, but for several exceptions, are adjusted in the direction to higher values. The influence of systematic errors is here dominating.

The observation can be summarized as follows.

- The status average (0.79) is near to the theoretical value (0.69). As a whole, the system (the model proper and assumed measurement errors) is well designed, actual errors are somewhat higher than assumed
- In 4 cases, a gross error has been detected, which is less than corresponds to the significance level 0.05 (17 cases)
- The maximum value of the status ca. 1.2 indicates that in the whole period, no gross error was present (see the discussion in Section 3.8)

The above observations can be explained by the presence of relevant systematic errors. This fact however does not make the usefulness of DR doubtful – both random and systematic errors are reconciled and their impact is thus diminished. We are reconciling single data sets and in that moment we need not bother about the previous results. The thus arisen time series are analyzed only later off-line and the conclusions help us first better understand the whole data treatment process, and further eliminate the possible errors identified thereby.

The so far given illustration is example of data without serious problems. Let us further give still other examples, where the data requires correction.

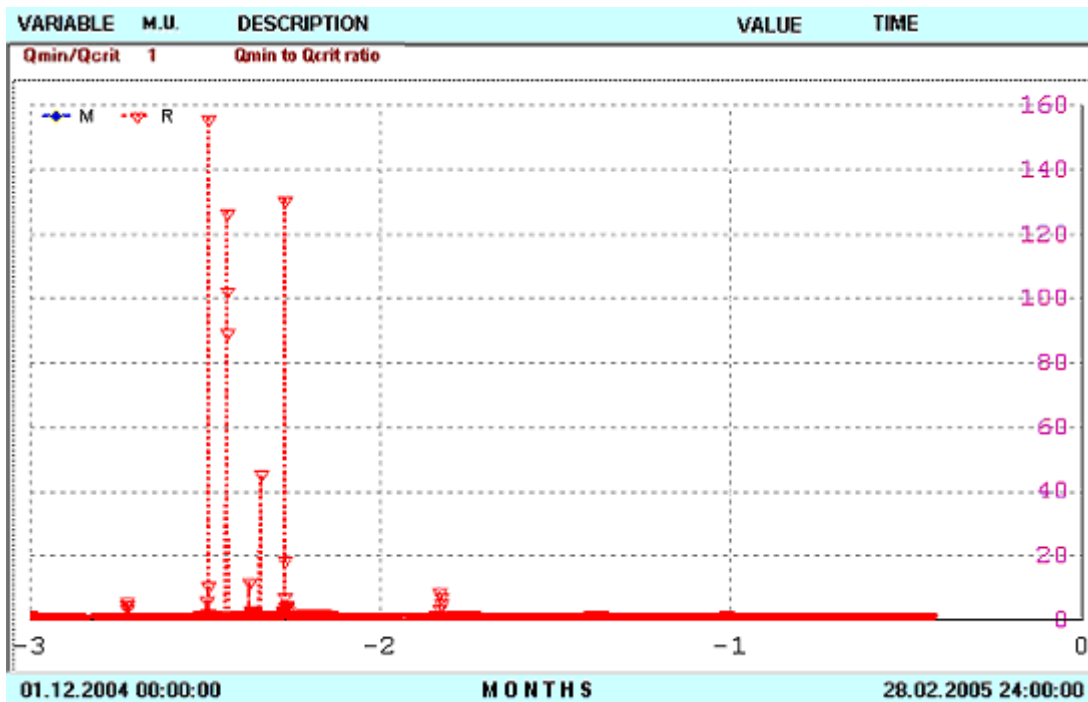


Fig. 3.11-3: Long-term trend of the Status – occurrence of isolated gross errors

In this figure one can see a long-time trend of the status (2.5 month) where several times, the status value exceeds 1 in the order of magnitude. In such cases, a 1st kind error does not come into consideration. It is most often a failure in data where data about one or several variables are absent. Because one deals with short periods, it is most likely not gross errors of the measuring instruments.

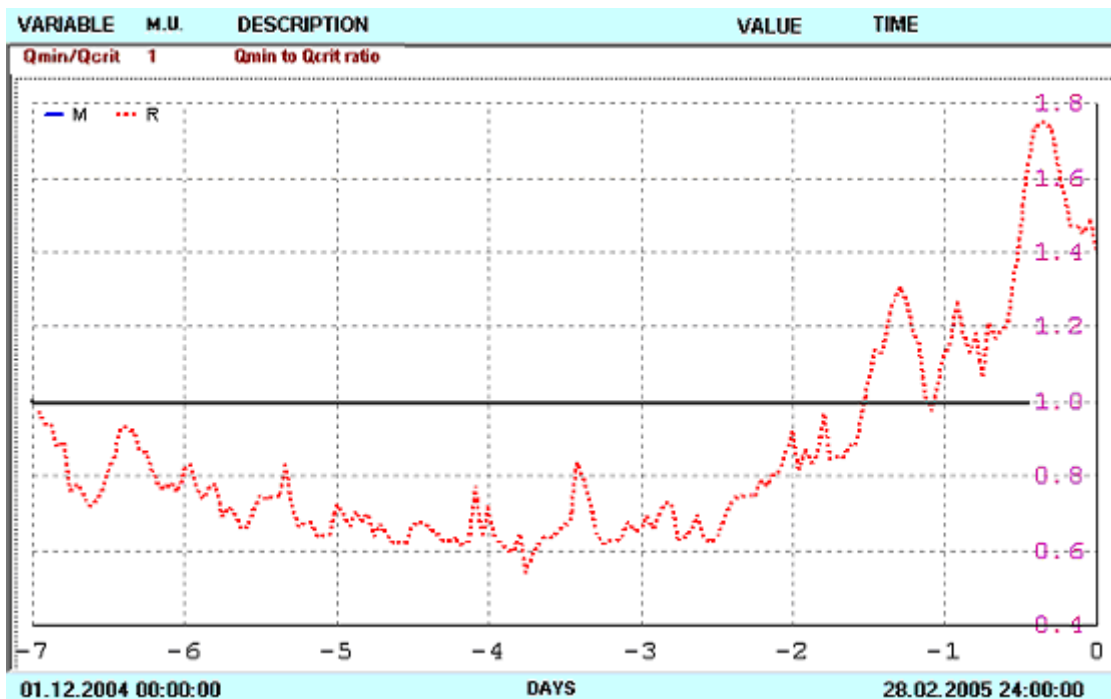


Fig. 3.11-4: Example of systematic increase in the Status

In this figure, one can see the systematic increase of the Status signaling successive deterioration of data quality. Until the cause is identified, most of the data are of no use.

3.12 Integration of DR into process information systems

Only in rare cases the balancing systems, DR included, can exist for themselves. Their routine application is conditioned by the automation of measured data input provided by production control and information systems. The up-to-date trend is to create a central *Process Data Warehouse* (relational database) denoted further as PDW, into which the data are imported.

The filling of the PDW is denoted by the abbreviation ETL (Extraction, Transformation and Loading) – see the monograph [7]. In the case of PDW, the key role in the process industry belongs to the data validation, which is part of the module Transformation.

In the PDW, process data are stored from the database of real-time process data or from that of aggregated data (average values etc.). Data from the PDW serves for a number of applications. Based thereon, adjustment of balances is carried out and the adjusted (reconciled) values are stored again in the PDW. They can then serve for other applications such as planning, scheduling, process monitoring, etc.

If there is no PDW as yet, the RECON system for DR can serve as the first step in creating it. The situation is depicted in the following figure.

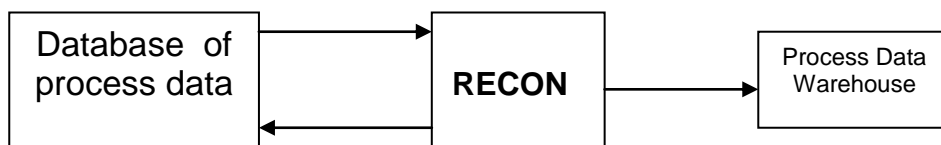


Fig. 3.12-1: Creating the data store in the frame of DR

During the on-line operation of the RECON program, all the three steps of the ETL can be carried out.

1. **EXTRACTION** – Read data from process data database and carry out their aggregation, i.e. creating average values with required frequency
2. **TRANSFORMATION** – Reconcile and validate the data
3. **LOADING** – Store the data in the data store.

In the data store, the data then can be at hand for tens of years. As PDW serves so-called *native* database of the RECON program preserving information on measured and reconciled values, data quality included. The simplest platform supported by RECON program is the MS ACCESS. For more extensive data systems, one can recommended relational databases of the client – server type – for example Oracle, Microsoft SQL or MySQL.

In the frame of the RECON program, these databases are supported by the means for database administration, data visualization and exports of selected data into the MS Excel. The data can be filtered according to whether one deals with good quality data or data that did not pass the test and require to be cleaned from gross errors.

For easier tracing, data can be ranged into thematic groups. Further available is the module for creating general reports in the Excel environment based on user-defined templates.

Besides these activities oriented towards their native database, RECON can of course simultaneously store the reconciled data into the original process data sources, as is depicted in the previous figure.

4 Examples of mass and component balances

Let us further give several typical examples of mass and component balances. This chapter will enable the reader to create simple models in the RECON program and check the results.

The structure of all sections is as follows.

1. **BALANCE SCHEME.** It is a copy of the scheme created in the graphical editor of program RECON.
2. **INPUT DATA.** It is a somewhat abridged extract from the data of the problem created in program RECON – menu *Flowsheet–Data review*. It contains all information necessary for the problem configuration.
3. **RESULTS.** It is an abridged extract from the results for the user's inspection, created in program RECON – menu *Calculate –Results*.

Although this manual cannot replace detailed manual to the RECON program, let us still briefly describe the procedure to be maintained in the problem creation.

1. Enter the name of the problem that is also the name of the file of the model being created
2. Enter the text of the problem title (long name) – not obligatory.
3. In the further panel, change the physical units (when necessary). Units selected in individual examples are given in the part INPUT DATA.
4. Enter the names of components (species). In the case of single-component balances, we recommend to enter 'mass'; full name need not be filled in.
5. The graphical editor screen turns up. The scheme drawing proper is recommended to be started by drawing first all nodes, which are conveniently placed on the screen (at later changes in size and placement of the nodes, the shape of the streams drawn can change). Helpful is also a lattice that can be called up in menu *Settings - Gridlines*. At the nodes, we fill in only Name (it is in the scheme) and Description (one need not fill in, or we devise some).
6. We then finish by drawing the streams. Their short names are given in the scheme, we invent again the description. Types of streams, values and possible errors (for the measured variables) can be found in part INPUT DATA. The problem definition for mass balance is thus complete.
7. After the configuration of all streams and nodes, one can carry out the computation.

Let us in addition explain certain abbreviations used in the RECON program.

F	type of variable - Fixed variable (known as errorless)
M	type of variable - Measured variable
MC	type of variable - Measured variable, adjustable (can be Corrected)
MN	type of variable - Measured variable, Nonadjustable
N	type of variable - uNmeasured variable
NO	type of variable - uNmeasured variable, Observable
NN	type of variable - uNmeasured variable, uNobservable

For most of the examples given in this chapter, the reader will have at hand also their model solutions, i.e. files with respective models. Names of the files are given in the text.

4.1 Mass balance at steady state without redundancy

In the following four examples, we'll show basic situations that can occur at the data validation. We'll make use of a simple scheme with four nodes and eight streams. One deals with a single-component balance. From the mathematical point of view, one deals with a linear model.

In the first example, in the problem are measured just so many variables how many are necessary for solving the set of balance equations (4 equations for 4 unmeasured variables). The degree of redundancy equals zero and in this case, it is thus impossible to carry out the data validation. The balance scheme is given in the next figure. The full lines represent measured streams, the dash-and dotted ones are unmeasured.

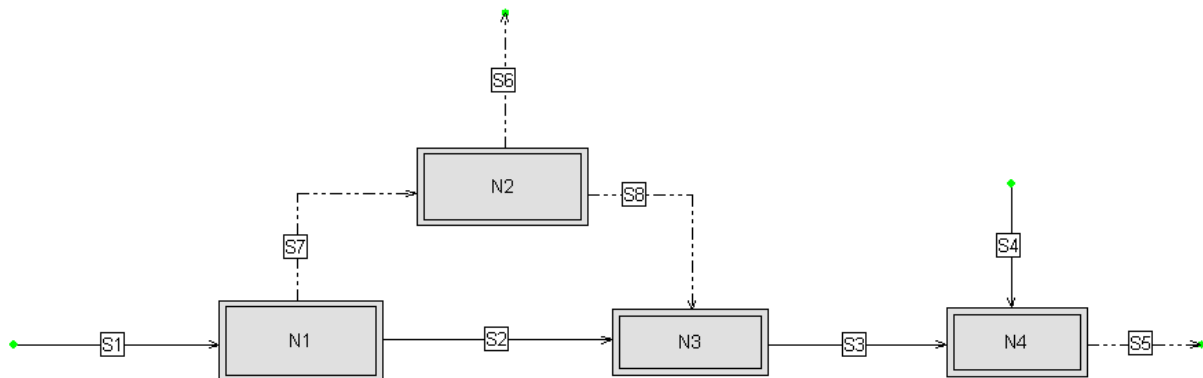


Fig. 4.1-1: Balance scheme (demo Example MC-1)

Input data

ID	TYPE	VALUE	MAX. ERROR
S1	M	100.1000	2.0000 %
S2	M	41.1000	4.0000 %
S3	M	79.0000	2.0000 %
S4	M	30.6000	10.0000 %
S5	N	10.0000	
S6	N	10.0000	
S7	N	10.0000	
S8	N	10.0000	

At the unmeasured variables (type N), their values represent the estimate necessary for starting the computation (so-called initial guess). This value has no influence on the final result. Still, it should be as close as possible to the actual value. The program finds thus easier the solution, and the risk of divergence of the computation

process is diminished (this risk can never be quite precluded). As the initial guess, RECON does not accept zero value.

Stream : S1 Sort of stream: Material

From node: **ENVIRON**
to node: **N1**

Mass flow [KG/S]

Stream	Description	Type	Value	Max.error
S1		M	100.1	2%

Concentrations [%]

Component	Description	Type	Value	Max.error
WATER		F	100	

Fig. 4.1-2: Example of stream definition panel

In this panel, we can enter or modify the name of the stream, its description, type (M,N, or F). In case of type M, we must enter the maximum error (absolute or in per cents). In the case of mass (single-component) balance only, the concentration of the one component is always 100 % (obligatorily fixed).

In the next figure, a node property definition is illustrated.

Node: N1 Sort of calculations:

ID: N1 Description:

Above sea level: Node pres.:

Balancing
 Hydraulic node
 Heat node
 Reaction node

Non-energetic streams incident with node

Stream	Function	Temperature	Pressure	Wetness
S1				
S2				
S7				

Reactions in node: Reaction + -

Fig.4.1-3: Illustration of the node definition panel

Here, we again utilize the possibility of entering the name of the node (ID), contingently its description.

RESULTS

ITERATION

Iter	Qeq	Qx	Qy	Qmin
START	2.8722E+01			
1	0.0000E+00	0.0000E+00	2.8748E+01	0.0000E+00
2	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00

Key:

Qeq mean residual of equations

Qx mean increment of measured variables in iteration

Qy mean increment of unmeasured variables in iteration
 Qmin least squares 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 4
Number of adjusted variables 0
Degree of redundancy     0
Number of unmeasured variables 4
Number of observable variables 4
Number of unobservable variables 0
Number of free variables  0
Number of equations       4
Number of independent equations 4
Number of user defined equations 0

Mean residual of the equations 0
Qmin                        0
Qcrit                      0
Status (Qmin/Qcrit)       0
```

S T R E A M S		[KG/S]		
Name	Type	Inp.value	Rec.value	Abs.error
S1	MN	100.100	100.100	2.002
S2	MN	41.100	41.100	1.644
S3	MN	79.000	79.000	1.580
S4	MN	30.600	30.600	3.060
S5	NO	10.000	109.600	3.444
S6	NO	10.000	21.100	2.550
S7	NO	10.000	59.000	2.591
S8	NO	10.000	37.900	2.280

In this first example, let us describe in detail individual information given in the extract.

- Variable *Qeq* is the residual of the equations. The starting value (START) gives the mean quadratic residual of the model equations with the values of measured and estimates (guesses) of unmeasured data.
- Variable *Qx* is mean quadratic increment of measured variables in the iteration. Because one deals with an example without data redundancy, the measured variables need not be adjusted and this value is zero.
- Variable *Qy* is mean quadratic increment of unmeasured variables in the iteration. In the given case (linear model), the values of unmeasured variables are adjusted in just one step (calculated from the measured values). The second iteration is only of checking importance.
- Variable *Qmin* (least squares function) is here zero, as no measured values corrections (adjustments) take place.

The global data need no commentary. Let us perhaps only state the fact that the degree of redundancy is zero, because one does not deal with a redundant system. .

The measured streams are marked by MN, which means that they are Measured and Not corrected (non-adjusted). At unmeasured variables, NO means Non-measured and Observable.

In the results (table Streams), *Inp.value* means measured value or initial estimate for unmeasured variables. *Rec.value* means either adjusted (reconciled) measured value, or computed unmeasured value. In the last column, we have the maximum error of the result.

For example the unmeasured variable S5 was initially estimated (guessed) by the value 10, the computed value was 109.60 and the maximum error was 3.44, all in kg/s. The given value of the error means that the actual (unknown) value lies with probability 95 % in the interval

$\langle 109.600 - 3.444; 109.600 + 3.444 \rangle$.

4.2 Mass balance at steady state with redundancy

The example in the preceding section is modified in the manner that streams No. 5 and 6 are now regarded as measured. This assumption reduces the number of unknowns to 2. As we now have 4 balance equations and 2 unknowns, the degree of redundancy according to Eq. (3.3-1) is $4 - 2 = 2$. Some results for this example were already given in Section 3.4 (Example 3.4-1). This information will now be completed.

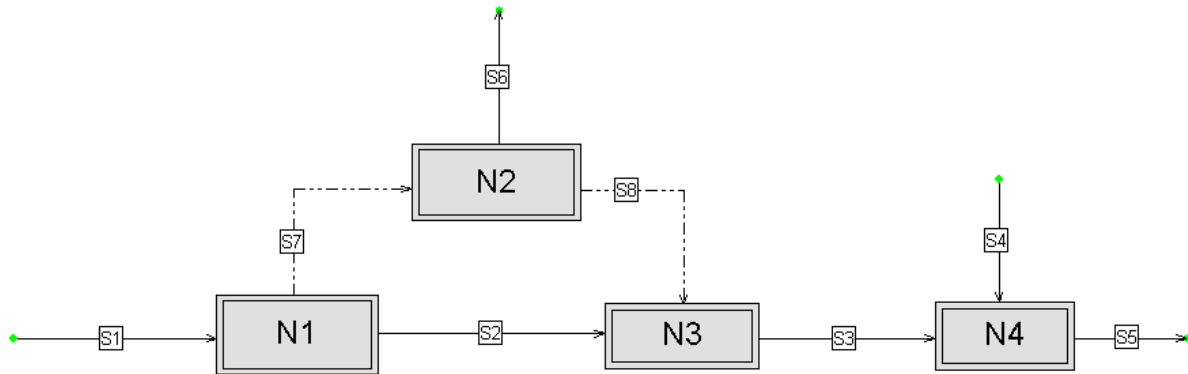


Fig. 4.2-1: Balance scheme (demo Example MC-2)

Dash-and-dotted lines at streams S7 and S8 designate unmeasured streams. The input data are the following:

MATERIAL STREAMS [KG/S]

ID	TYPE	VALUE	MAX. ERROR
S1	M	100.1000	2.0000 %
S2	M	41.1000	4.0000 %
S3	M	79.0000	2.0000 %
S4	M	30.6000	10.0000 %
S5	M	108.3000	4.0000 %
S6	M	19.8000	4.0000 %
S7	N	10.0000	
S8	N	10.0000	

The results of data reconciliation are:

ITERATION

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

Qeq mean residual of equations
 Qx mean increment of measured variables in iteration
 Qy mean increment of unmeasured variables in iteration
 Qmin least squares function

GLOBAL DATA

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

Number of measured variables	6
Number of adjusted variables	5
Degree of redundancy	2
Number of unmeasured variables	2
Number of observable variables	2
Number of unobservable variables	0
Number of free variables	0
Number of equations	4
Number of independent equations	4
Number of user defined equations	0
Mean residual of the equations	3.5527E-15
Qmin	1.3081E+00
Qcrit	5.9739E+00
Status (Qmin/Qcrit)	0.218963

S T R E A M S [KG/S]

Name	Type	Inp.value	Rec.value	Abs.error
S1	MC	100.100	99.287	1.300
S2	MN	41.100	41.100	1.644
S3	MC	79.000	79.359	1.239
S4	MC	30.600	30.048	2.533
S5	MC	108.300	109.407	2.632
S6	MC	19.800	19.927	0.755
S7	NO	10.000	58.187	2.096
S8	NO	10.000	38.259	2.058

In the information about the course of computation we see that in contrast to the preceding example, the measured values are now adjusted (column Q_x in part Iteration). In the second iteration, the increments of variables and residuals are practically zero (given by the number of digits with which the numbers are stored in the computer).

The value of the weighted sum of squares of the adjustments Q_{min} equals 1.308, which is less than the critical value $Q_{crit} = 5.9739$ (for the chi-square distribution with two degrees of freedom and significance level 95 %). The situation is well characterized by so-called *Status* of the data quality defined by equation (3.11-1), which is the ratio Q_{min}/Q_{crit} . So long as $Q_{min} < Q_{crit}$, the data are in order (no gross error presence detected). In this case, the Status must clearly be smaller than or equal to the value 1.

Concerning this case there are several studies in Chapter 3, which will not be repeated here. Only for brief information:

- enhancing the precision of results as a consequence of reconciliation (Example 3.5-1)
- gross error presence detection (Example 3.6-1)
- gross error identification (Example 3.7-1)
- efficiency of gross errors detection (Example 3.8-1)
- propagation of errors in data processing and measurement system optimization (Example 3.9-2)
- parametric sensitivity (Example 3.10-1).

4.3 Mass balance in steady state with unobservable variables

The aim of this example is to show the consequences of insufficient instrumentation of the measurement system.

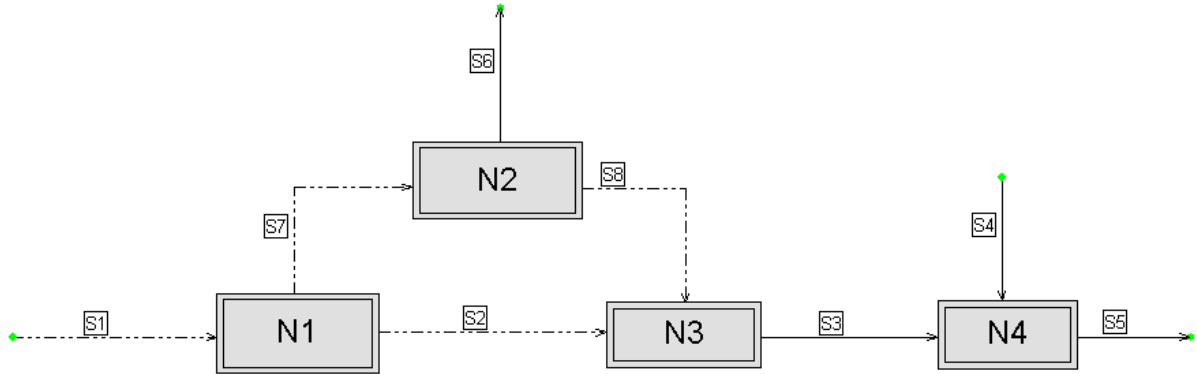


Fig. 4.3-1: Balance scheme (demo Example MC-4)

We have here 4 unmeasured streams (dash-and-dotted), so one could believe that they could be computed from the 4 balance equations. The input data are:

M A T E R I Á L S T R E A M S [KG/S]

ID	TYPE	VALUE	MAX.ERROR
S1	N	100.1000	
S2	N	41.1000	
S3	M	79.0000	2.0000 %
S4	M	30.6000	10.0000 %
S5	M	108.3000	4.0000 %
S6	M	19.8000	4.0000 %
S7	N	10.0000	
S8	N	10.0000	

The abridged results read:

```

Number of measured variables          4
Number of adjusted variables          3
Degree of redundancy                  1
Number of unmeasured variables        4
Number of observable variables         1
Number of unobservable variables       3
Number of free variables               1
Number of equations                   4
Number of independent equations        4
  
```

W A R N I N G

1. Some unobservable variables detected.
Please, use menu item 'Calculate > Classification'

S T R E A M S [KG/S]

Name	Type	Inp.value	Rec.value	Abs.error
S1	NO	100.100	98.694	1.709
S2	NN	UNOBSERVABLE		
S3	MC	79.000	78.894	1.514
S4	MC	30.600	30.203	2.550
S5	MC	108.300	109.097	2.696
S6	MN	19.800	19.800	0.792
S7	NN	UNOBSERVABLE		
S8	NN	UNOBSERVABLE		

Here, the situation differs from the case examined in Section 4.1. We have now met with the situation where in one part of the scheme, the measurement is quite sufficient and redundancy occurs (around node N4), while in other places, there is lack of measurement (around nodes N1, N2, N3). This can be met in practice quite often, because certain regions of production units are, due to their importance, endowed with meters considerably more than regions of smaller importance. The consequence is then the notice (warning) in the extract from the program, viz. that unobservable variables are present.

Let us first observe that in this case, there no longer holds the relation (3.3-1) that under certain circumstances enables us computing the degree of redundancy. This relation gives degree of redundancy 0 (no redundancy), while in reality all streams connected with node N4 are measured, and also reconciled.

In the basic information on the results, there is the item *Number of free variables* ; this is the necessary number of unmeasured variables to be measured or otherwise fixed, so as to make all variables observable. In the given case, the number of free variables is 1, thus just one unmeasured variable must be given (preferably measured) in addition to make the system fully observable. The selection of this variable is, however, not arbitrary. The problem is solved by the RECON program in menu *Calculations – Classification*. The result reads as follows.

U N O B S E R V A B L E V A R I A B L E S

Type Variable

```
MF  S2
MF  S7
MF  S8
```

```
Unobservable variables: 3
1 must be measured or fixed
```

Among the given streams S2, S7, S8, we can choose. Let us note that it makes no sense to measure in addition the further unmeasured stream S1. If we did so, the observability problem would not be solved, only the degree of redundancy would increase (as the reader can verify).

4.4 Mass balance at steady state – an unsolvable system

Besides the variables of types M (measured) and N (not measured), the program RECON knows also so-called *fixed* variables thus constants that do not change in the course of data processing. They also can be conceived as measured variables, with null error. Because the program cannot freely manipulate with their values, in certain situations the data reconciliation process cannot be brought to end (to attain zero residuals). Such a case takes place with the following scheme.

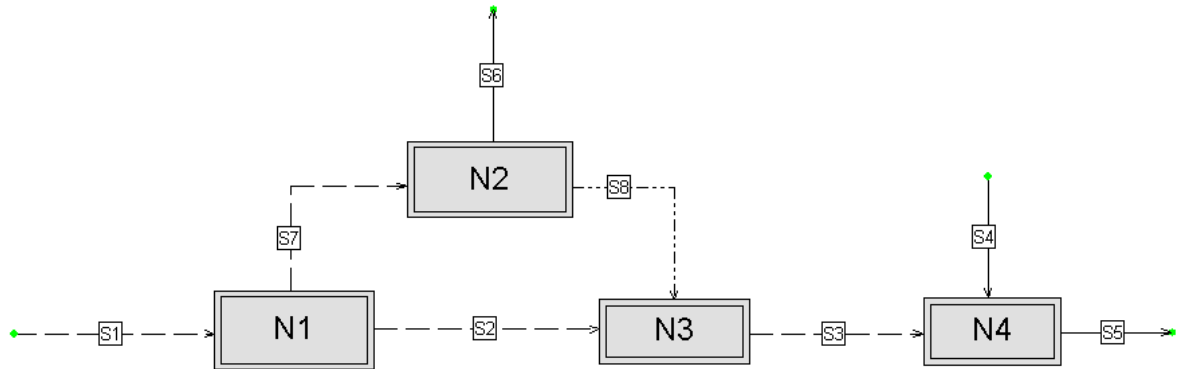


Fig. 4.4-1: Balance scheme (demo Example MC-5)

The dashed lines represent here fixed streams. It is obvious at first sight that in the node N1 balance, only fixed streams occur. If the values of their flowrates are chosen arbitrarily, the balance around this node will most likely not be satisfied. The input data are following.

M A T E R I A L S T R E A M S [KG/S]			
ID	TYPE	VALUE	MAX.ERROR
S1	F	100.1000	
S2	F	41.1000	
S3	F	79.0000	
S4	M	30.6000	10.0000 %
S5	M	108.3000	4.0000 %
S6	M	19.8000	4.0000 %
S7	F	58.0000	
S8	N	10.0000	

During the computation with these data, the following results turn up.

I T E R A T I O N				
Iter	Qeq	Qx	Qy	Qmin
START	9.9258E+00			
1	2.5000E-01	3.3809E-01	2.7900E+01	7.8199E-01
2	2.5000E-01	4.2509E-15	0.0000E+00	7.8199E-01
3	2.5000E-01	4.2509E-15	0.0000E+00	7.8199E-01
4	2.5000E-01	4.2509E-15	0.0000E+00	7.8199E-01
5	2.5000E-01	4.2509E-15	0.0000E+00	7.8199E-01

Task does not converge !!!

It is seen that one has not been able to attain zero residuals of the equations. If employing the service of menu *Calculate – Solvability*, we obtain the following message.

MESSAGE ON SOLVABILITY
The following fixed variables are not consistent:

Type of variable

MF S1
MF S2
MF S7

Legend:

MF Mass flow

Note

From these variables 1 must be re-classified to 'M' (measured) or 'N' (unmeasured)
Please correct your task.

4.5 Multicomponent balance of the LPG separation

The mixture of hydrocarbons LPG (Liquefied Petroleum Gas) is separated in the system of three distillation columns

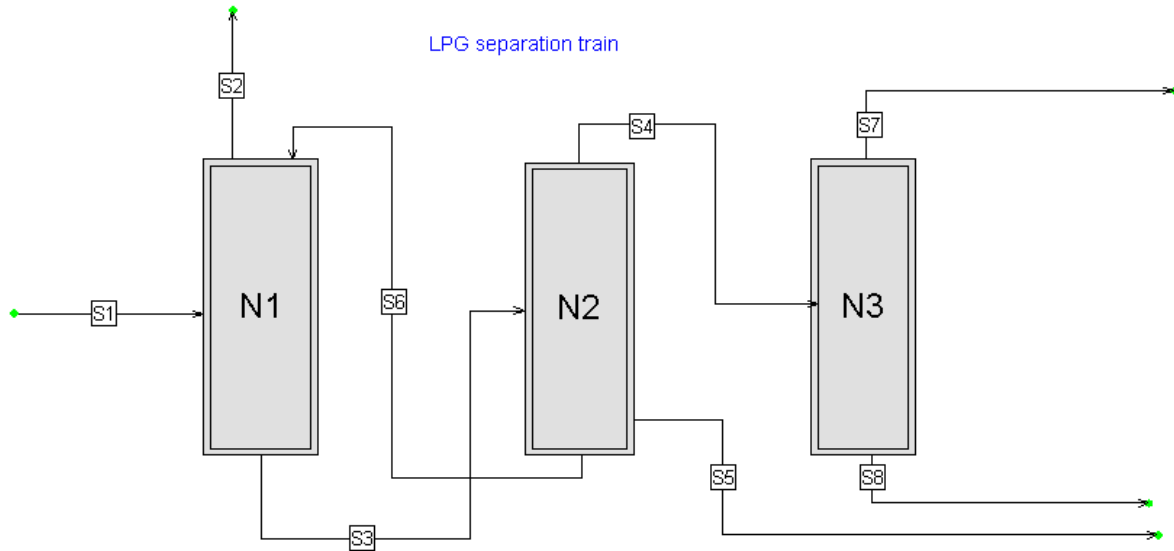


Fig. 4.5-1: Balance scheme (demo example MC-6)

We have here altogether 5 components, 3 balancing nodes (distillation columns) and 8 streams.

Components:

C1	ethane
C2	propane
C3	i-butane
C4	n-butane
C5	pentane+ (pentane and higher hydrocarbons)

The hydrocarbon mixture S1 is led into node N1, so-called absorber – desorber. In its upper part, heavier hydrocarbons are stripped by the pentane fraction S6, giving rise to so-called lean (residual) gas consisting mainly of ethane. In column N2, hydrocarbons C4 and lighter (stream S4) are separated from pentane and heavier hydrocarbons (streams S5 and S6). In column N3, propane (stream S7) is separated from the butanes (stream S8).

All flowrates and their composition are measured with the exception of the stream S3 composition. The maximum errors at the measurement of concentrations are expressed as so-called *relative per cents*, i.e. per cents from the concentration measured itself also in per cents. The input data read as follows.

M A T E R I A L S T R E A M S [kg/h], [%]

ID	COMPONENT	TYPE	VALUE	MAX.ERROR
S1	Flowrate	M	8620.0000	4.0000 %
	C1	M	10.500000	5.000000 %
	C2	M	32.000000	3.000000 %
	C3	M	43.600000	3.000000 %

	C4	M	3.000000	15.000000	%
	C5	M	9.500000	5.000000	%
S2	Flowrate	M	1040.0000	6.0000	%
	C1	M	85.700000	2.000000	%
	C2	M	2.100000	15.000000	%
	C3	M	2.300000	15.000000	%
	C4	M	1.300000	15.000000	%
	C5	M	9.900000	5.000000	%
S3	Flowrate	M	17800.0000	4.0000	%
	C1	N	0.100000		
	C2	N	15.000000		
	C3	N	20.000000		
	C4	N	5.000000		
	C5	N	60.000000		
S4	Flowrate	M	6860.0000	4.0000	%
	C1	M	0.200000	40.000000	%
	C2	M	41.200000	3.000000	%
	C3	M	54.200000	3.000000	%
	C4	M	2.700000	5.000000	%
	C5	M	0.600000	40.000000	%
S5	Flowrate	M	810.0000	2.0000	%
	C1	F	0.00000E+0		
	C2	M	0.400000	40.000000	%
	C3	M	1.800000	15.000000	%
	C4	M	8.200000	5.000000	%
	C5	M	90.200000	1.000000	%
S6	Flowrate	M	10400.0000	4.0000	%
	C1	F	0.00000E+0		
	C2	F	0.00000E+0		
	C3	M	0.200000	40.000000	%
	C4	M	3.300000	15.000000	%
	C5	M	95.900000	0.500000	%
S7	Flowrate	M	2850.0000	2.0000	%
	C1	M	0.600000	40.000000	%
	C2	M	96.200000	0.500000	%
	C3	M	3.700000	15.000000	%
	C4	F	0.00000E+0		
	C5	F	0.00000E+0		
S8	Flowrate	M	4060.0000	2.0000	%
	C1	M	0.100000	40.000000	%
	C2	M	2.500000	15.000000	%
	C3	M	91.400000	1.000000	%
	C4	M	4.500000	5.000000	%
	C5	M	0.700000	40.000000	%

At the configuration of this problem, we must first enter the components C1 to C5. The full name is not filled in.

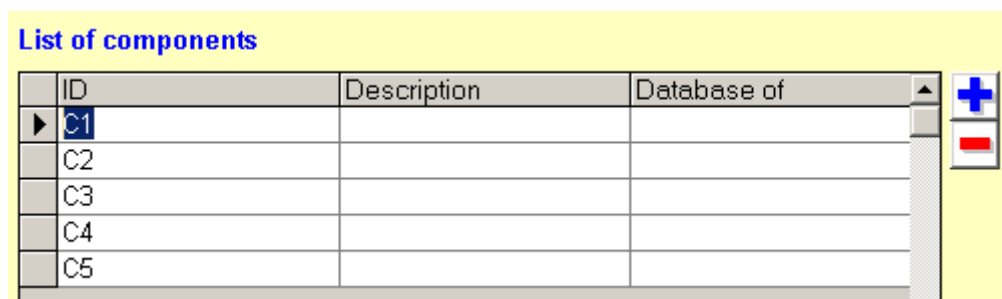


Fig. 4.5-2: Panel for components definition

At the creation of streams, besides the flowrates one also fills in the component concentrations.

Stream : S1 Sort of stream: Material

From node: **ENVIRON** Dependent stream
to node: **N1**

Mass flow [kg/h]

Stream	Description	Type	Value	Max.error
S1		M	8620	4%

Concentrations [%]

Component	Description	Type	Value	Max.error
C1		M	10.5	5%
C2		M	32	3%
C3		M	43.6	3%

Fig. 4.5-3: Panel for filling-in the streams parameters

We further give the extract of the results. At the stream properties, for the sake of space we restrict ourselves to one stream.

Problem: MC-6 (Component balance of system of 3 dist. columns)

I T E R A T I O N

Iter	Qeq	Qx	Qy	Qmin
START	6.6866E-03			
1	7.9806E-05	5.5151E-04	4.6870E-03	2.1447E+01
2	1.5241E-09	1.9499E-06	5.5365E-05	2.1364E+01

Key:

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

G L O B A L D A T A

Number of nodes	3
Number of streams	8
Number of components	5
Number of measured variables	38
Number of adjusted variables	38
Degree of redundancy	18
Number of unmeasured variables	5
Number of observable variables	5
Number of unobservable variables	0
Number of free variables	0
Number of equations	23
Number of independent equations	23
Number of user defined equations	0
Mean residual of the equations	1.5241E-09
Qmin	2.1364E+01
Qcrit	2.8919E+01
Status (Qmin/Qcrit)	0.738748

S T R E A M S [kg/h], [%]

Name of stream: S1

No.	Name	Type	Inp.value	Rec.value	Abs.error
Flowrate	MC		8620.000	8756.334	102.642
1	C1	MC	10.500	10.429	0.368
2	C2	MC	32.000	32.677	0.413
3	C3	MC	43.600	44.048	0.459
4	C4	MC	3.000	3.017	0.084
5	C5	MC	9.500	9.829	0.196

4.6 Component balance of the chlorination of methane

The following section describes the balance of a single node – reactor – where several chemical reactions take place.

Chlorination reactor

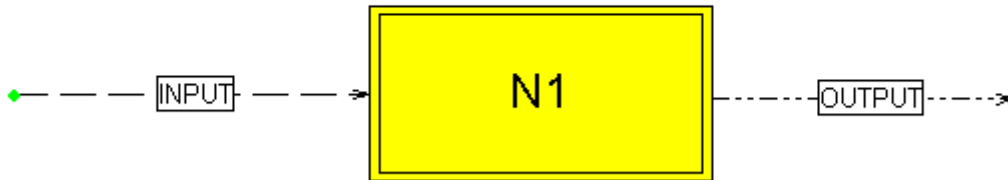
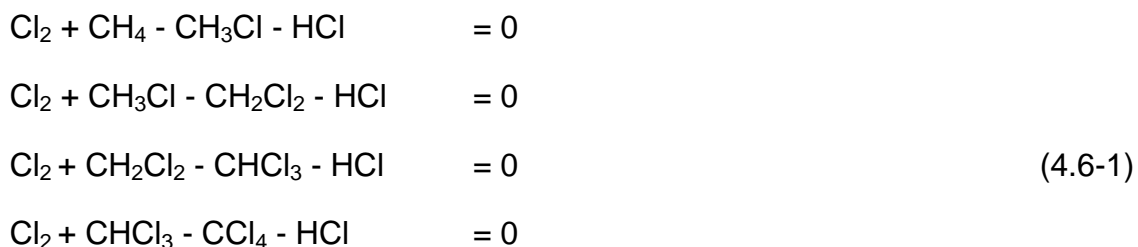


Fig. 4.6-1: Balance scheme (demo example MC-7)

Let us consider the chlorination of methane. From the balancing point of view, altogether 7 components occur in the system (the balancing viewpoint means that one ignores intermediate reaction products, radicals etc., which may participate in the reactions but do not belong to the initial reaction species, nor to the products):

Component No.	Formula
1	Cl ₂
2	CH ₄
3	CH ₃ Cl
4	CH ₂ Cl ₂
5	CHCl ₃
6	CCl ₄
7	HCl

The chlorination reactor has one inlet and one outlet stream. For the seven components, four independent stoichiometric equations can be written.



The set of equations (4.6-1) constitutes so-called *maximal set of stoichiometric equations* for the given system of species. Any further reaction must be linearly dependent on the above equations (it can be formed as a linear combination of the latter). For example the stoichiometric equation



is sum of the first two equations in (4.6-1). We'll return later to the problem that would arise, if we added this equation to the maximal set and wanted then use it in the balancing.

Let us now go back to the original problem formulation in the RECON program. We first give the components occurring in the problem.

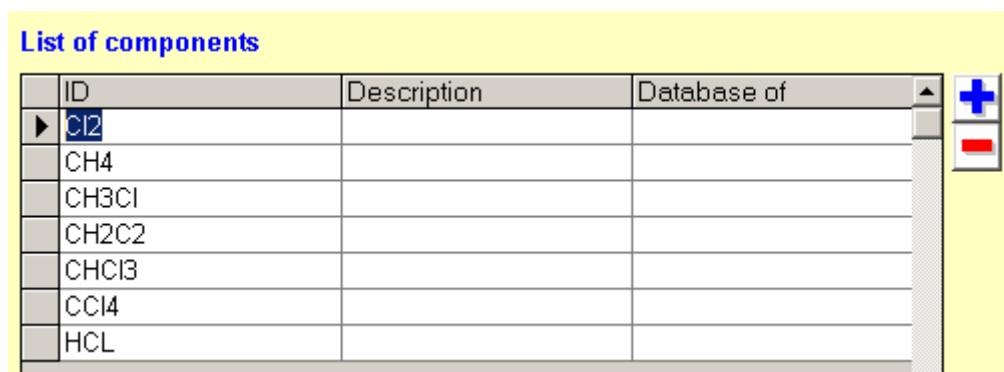


Fig. 4.6-2: Panel for defining the components

Further, we have to enter the values of flowrates and concentrations.

M A T E R I A L S T R E A M S [mol/s], [%]				
ID	COMPONENT	TYPE	VALUE	MAX.ERROR
INPUT	Flowrate	M	100.0000	3.0000 %
	C12	M	62.100000	10.000000 %
	CH4	M	26.900000	10.000000 %
	CH3Cl	M	10.800000	10.000000 %
	CH2Cl2	M	0.780000	10.000000 %
	CHCl3	M	0.380000	10.000000 %
	CCl4	F	0.00000E+0	
	C12	F	0.00000E+0	
OUTPUT	Flowrate	N	100.0000	
	C12	F	0.00000E+0	
	CH4	M	1.570000	10.000000 %
	CH3Cl	M	6.830000	10.000000 %
	CH2Cl2	M	24.800000	10.000000 %
	CHCl3	M	4.820000	10.000000 %
	CCl4	M	0.390000	10.000000 %
	C12	M	59.300000	10.000000 %

Our reactor is a laboratory one, where only samples of inlet and outlet streams are taken and analyzed. The flowrate is measured at the inlet only. One sets up the material balance in units of species (matter) variable. After entering the properties of streams, one has to enter information on the chemical reactions. This is done in two steps. First, one creates so-called *reaction bank*.

List of reactions

Heat of reaction: Positive for exothermic reactions, [KJ/mol]
Stoich.coefficients: Positive for reactants, negative for products

ID	Description	Heat	Cl2	CH4	CH3Cl	CH2Cl2	CHCl3	CCl4	HCl
R1		0	1	1	-1	0	0	0	-1
R2		0	1	0	1	-1	0	0	-1
R3		0	1	0	0	1	-1	0	-1
R4		0	1	0	0	0	1	-1	-1

Fig. 4.6-3: Panel for creating the reaction bank

In the second step, with the reaction nodes one associates the reactions defined in the bank.

Node: N1

ID: Description:

Above sea level: Node pres.:

Sort of calculations:

- Balancing
- Hydraulic node
- Heat node
- Reaction node

Non-energetic streams incident with node

Stream	Function	Temperature	Pressure	Wetness
INPUT				
OUTPUT				

Reactions in node

Reaction
R1
R2
R3
R4

Fig. 4.6-4: Associating reactions with reaction node

At the panel of node in Fig. 4.6-4, we first mark the window *Reaction node* on the right above. One then can choose a reaction from the list in the reaction bank.

The data input is thus finished. The results of the data reconciliation are following:

```

GLOBAL DATA
Number of nodes                1
Number of streams              2
Number of components           7
Number of reactions            4
Number of reaction nodes       1
Number of measured variables    12
Number of adjusted variables    11
Degree of redundancy           4
Number of unmeasured variables  5
Number of observable variables  5
Number of unobservable variables 0
Number of free variables        0

Number of equations            9
Number of independent equations 9
Number of user defined equations 0
Mean residual of the equations  4.9403E-07
Qmin                           9.4573E-01
Qcrit                           9.5145E+00
Status (Qmin/Qcrit)            0.099400

```

S T R E A M S [mol/s], [%]

Name of stream: INPUT

No.	Name	Type	Inp.value	Rec.value	Abs.error
	Flowrate	MN	100.000	100.000	3.000
1	Cl2	MC	62.100	60.716	0.463
2	CH4	MC	26.900	27.328	0.742
3	CH3Cl	MC	10.800	10.797	1.031
4	CH2Cl2	MC	0.780	0.780	0.078
5	CHCl3	MC	0.380	0.380	0.038
6	CCl4	F	0.00E+0	0.00E+0	
7	Cl2	F	0.00E+0	0.00E+0	

Name of stream: OUTPUT

No.	Name	Type	Inp.value	Rec.value	Abs.error
	Flowrate	NO	100.000	99.993	3.000
1	Cl2	F	0.00E+0	0.00E+0	
2	CH4	MC	1.570	1.570	0.157
3	CH3Cl	MC	6.830	6.862	0.670
4	CH2Cl2	MC	24.800	25.598	0.818
5	CHCl3	MC	4.820	4.864	0.466
6	CCl4	MC	0.390	0.390	0.039
7	Cl2	MC	59.300	60.716	0.463

R R E A C T I O N E X T E N T S

Node	Reaction	Extent
N1	R1	25.756
	R2	29.691
	R3	4.875
	R4	0.390

Let us go back once more to the problems concerning the set of stoichiometric reactions used in the balancing. To the set of equations (4.6-1), let us add the dependent equation (4.6-2). The reaction bank is then of the form

List of reactions

Heat of reaction: Positive for exothermic reactions, [KJ/mol]
Stoich.coefficients: Positive for reactants, negative for products

ID	Description	Heat	Cl2	CH4	CH3Cl	CH2Cl2	CHCl3	CCl4	HCl
R1		0	1	1	-1	0	0	0	-1
R2		0	1	0	1	-1	0	0	-1
R3		0	1	0	0	1	-1	0	-1
R4		0	1	0	0	0	1	-1	-1
R5		0	2	1	0	-1	0	0	-2

If we associate all these 5 reactions with reactor N1, the computation will run with the result that certain reaction extents remain unobservable.

R E A C T I O N E X T E N T S

Node	Reaction	Extent
N1	R1	UNOBSERVABLE
	R2	UNOBSERVABLE
	R3	4.875
	R4	0.390
	R5	UNOBSERVABLE

All other results remain however the same as before adding the dependent reaction (4.6 2). This means only that the extents of the dependent reactions cannot be identified by the measured data (although all the reactions can run in the given species system). On the other hand, if some of the independent reactions of the set (4.6-1) were considered absent, the results would change and probably, the presence of a gross error would be detected.

4.7 Ammonia synthesis reactor – master and dependent streams

The following section describes the multi-component balance of a reactor for synthesis of ammonia.

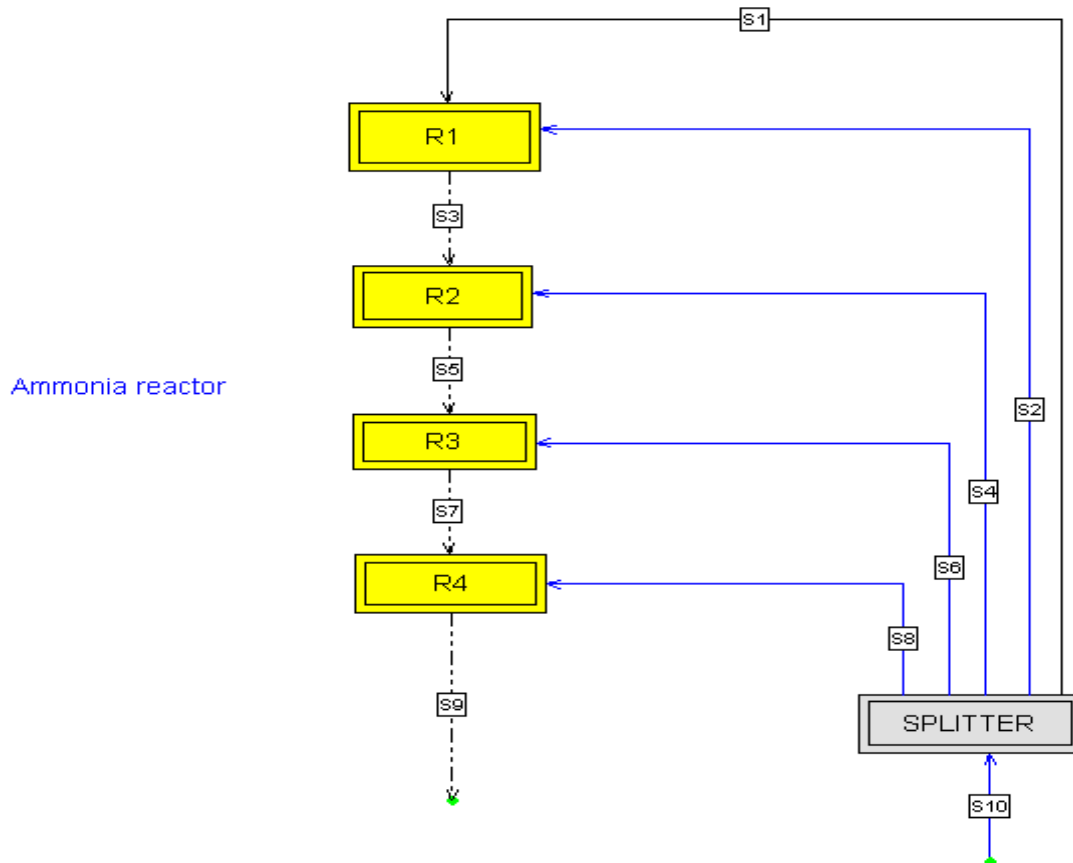


Fig. 4.7-1: Balance flow-sheet (demo example MC-8)

The synthesis gas (stream S10) consists of four components:

Component No	Formula
1	H ₂
2	N ₂
3	NH ₃
4	CH ₄ Ar

The pseudocomponent CH₄Ar represents inert species in the synthesis gas, mostly methane and rare gases (stemming from the technology of its preparation).

The ammonia reactor has four beds (sections). The reactants from sections are quenched by introducing parts of the cold raw synthesis gas (streams S2, S4, S6 and S8). There is only one chemical reaction in the system:



Let us now go back to the original problem formulation in the RECON program. We first define the components occurring in the problem.

This reaction must be put into the reaction bank and associated with reactors R1, R2, R3 and R4 in the way described in the preceding example.

There are 6 streams incident with the SPLITTER node (S1, S2, S4, S6, S8 and S10). These streams have the same composition. We can select for example the S1 stream as the *master* stream, the 5 remaining streams will be so-called *dependent* streams – see Subsection 2.3.2 (the choice of the master stream is arbitrary and does not influence the results).

Concentrations are entered only for the master stream, the dependent streams are only marked as dependent on their panel of parameters.

Stream : S2

From node: **SPLITTER** Dependent stream
to node: **R1** S1

Mass flow [MOL/S]

Sort of stream: Material

Stream	Description	Type	Value	Max.error
S2		M	655	5%

Fig. 4.7-2: Panel for defining a dependent stream

Further, we have to enter the values of flowrates and concentrations.

M A T E R I A L S T R E A M S [MOL/S], [%]

ID	COMPONENT	TYPE	VALUE	MAX.ERROR
S1	Flowrate	M	655.0000	5.0000 %
	H2	M	63.800000	2.000000 %
	N2	M	21.100000	2.000000 %
	NH3	M	3.000000	2.000000 %
	CH4AR	M	12.000000	2.000000 %
S10	Flowrate	M	2300.0000	5.0000 %
Composition identical with stream 'S1'				
S2	Flowrate	M	655.0000	5.0000 %
Composition identical with stream 'S1'				
S3	Flowrate	N	1200.0000	
	H2	M	57.200000	2.000000 %
	N2	M	19.000000	2.000000 %
	NH3	M	10.900000	2.000000 %
	CH4AR	M	12.900000	2.000000 %
S4	Flowrate	M	330.0000	5.0000 %
Composition identical with stream 'S1'				
S5	Flowrate	N	1400.0000	
	H2	M	53.400000	2.000000 %
	N2	M	17.700000	2.000000 %
	NH3	M	15.400000	2.000000 %
	CH4AR	M	13.500000	2.000000 %
S6	Flowrate	M	360.0000	5.0000 %
Composition identical with stream 'S1'				
S7	Flowrate	N	1800.0000	
	H2	M	51.700000	2.000000 %
	N2	M	17.100000	2.000000 %

	NH3	M	17.500000	2.000000 %
	CH4AR	M	13.700000	2.000000 %
S8	Flowrate	M	300.0000	5.0000 %
Composition identical with stream 'S1'				
S9	Flowrate	N	2000.0000	
	H2	M	51.700000	2.000000 %
	N2	M	16.800000	2.000000 %
	NH3	M	18.700000	2.000000 %
	CH4AR	M	13.800000	2.000000 %

The data input is thus finished. The results of the data reconciliation are following (abbreviated):

GLOBAL DATA

Number of nodes	5
Number of streams	10
Number of components	4
Number of reactions :	1
Number of react. nodes :	4
Number of measured variables	26
Number of adjusted variables	26
Number of non-measured variables	8
Number of observable variables	8
Number of unobservable variables	0
Number of free variables	0
Number of equations	22
Number of independent equations	22
Number of user equations	0
Degree of redundancy	14

Mean residue of equations	5.8649E-12
Qmin	3.5910E+00
Qcrit	2.3697E+01
Status (Qmin/Qcrit)	0.151541

STREAMS [MOL/S], [%]

Stream name: S1

From node 'SPLITTER' to node 'R1'

No.	Name	Type	Inp.value	Rec.value	Abs.error
	Flowrate	MC	655.000	654.835	31.647
1	H2	MC	63.800	63.854	0.175
2	N2	MC	21.100	21.149	0.142
3	NH3	MC	3.000	3.000	0.060
4	CH4AR	MC	12.000	11.996	0.106

Stream name: S2

From node 'SPLITTER' to node 'R1'

Composition identical with stream 'S1'

No.	Name	Type	Inp.value	Rec.value	Abs.error
	Flowrate	MC	655.000	654.835	31.647

Stream name: S3

From node 'R1' to node 'R2'

No.	Name	Type	Inp.value	Rec.value	Abs.error
	Flowrate	NO	1200.000	1216.440	40.141
1	H2	MC	57.200	57.252	0.254
2	N2	MC	19.000	18.938	0.161
3	NH3	MC	10.900	10.895	0.212
4	CH4AR	MC	12.900	12.916	0.116

Etc.

EXTENTS OF REACTIONS

NODE	REACTION	EXTENT
R1	R1	46.615
R2	R1	41.528
R3	R1	35.245
R4	R1	27.795

4.8 Multicomponent balance of a flotation circuit

This example from mineral industry describes the multi-component balance of a flotation circuit (adapted from H.W.Smith and N.M.Ichiyen: *Computer Adjustment of Metallurgical Balances. CIM Bulletin for September, 1973, pp. 97-100*).

Copper and zinc are separated from suspensions by flotation in the following process:

Flotation circuit

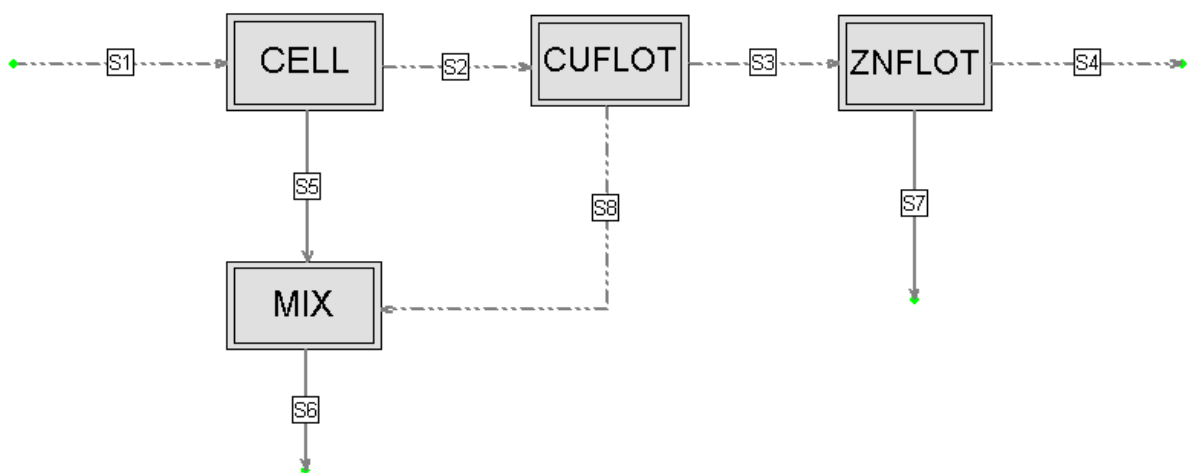


Fig. 4.8-1: Balance flow-sheet (demo example MC-9)

The composition of all streams is simplified to 3 components:

Component No	Formula
1	Cu
2	Zn
3	REST

The pseudocomponent REST represents the remaining elements in the system, including water. The concentration of the REST is always unmeasured and completes the composition to 100 %.

Further are values of flowrates and concentrations.

MATERIAL STREAMS			[T/H], [%]	
ID	COMPONENT	TYPE	VALUE	MAX.ERROR
S1	Flowrate	N	9.0000	
	CU	M	5.210000	2.00000E-2
	ZN	M	5.410000	2.00000E-2
	REST	N	95.000000	

```

S2      Flowrate  N      8.0000
        CU        M      0.450000  5.00000E-2
        ZN        M      4.920000  0.200000
        REST     N      95.000000

S3      Flowrate  N      7.0000
        CU        M      0.130000  2.00000E-2
        ZN        M      5.120000  0.200000
        REST     N      95.000000

S4      Flowrate  N      6.0000
        CU        M      9.00000E-2  1.00000E-2
        ZN        M      0.410000  2.00000E-2
        REST     N      99.000000

S5      Flowrate  M      2.2100  10.0000 %
        CU        M      19.860000  2.000000
        ZN        M      7.090000  1.000000
        REST     N      75.000000

S6      Flowrate  M      2.5000  10.0000 %
        CU        M      21.440000  2.000000
        ZN        N      4.950000
        REST     N      75.000000

S7      Flowrate  M      0.6600  10.0000 %
        CU        M      0.510000  2.00000E-2
        ZN        M      52.100000  2.000000
        REST     N      50.000000

S8      Flowrate  N      0.3500
        CU        M      21.600000  1.000000
        ZN        M      2.100000  2.000000
        REST     N      96.000000

```

The data input is thus finished. The results of the data reconciliation are following (abbreviated):

G L O B A L D A T A

```

Mean residue of equations      2.4463E-11
Qmin                          3.9550E+00
Qcrit                         1.2584E+01
Status (Qmin/Qcrit)           0.314289

```

```

S T R E A M S   [T/H], [%]
Stream name: S1
From node 'ENVIRON' to node 'CELL'

```

No.	Name	Type	Inp.value	Rec.value	Abs.error
	Flowrate	NO	9.000	9.714	0.673
1	CU	MC	5.210	5.210	0.020
2	ZN	MC	5.410	5.410	0.020
3	REST	NO	95.000	89.380	0.028

```

Stream name: S2
From node 'CELL' to node 'CUFLOT'

```

No.	Name	Type	Inp.value	Rec.value	Abs.error
	Flowrate	NO	8.000	7.429	0.593
1	CU	MC	0.450	0.453	0.050
2	ZN	MC	4.920	4.983	0.126
3	REST	NO	95.000	94.564	0.134

```

Stream name: S3
From node 'CUFLOT' to node 'ZNFLOT'

```

No.	Name	Type	Inp.value	Rec.value	Abs.error
	Flowrate	NO	7.000	7.316	0.583
1	CU	MC	0.130	0.128	8.53E-3
2	ZN	MC	5.120	5.035	0.128
3	REST	NO	95.000	94.837	0.129

5 Models of thermal unit operations

In the previous chapter, we have given several examples of mass and species balances. With the energy balances, the situation is more complicated for a number of reasons. Besides the unavoidable energy losses into the environment, there is another problem. It is the modeling of the mass streams enthalpy dependency on the state variables; indeed, our knowledge about these dependencies is never perfect. That's why already the name of the chapter speaks of models.

Further, we'll describe models of basic heat operations. From such units, one can then set up the models of more complex systems. The present chapter will enable the reader to create simple models in the program RECON and check himself the results.

The structure of all sections is as follows.

1. **BALANCING SCHEME.** It is a copy of the scheme created in the graphical editor of the RECON program. To individual streams are in addition written identifiers of state variables (T for temperature, P for pressure and X for wetness) .If a given variable is denoted by these letters, unmeasured variables are given in italics. Details can be found in the part INPUT DATA of the example.
2. **INPUT DATA.** This is a somewhat abridged extract of data for the problem created in program RECON – menu *Flowsheet – Data review*. It comprises all information needed for the problem configuration.
3. **PANEL OF THE NODE MODEL PARAMETERS DEFINITION.** It is the panel, where the node parameters for energy balance are defined. In case of more complex schemes, there can be more such panels.
4. **RESULTS.** It is an abridged extract from the results for the user's inspection, created in program RECON – menu *Calculate – Results*. All examples in this chapter are without gross errors, so the sum of squares of adjustments thus Q_{min} values are not given.

Although this manual is not conceived as a substitute for instructions to the RECON program, let us still indicate a procedure to be maintained by the reader when creating the problems (tasks).

1. Enter the name of the task, which is simultaneously the name of the file for the model to be created.
2. Enter the text of the file title (long name).
3. In the further panel, change physical units (when necessary). The units chosen in individual cases are given in part INPUT DATA.
4. Enter the component name. It is recommended to enter H₂O and not to fill in the full name.
5. The graphical editor surface turns up. Before one starts drawing, it is recommended to enter the state variables values (temperatures, pressures and wetnesses) in the menu *Accessories*. Names, types and values can be found in part INPUT DATA. In the standard manner are given beforehand the wetnesses STEAM (wetness = 0) and WATER (wetness = 100).

6. The set up of models including energy balance is a little bit more difficult than for models with a mass balance only. It is recommended to start with the mass balance part of the model, which is usually a basis of energy balancing. After checking the mass balance part of the model, energy balance functions can be added gradually.
7. The scheme drawing proper is recommended to be started by drawing all nodes; the latter are to be conveniently placed on the screen (at later changes in size and placing, one can disturb the streams already drawn). At the nodes, one fills in only Name (it is in the scheme) and Description (invent some).
8. Then draw the streams. Their short names are given in the scheme, a description is again invented. The types of streams, values and possible errors (for measured variables) can be found in part INPUT DATA. If one deals with an energy stream, this must be marked in the panel of the stream on the right above. Thereby, the definition of the problem concerning the mass balance is complete.
9. In order to create the energy balance equation for the given node, we must designate this node on its panel as a *heat node* (square on the right above). In this manner, one makes the center of the panel accessible for the configuration of heat functions, temperatures, pressures and wetnesses. **At this moment, these state variables must naturally be already defined.** Let us recall that one fills in only those state variables that are relevant for the given heat function. So as to explain this important part, all panels are figured in the text.
10. Having finished the configuration of all heat nodes, one can carry out the computation.

Let us still explain certain abbreviations used in the RECON program.

dT	mean temperature difference in heat exchanger
F	type of variable - fixed variable (known as errorless)
HTC	heat transfer coefficient
M	type of variable - Measured variable
MC	type of variable – Measured variable, adjustable (Corrected by data reconciliation)
MN	type of variable – measured variable, nonadjustable
MPag	unit of pressure (final g means pressure difference against atmospheric pressure, the same for further pressure units)
N	type of variable – uNmeasured variable
NO	type of variable – uNmeasured variable, Observable
NN	type of variable – uNmeasured variable, uNobservable (non-observable)
Q	heat flow

Let us note in addition that certain examples have null degree of redundancy and one deals then only with solving the set of equations. For simple cases formed by one or several nodes, this is typical in practice. Redundancy arises just by connecting simple unit operations in larger systems.

For the majority of examples given in this chapter, the reader will have at hand also their model solutions, i.e. files with respective models. Names of the files correspond to numbers of sections in this chapter. E.g. to example from Section 5.1 corresponds Example E-1, to example from Section 5.2 example E-2, etc.

5.1 Mixer of thermal streams

Mixing two or several streams is a basic unit operation. The model generates two balance equations – the mass and energy balances. From the model, one can compute two unmeasured variables at most, e.g. the outlet flowrate and temperature.

The following example is that of mixing two streams of water with different temperatures. Before creating the scheme in the graphical editor, the user must define 3 temperatures and 1 pressure (menu *Accessories*). The pressure in the whole system is that of the atmosphere, and with respect to the temperatures, the water is below the saturation line. The model is formed by two equations (mass and energy balances). Because all variables are measured, the problem has two degrees of redundancy.

Start by drawing the flowsheet for mass balance only (node M1 and streams S1, S2 and S3). At this moment the color of the node will be gray. After checking the box *Heat node* on the node's panel, fill in heat functions and their parameters. After that the color of the node will change, showing that the heat balance is active.

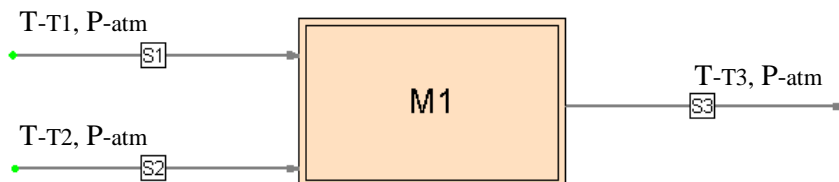


Fig. 5.1-1: Balance scheme (demo Example E-1)

INPUT DATA

NAME	TYPE	VALUE	MAX. ERROR
M A T E R I A L S T R E A M S [KG/S]			
S1	M	60.0000	1.0000
S2	M	40.0000	2.0000
S3	M	102.0000	2.0000
T E M P E R A T U R E S [C]			
T1	M	60.0000	1.0000
T2	M	40.0000	1.0000
T3	M	51.0000	1.0000
P R E S S U R E S [KPA]			
atm	F	101.3250	

Node: M1

ID: Description:

Above sea level: Node pres.:

Sort of calculations:

- Balancing
- Hydraulic node
- Heat node
- Reaction node

Non-energetic streams incident with node

Stream	Function	Temperature	Pressure	Wetness
S1	H2O(T,P)	T1	atm	
S2	H2O(T,P)	T2	atm	
S3	H2O(T,P)	T3	atm	

Reactions in node

Reaction:

Fig. 5.1-2: Panel of node model definition

RESULTS (abbreviated)

GLOBAL DATA

Degree of redundancy	2
Number of equations	2
Qmin	3.764E+00
Qcrit	5.974E+00

VARIABLES

Name	Type	Inp. value	Rec. value	Max. error
S T R E A M S [KG/S]				
S1	MC	60.000	60.148	0.938
S2	MC	40.000	41.053	1.472
S3	MC	102.000	101.201	1.484
T E M P E R A T U R E S [C]				
T1	MC	60.000	59.653	0.880
T2	MC	40.000	39.769	0.946
T3	MC	51.000	51.589	0.600
P R E S S U R E S [KPA]				
atm	F	101.325	101.325	

The mixer is a basic node type. In practice, it can have several inlets, and also more outlets. For a mixer in proper sense, all outlet streams should be in the same thermodynamic state.

5.2 Simple heat exchanger

This kind of exchanger is characterized by the fact that it has only two streams, which exchange heat. One stream is so-called *hot* and the other *cold* stream. This model generates just one equation (heat balance). It is worth mentioning that with this model, one gives the heat exchange area and number of passes. From the model one then computes, besides the heat flow in the exchanger, also the overall heat transfer coefficient.

In our example, one deals with the heating/cooling of water with temperature below the saturation line and under atmospheric pressure. Before creating the scheme in graphical editor, one has to enter 4 temperatures and 1 pressure.

After that draw both streams. You will be warned that these streams are loops going from Environment to environment. This is allowed for heat exchangers only. After that draw the exchanger E1 on the crossing of streams.

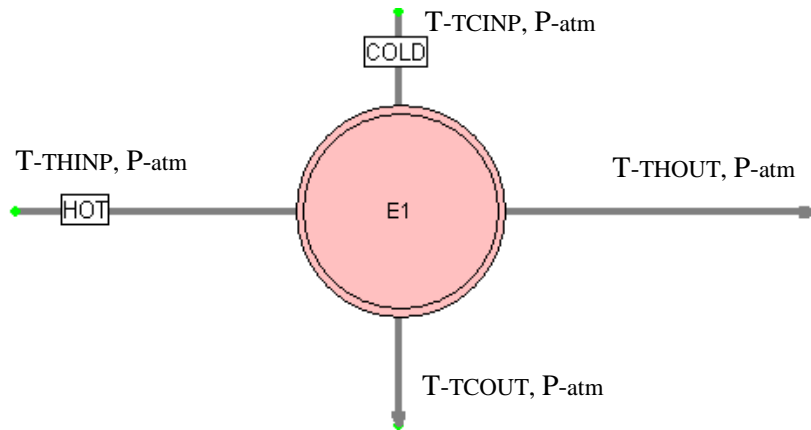


Fig. 5.2-1: Balance scheme (demo Example E-2)

INPUT DATA

NAME	TYPE	VALUE	MAX. ERROR
M A T E R I A L S T R E A M S [t/h]			
COLD	M	20.0000	2.0000 %
HOT	M	10.0000	2.0000 %
T E M P E R A T U R E S [C]			
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]			
atm	F	100.0000	

E X C H A N G E R S [M^2]

NAME	Stream	End	Function	Temperature	Pressure	Wetness	Area
E1	HOT	Inlet	H2O(T,P)	THINP	atm		100
		Outlet	H2O(T,P)	THOUT	atm		
	COLD	Inlet	H2O(T,P)	TCINP	atm		
		Outlet	H2O(T,P)	TCOUT	atm		

Exchanger : E1 Balancing

ID	Description	Area [M^2]	Passes
E1	Simple hear exchanger	100	1 - 2

Exchanger parameters:

HOT Hot stream

	Function	Temperature	Pressure	Wetness
Inlet	H2O(T,P)	THINP	atm	
Outlet	H2O(T,P)	THOUT	atm	

COLD Cold stream

	Function	Temperature	Pressure	Wetness
Inlet	H2O(T,P)	TCINP	atm	
Outlet	H2O(T,P)	TCOUT	atm	

Fig. 5.2-2 Panel of the model parameters definition

RESULTS

GLOBAL DATA

Degree of redundancy	1
Number of equations	1
Qmin	1.481E+00
Qcrit	3.841E+00

VARIABLES

Name	Type	Inp.value	Rec.value	Max.error
S T R E A M S [t/h]				
COLD	MC	20.000	20.056	0.389
HOT	MC	10.000	9.970	0.194
T E M P E R A T U R E S [C]				
TCINP	MC	20.000	19.629	0.802
TCOUT	MC	39.000	39.370	0.803
THINP	MC	90.000	89.814	0.954
THOUT	MC	50.000	50.185	0.955
P R E S S U R E S [KPA]				
atm	F	100.000	100.000	

EXCHANGERS [MJ/h], [C], [MJ/h/M^2/C]

Name	Q(hot)	Q(cold)	Q(rec)	dT	HTC
E1	1675.797	1588.655	1655.290	39.672	0.417

Here, Q(hot) and Q(cold) mean transferred heats calculated from hot and cold streams separately, while Q(rec) is heat resulting from reconciled values.

5.3 General heat exchanger

A general heat exchanger differs from the simple one mainly by the fact that it can comprise more than two streams. It is formed by two balancing nodes, one of which represents the hot side, the other one the cold side of the exchanger. Both nodes are connected by an energy stream representing the heat flow from the hot to the cold side through the heat exchange surface. This arrangement makes possible to model situations, for the description of which the simple exchanger does not suffice.

The model generates altogether 4 balance equations – the mass and energy balances around both of the nodes. If compared with the simple exchanger, it has now three unknowns – the heat flow through the exchanger and two unknown flowrates at the outlets from both nodes. One degree of redundancy is available for the reconciliation, so long as the other variables are known (thus in the same manner as in the case of a simple exchanger).

Let us start from the preceding example, i.e. the heating/cooling of water with temperature below the saturation line at atmospheric pressure. In addition, let us consider the heat loss from the cold side, say QLOSS estimated a priori as 20 MJ/h; the error of this estimate is assumed to be 4 MJ/h.

Before drawing the scheme, it is necessary to define 4 temperatures and 1 pressure.

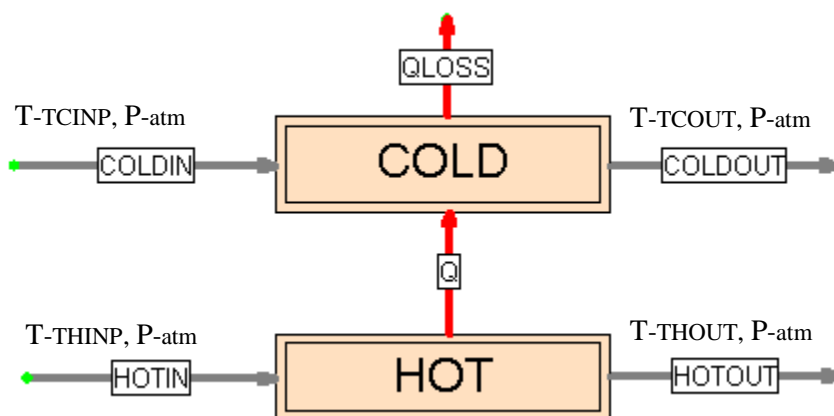


Fig. 5.3-1: Balance scheme (demo Example E-3)

Comparing with a simple exchanger, we see that both parts of the exchanger have now two incident streams – inlet and outlet. If some of the flowrates is measured then only once, e.g. at the inlet. The outlet flowrate is then calculated from the mass balance.

INPUT DATA

NAME	TYPE	VALUE	MAX.ERROR
M A T E R I A L S T R E A M S [t/h]			
COLDIN	M	20.0000	2.0000 %
COLDOUT	N	20.0000	
HOTIN	M	10.0000	2.0000 %
HOTOUT	N	10.0000	
F L O W S O F E N E R G Y [MJ/h]			
Q	N	1000.0000	
QLOSS	M	20.0000	4.0000
T E M P E R A T U R E S [C]			
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]
atm        F      100.0000

```

PANELS OF THE MODEL PARAMETERS DEFINITION

Node: COLD

Sort of calculations:

- Balancing
- Hydraulic node
- Heat node
- Reaction node

ID	Description
COLD	cold side of a heat exchanger

Above sea level: Node pres.:

Non-energetic streams incident with node

Stream	Function	Temperature	Pressure	Wetness
COLDIN	H2OL(T,P)	TCINP	atm	
COLDOUT	H2OL(T,P)	TCOUT	atm	

Reactions in node: Reaction

Fig. 5.3-2: Cold side of exchanger

Node: HOT

Sort of calculations:

- Balancing
- Hydraulic node
- Heat node
- Reaction node

ID	Description
HOT	hot side of a heat exchanger

Above sea level: Node pres.:

Non-energetic streams incident with node

Stream	Function	Temperature	Pressure	Wetness
HOTIN	H2OL(T,P)	THINP	atm	
HOTOUT	H2OL(T,P)	THOUT	atm	

Reactions in node: Reaction

Fig. 5.3-3: Hot side of exchanger

RESULTS

GLOBAL DATA

```

Degree of redundancy      1
Number of equations      4
Qmin                      8.792E-01
Qcrit                     3.841E+00

```

VARIABLES

Name	Type	Measured	Reconciled	Max.error of reconciled val.
S T R E A M S [t/h]				
COLDIN	MC	20.000	20.043	0.389
COLDOUT	NO	20.000	20.043	0.389
HOTIN	MC	10.000	9.977	0.194
HOTOUT	NO	10.000	9.977	0.194
F L O W S O F E N E R G Y [MJ/h]				
Q	NO	1000.000	1659.996	59.427

QLOSS	MC	20.000	20.055	3.998
T E M P E R A T U R E S [C]				
TCINP	MC	20.000	19.714	0.802
TCOUT	MC	39.000	39.285	0.803
THINP	MC	90.000	89.857	0.954
THOUT	MC	50.000	50.143	0.955
P R E S S U R E S [KPA]				
atm	F	100.000	100.000	

The results can be compared with the preceding example of a simple exchanger. The only difference is here the energy loss stream. In the case of a general heat exchanger, the mean temperature difference is no longer computed, nor the overall heat transfer coefficient. If the user needs these values, he must define them himself with the aid of user defined equations.

5.4 Steam generator

In practice, steam can be generated in different manners (e.g. by hot combustion products, other steam of higher pressure, or a hot heat exchanging medium). The thermodynamic state at the boiling proper inside the generator can be defined by temperature or pressure. One here assumes phase equilibrium between liquid and vapor phases.

This unit operation can be, according to its complexity, modeled as a simple or general heat exchanger. As the steam generator in a nuclear power plant will be the subject of a more extensive case study below, let us now prepare this model for further use.

Heat is supplied to the steam generator (SG) by high-pressure hot water circulating between the nuclear reactor and the high pressure water part of the SG. Into the steam part is supplied feed water; steam and blowdown water form the outlets. The steam contains a small amount of liquid phase. Because we here have more than 2 streams, let us apply the general heat exchanger model.

In this model, altogether 4 balance equations are generated. If we measure the flowrates of all mass streams connected with the steam space and the hot water flowrate at the inlet into the high-pressure water space, we have altogether 2 unknown streams, viz. hot water outlet and heat flow. If we further measure all temperatures and pressures, two degrees of redundancy are available for reconciliation and data validation.

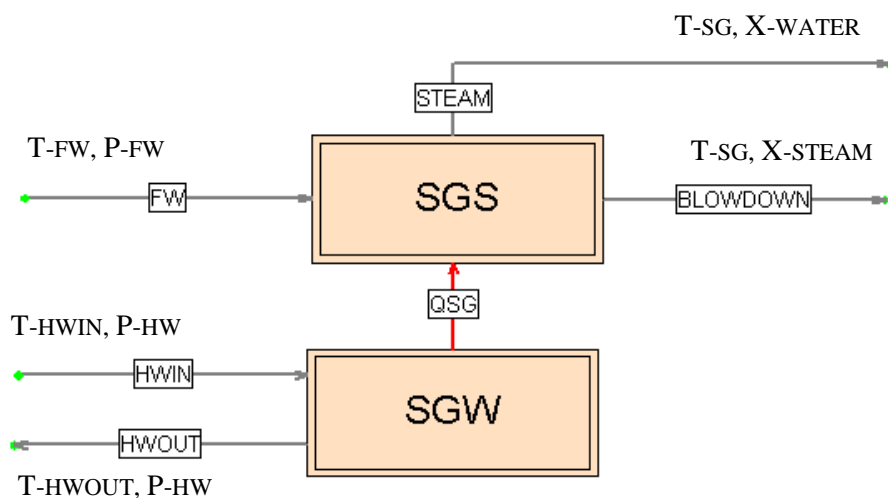


Fig. 5.4-1: Balance scheme (demo Example E-4)

INPUT DATA

Besides the mass and heat flowrates, the problem involves 4 temperatures (hot water temperatures HWIN and HWOUT, temperature in steam generator SG, equal for outlet steam and blowdown, and temperature of feed water FW). Further involved are two pressures (FW for feed water and HW for hot water). In addition, we here have two wetness values for liquid water (WATER) and (wet) steam (STEAM).

NAME	TYPE	VALUE	MAX. ERROR
M A T E R I A L S T R E A M S [KG/S]			
BLOWDOWN	M	6.1200	5.0000 %
FW	M	444.5000	2.0000 %
HWIN	M	5650.0000	5.0000 %
HWOUT	N	5000.0000	
STEAM	M	445.0000	2.0000 %
E N E R G Y F L O W S [KJ/S]			
QSG	N	800000.0000	
T E M P E R A T U R E S [C]			
HWIN	M	295.2000	1.0000
HWOUT	M	265.8000	1.0000
SG	M	257.6000	1.0000
FW	M	221.6000	1.0000
P R E S S U R E S [KPA]			
FW	M	10000.0000	0.5000 %
HW	M	10000.0000	0.5000 %
W E T N E S S E S [%]			
STEAM	F	0.2500	
WATER	F	100.0000	

PANELS OF MODEL PARAMETERS DEFINITION

Node: SGS

ID	Description
SGS	Steam generator steam space

Above sea level: Node pres.:

Sort of calculations:

- Balancing
- Hydraulic node
- Heat node
- Reaction node

Non-energetic streams incident with node				
Stream	Function	Temperatur	Pressure	Wetness
BLOWDOWN	H2O(T,X)	SG		WATER
FW	H2O(T,P)	FW	FW	
STEAM	H2O(T,X)	SG		STEAM

Reactions in node: Reaction

Fig. 5.4-2: Steam space

Node: SGW

ID	Description
SGW	steam generator water side

Above sea level: Node pres.:

Sort of calculations:

- Balancing
- Hydraulic node
- Heat node
- Reaction node

Non-energetic streams incident with node				
Stream	Function	Temperatur	Pressure	Wetness
HWIN	H2O(T,P)	HWIN	HW	
HWOUT	H2O(T,P)	HWOUT	HW	

Reactions in node: Reaction

Fig. 5.4-3: Water space

RESULTS

GLOBAL DATA

Degree of redundancy	2
Number of equations	4
Qmin	4.001E+00
Qcrit	5.974E+00

VARIABLES

Name	Type	Inp.value	Rec.value	Max.error
S T R E A M S [KG/S]				
BLOWDOWN	MC	6.120	6.115	0.306
FW	MC	444.500	448.863	6.172
HWIN	MC	5650.000	5471.834	200.270
HWOUT	NO	5000.000	5471.834	200.270
STEAM	MC	445.000	442.748	6.172
E N E R G Y F L O W S [KJ/S]				
QSG	NO	800000.000	816004.219	11530.738
T E M P E R A T U R E S [C]				
FW	MC	221.600	221.570	0.999
HWIN	MC	295.200	294.745	0.863
HWOUT	MC	265.800	266.161	0.890
SG	MC	257.600	257.597	1.000
P R E S S U R E S [KPA]				
FW	MC	10000.000	9999.995	50.000
HW	MC	10000.000	10000.159	50.000
W E T N E S S E S [%]				
STEAM	F	0.250	0.250	
WATER	F	100.000	100.000	

The steam generator balancing is very important in practice. This example will again be scrutinized from other points of view in one of the Case studies below.

5.5 Condensation of steam

During the condensation, steam transfers its condensation heat to the cold stream, which is mostly liquid of lower temperature. So long as just this heating is the purpose, the whole apparatus is called heater. The steam in the condenser changes to condensate, which can be water at the saturation line, possibly also somewhat subcooled (in this section, sub-cooling will not be considered for simplicity). The difference between the inlet steam and outlet condensate enthalpies determines the amount of heat transferred.

The state at the condensation proper inside the exchanger can be defined either by temperature or by pressure. One here assumes phase equilibrium between steam and condensate. A heat exchanger with condensation can be, according to its complexity, modeled as a simple or general exchanger. Let us further use the general form of exchanger.

The example describes a high-pressure heater of feed water heated by steam withdrawn as side stream from a high-pressure turbine. With this arrangement, only feed water flowrate is usually measured. One further measures the withdrawn steam (STEAM) pressure (which approximates the pressure in the steam space of the heater) and the feed water temperatures before (FW-IN) and after (FW-OUT) the heater. Further known is also the wetness of the steam withdrawn.

This model generates 4 balance equations, but simultaneously we here have 4 unknowns (flowrates of steam, condensate and outlet feed water, and the heat flow). The degree of redundancy is 0 and no data reconciliation takes place. At this measurement constellation, the model is only able to compute all the unmeasured variables.

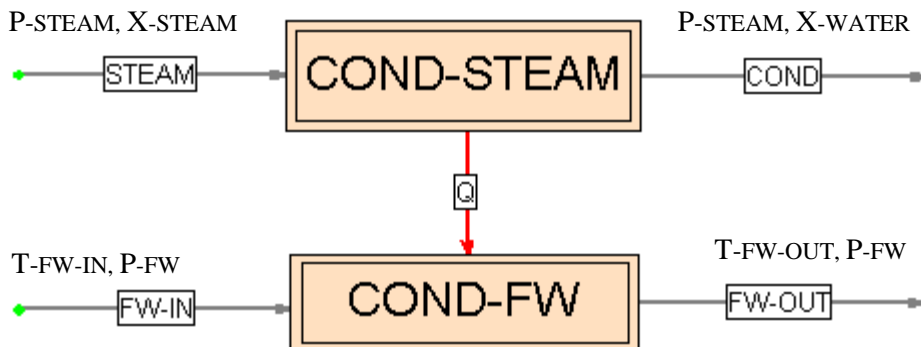


Fig. 5.5-1: Balance scheme (demo Example E-5)

INPUT DATA

NAME	TYPE	VALUE	MAX.ERROR
M A T E R I A L S T R E A M S [T/H]			
COND	N	60.0000	
FW-IN	M	1320.0000	2.0000 %
FW-OUT	N	1300.0000	
STEAM	N	60.0000	
E N E R G Y F L O W S [GJ/H]			
Q	N	60.0000	

```

T E M P E R A T U R E S      [C]
FW-IN      M      191.0000      1.0000
FW-OUT     M      223.0000      1.0000

P R E S S U R E S      [MPAG]
FW         M      6.5000      0.1000
STEAM     M      2.8400      2.000E-2

W E T N E S S E S      [%]
steam     F      4.6000
water     F      100.0000

```

PANEL OF MODEL PARAMETERS DEFINITION

Node: COND-STEAM

Sort of calculations:

- Balancing
- Hydraulic node
- Heat node
- Reaction node

ID	Description
COND-STEAM	condenser - steam side

Above sea level: Node pres.:

Non-energetic streams incident with node

Stream	Function	Temperature	Pressure	Wetness
COND	H2O(P,X)		STEAM	water
STEAM	H2O(P,X)		STEAM	steam

Reactions in node: Reaction

Fig. 5.5-2: Steam side

Node: COND-FW

Sort of calculations:

- Balancing
- Hydraulic node
- Heat node
- Reaction node

ID	Description
COND-FW	condenser - feed water side

Above sea level: Node pres.:

Non-energetic streams incident with node

Stream	Function	Temperature	Pressure	Wetness
FW-IN	H2O(L,T,F)	FW-IN	FW	
FW-OUT	H2O(L,T,F)	FW-OUT	FW	

Reactions in node: Reaction

Fig. 5.5-3: Feed water side

RESULTS

```

G L O B A L   D A T A
Degree of redundancy      0
Number of equations      4
Qmin                     0.000E+00
Qcrit                    0.000E+00

```

V A R I A B L E S

Name	Type	Inp.value	Rec.value	Max.error
S T R E A M S [T/H]				
COND	NO	60.000	110.800	5.383
FW-IN	MN	1320.000	1320.000	26.400
FW-OUT	NO	1300.000	1320.000	26.400
STEAM	NO	60.000	110.800	5.383
F L O W S O F E N E R G Y [GJ/H]				
Q	NO	60.000	190.266	9.243
T E M P E R A T U R E S [C]				
FW-IN	MN	191.000	191.000	1.000
FW-OUT	MN	223.000	223.000	1.000
P R E S S U R E S [MPAG]				
FW	MN	6.500	6.500	0.100
STEAM	MN	2.840	2.840	0.020
W E T N E S S E S [%]				
steam	F	4.600	4.600	
water	F	100.000	100.000	

In this case, one has not dealt with data reconciliation, but only with direct computation (the degree of redundancy was 0). It is obvious that the values of all measured variables before and after the procedure must remain the same. However, this does not hold for the unmeasured ones. Quite often, at the problem configuration the values of unmeasured variables are not known even approximately. There then arises the question, what the software will do with bad initial guesses of the unmeasured variables.

Special attention is required in cases where the enthalpy function does not uniquely determine the temperature or pressure (the enthalpy of steam as dependent on temperature or pressure passes through the maximum and two values of temperature or pressure can correspond to one enthalpy value).

5.6 Expander

Expanders serve for reducing the pressure of hot water to a lower value (frequently as part of a cascade of condensate from the steam system). Some part of the heat energy then gives rise to steam phase by evaporation. Then from one liquid stream, two streams are formed – water and steam. We here assume that the two streams are in the state of thermodynamic equilibrium. In addition the situation is complicated by the fact that the steam phase can contain some portion of liquid as a consequence of imperfect water droplet separation (this depends on the construction of the expander).

The following example will be simple – condensate on the saturation line expands from higher to lower pressure. The condensate is defined by the temperature, in the expander proper is maintained constant pressure. The arising steam contains 0.2 % liquid phase. From the flowrates, only inlet into the expander is measured.

The model generates two balance equations and we here have two unmeasured outlet flowrates. The degree of redundancy is thus 0. The model only serves for computing the phase distribution in the expander.

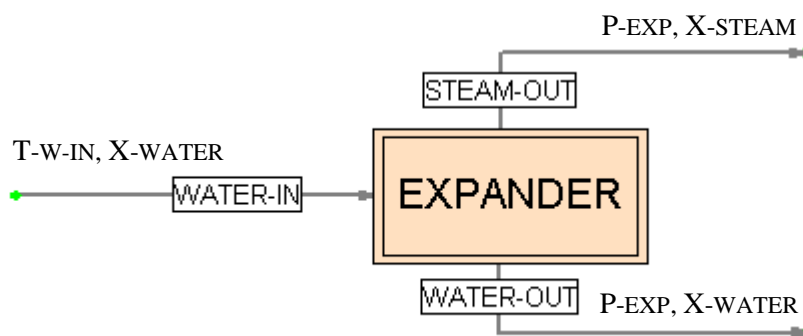


Fig. 5.6-1: Balance scheme (demo Example E-6)

INPUT DATA

NAME	TYPE	VALUE	MAX. ERROR
M A T E R I A L S T R E A M S [T/H]			
STEAM-OUT	N	2.0000	
WATER-IN	M	26.0000	5.0000 %
WATER-OUT	N	20.0000	
T E M P E R A T U R E S [C]			
W-IN	M	258.0000	1.0000
P R E S S U R E S [MPAG]			
EXP	M	0.6600	5.000E-3
W E T N E S S E S [%]			
steam	F	0.200E+0	
water	F	100.0000	

Node: EXPANDER

ID: EXPANDER Description: expander

Above sea level: Node pres.:

Sort of calculations:

- Balancing
- Hydraulic node
- Heat node
- Reaction node

Non-energetic streams incident with node

Stream	Function	Temperature	Pressure	Wetness
STEAM-OU	H2O(P,X)		EXP	steam
WATER-IN	H2O(T,X)	W-IN		water
WATER-OU	H2O(P,X)		EXP	water

Reactions in node

Reaction: + -

Fig. 5.6-2: Panel of the model parameters definition

RESULTS

GLOBAL DATA

Degree of redundancy	0
Number of equations	2
Qmin	0.000E+00
Qcrit	0.000E+00

VARIABLES

Name	Type	Measured	Reconciled	Max.error of reconciled val.
STREAMS [T/H]				
STEAM-OUT	NO	2.000	5.236	0.269
WATER-IN	MN	26.000	26.000	1.300
WATER-OUT	NO	20.000	20.764	1.041
TEMPERATURES [C]				
W-IN	MN	258.000	258.000	1.000
PRESURES [MPAG]				
EXP	MN	0.660	0.660	5.00E-3
WETNESSES [%]				
steam	F	0.00E+0	0.20E+0	
water	F	100.000	100.000	

5.7 Throttling of steam

Throttling of steam is an adiabatic process, where the steam pressure is reduced by an obstacle in the flow. The throttling organ can have a fixed cross section area (e.g. an orifice), or a variable one (control valve). If heat loss into the environment is neglected, one can model in this way also pressure drop in flow of fluids in a pipeline.

The model generates 2 balance equations (mass and energy balances). In the following example, we'll have 2 unknown variables – flowrate and wetness of the steam at the throttling organ outlet. One will thus again deal with mere solution of the model equations without data reconciliation.

Concerning the stream energies, also their kinetic components may play a role (due to the pressure drop, also the stream velocity is changed). In case of wet steam, also wetness is changed. Although the operation is simple, the results of computation need not be trivial as the following examples will show.



Fig. 5.7-1: Balance scheme (demo Example E-7)

INPUT DATA

NAME	TYPE	VALUE	MAX.ERROR
M A T E R I A L S T R E A M S [T/H]			
STEAM-IN	M	440.0000	4.0000 %
STEAM-OUT	N	400.0000	
P R E S S U R E S [MPAG]			
IN	M	2.6500	5.000E-2
OUT	M	2.3500	5.000E-2
W E T N E S S E S [%]			
STEAM-IN	F	0.2500	
STEAM-OUT	N	0.2500	

Node: THR

ID	Description
THR	throttling element

Above sea level Node pres.

Sort of calculations:

- Balancing
- Hydraulic node
- Heat node
- Reaction node

Non-energetic streams incident with node

Stream	Function	Temperature	Pressure	Wetness
STEAM-IN	H2O(P,X)		IN	STEAM-I
STEAM-OU	H2O(P,X)		OUT	STEAM-C

Reactions in node

Reaction

Fig. 5.7-2: Panel of model parameters definition

RESULTS

GLOBAL DATA

```

Degree of redundancy          0
Number of equations          2
Qmin                          0.000E+00
Qcrit                         0.000E+00
  
```

VARIABLES

Name	Type	Inp.value	Rec.value	Max.error
S T R E A M S [T/H]				
STEAM-IN	MN	440.000	440.000	17.600
STEAM-OUT	NO	400.000	440.000	17.600
P R E S S U R E S [MPAG]				
IN	MN	2.650	2.650	0.050
OUT	MN	2.350	2.350	0.050
W E T N E S S E S [%]				
STEAM-IN	F	0.250	0.250	
STEAM-OUT	NO	0.250	0.187	0.016

One can see that due to throttling, the steam wetness has been lowered.

We have done still one computation with higher pressures of the steam; the values of other variables have remained the same.

```

P R E S S U R E S [MPAG]
IN      M      4.6500  5.000E-2
OUT     M      4.3500  5.000E-2
  
```

The resulting wetnesses now are

```

W E T N E S S E S [%]
STEAM-IN  F      0.250  0.250
STEAM-OUT NO      0.250  0.373  0.029
  
```

In this case, the wetness of the outlet steam has increased from the value 0.187 % obtained in the preceding example, to the value 0.373 %. The, as it seems to be contradiction of the two results follows from the saturated steam enthalpy dependency on pressure. While in the first case enthalpy increased with pressure, with higher pressures in the second example we have already got into the region where with increasing pressure, saturated steam enthalpy decreases (see the $p-i$

diagram for water vapor). The dividing point is ca. 235 deg C, resp. 3.06 MPa. In the neighborhood of this point (of the respective values on the saturation line), the temperature resp. pressure dependency of the saturated steam enthalpy is very flat and passes through the maximum. This fact is quite relevant for the precision of balancing calculations and will be documented with more details in one of the case studies.

5.8 Wetness separator

The separator of wetness (small water droplets) serves for separating the liquid phase from the wet steam. Since the steam wetness is difficult to measure, one usually starts from the assumptions of given (fixed) inlet wetness and separation efficiency expressed e.g. by the steam wetness at the outlet from the separator. The model generates 2 equations. Input variables can then be for example inlet steam flowrate and steam wetnesses at the inlet and outlet. It is then possible to compute for example the flowrate of the liquid phase separated and that of steam at the outlet from the separator. In the following example, we'll assume that the whole separator works under one pressure.

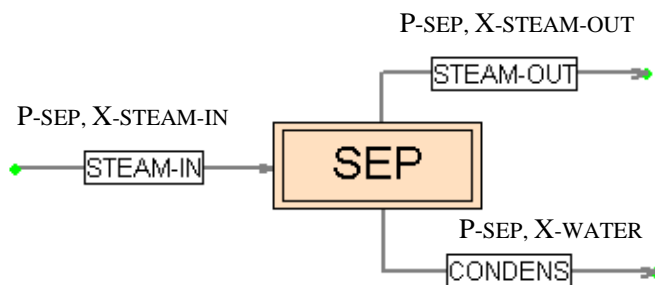


Fig. 5.8-1: Balance scheme (demo Example E-8)

INPUT DATA

NAME	TYPE	VALUE	MAX. ERROR
M A T E R I A L S T R E A M S [T/H]			
CONDENS	N	10.0000	
STEAM-IN	M	64.2000	4.0000 %
STEAM-OUT	N	40.0000	
P R E S S U R E S [KPAG]			
SEP	M	0.6600	5.000E-3
W E T N E S S E S [%]			
STEAM-IN	F	6.2000	
STEAM-OUT	F	0.3000	
water	F	100.0000	

Node: SEP

ID	Description
SEP	separator

Above sea level: Node pres.:

Sort of calculations:

Balancing

Hydraulic node

Heat node

Reaction node

Non-energetic streams incident with node				
Stream	Function	Temperature	Pressure	Wetness
CONDENS	H2O(P,X)		SEP	water
STEAM-IN	H2O(P,X)		SEP	STEAM-I
STEAM-OUT	H2O(P,X)		SEP	STEAM-C

Reactions in node

Reaction:

Fig. 5.8-2: Panel of the model parameters definition

RESULTS

GLOBAL DATA

Degree of redundancy	0
Number of equations	2
Qmin	0.000E+00
Qcrit	0.000E+00

VARIABLES

Name	Type	Inp.values	Rec.values	Max.error
S T R E A M S [T/H]				
CONDENS	NO	10.000	3.799	0.152
STEAM-IN	MN	64.200	64.200	2.568
STEAM-OUT	NO	40.000	60.401	2.416
P R E S S U R E S [KPAG]				
SEP	MN	0.660	0.660	5.00E-3
W E T N E S S E S [%]				
STEAM-IN	F	6.200	6.200	
STEAM-OUT	F	0.300	0.300	
water	F	100.000	100.000	

5.9 Input / output of energy (pump)

One frequently meets with the need to supply / withdraw energy into / from the system. An example can be e.g. withdrawing heat into the environment or work performed on the shaft of a turbine. Such kind of problems will be illustrated by the energy balance of a pump.

The pump serves for enhancing the pressure of the fluid stream. The necessary energy is usually supplied by an electrical motor. The whole input power is transferred to the pump with certain efficiency (e.g. 95 %). The remaining part of electric energy is transformed to heat withdrawn into the environment by the cooling of the motor. The part that enters the pump energy balance is only energy transferred by the shaft connecting the motor and the pump.

Mechanic energy is transformed in the pump proper to enthalpy thus pressure and heat energies (heating of the fluid). The ratio of the two energies depends on the construction of the pump, its mechanical state and the working régime.

We'll further describe the energy balance of the pump proper. The input mechanic energy (power) is measured by the electric energy input multiplied by its efficiency.

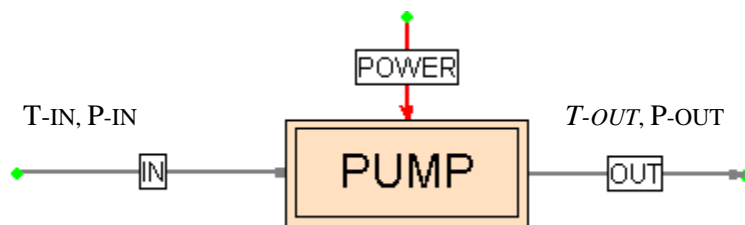


Fig. 5.9-1: Balance scheme (demo Example E-9)

INPUT DATA

NAME	TYPE	VALUE	MAX.ERROR
M A T E R I A L S T R E A M S [T/H]			
IN	M	836.0000	3.0000 %
OUT	N	800.0000	
F L O W S O F E N E R G Y [MWH/H]			
POWER	M	0.5500	10.0000 %
T E M P E R A T U R E S [C]			
IN	M	38.8000	1.0000
OUT	N	38.0000	
P R E S S U R E S [MPAG]			
IN	M	1.0800	1.000E-2
OUT	M	1.9800	1.000E-2

Node: PUMP

ID	Description
PUMP	the pump

Above sea level: Node pres.:

Sort of calculations:

Balancing

Hydraulic node

Heat node

Reaction node

Non-energetic streams incident with node

Stream	Function	Temperature	Pressure	Wetness
IN	H2O(L,T,F)	IN	IN	
OUT	H2O(L,T,F)	OUT	OUT	

Reactions in node: Reaction

Fig. 5.9-2: Panel of the model parameters definition

RESULTS

GLOBAL DATA

Degree of redundancy	0
Number of equations	2
Qmin	0.000E+00
Qcrit	0.000E+00

VARIABLES

Name	Type	Inp.value	Rec.value	Max.error
STREAMS [T/H]				
IN	MN	836.000	836.000	25.080
OUT	NO	800.000	836.000	25.080
FLOWS OF ENERGY [MWH/H]				
POWER	MN	0.550	0.550	0.055
TEMPERATURES [C]				
IN	MN	38.800	38.800	1.000
OUT	NO	38.000	39.176	1.002
PRESSURES [MPAG]				
IN	MN	1.080	1.080	0.010
OUT	MN	1.980	1.980	0.010

This simple example has shown a simple interconnection of the energy system with the 'ambient world'. At the mechanic energy, one dealt only with a rough estimate (the 10 % uncertainty can even be too optimistic). On the other hand we have obtained the temperature increase after the pump by less than 0.4 deg C, which is certainly inside the limits of uncertainty at temperature measurement. Fortunately, the heat equivalent of mechanic work is small.

5.10 Turbine segment

For the needs of balancing, the turbine segment (TS) is defined as part of the turbine, endowed with just one side steam withdrawal (bleeding). The thus defined TS can comprise one or several circulating wheels, which is however irrelevant for the balancing. The main goal of the TS balance is to find the work (power) exerted on the turbine shaft, which is otherwise a practically unmeasurable variable.

Let us further suppose that we know the pressure and temperature or (in the case of wet steam) wetness of steam at the inlet to some TS. In the TS, steam exerts work and gets on into the following segment (or out of the turbine), while some part of it is withdrawn as a side stream (extracted). For the outlet streams, we also suppose the knowledge of pressure and temperature (or again wetness for wet steam). If some of these variables were not known, observability problems could arise with the unmeasured variables. Because the turbine is part of a larger system, the observability can be improved by the integration of the TS model into the whole model.

In the following example, we put together the models of TS and feed water preheating. The steam extracted from the turbine condenses in the exchanger and preheats thereby the feed water. From this preheating, the balance enables us to compute the unmeasured amount of extracted steam.

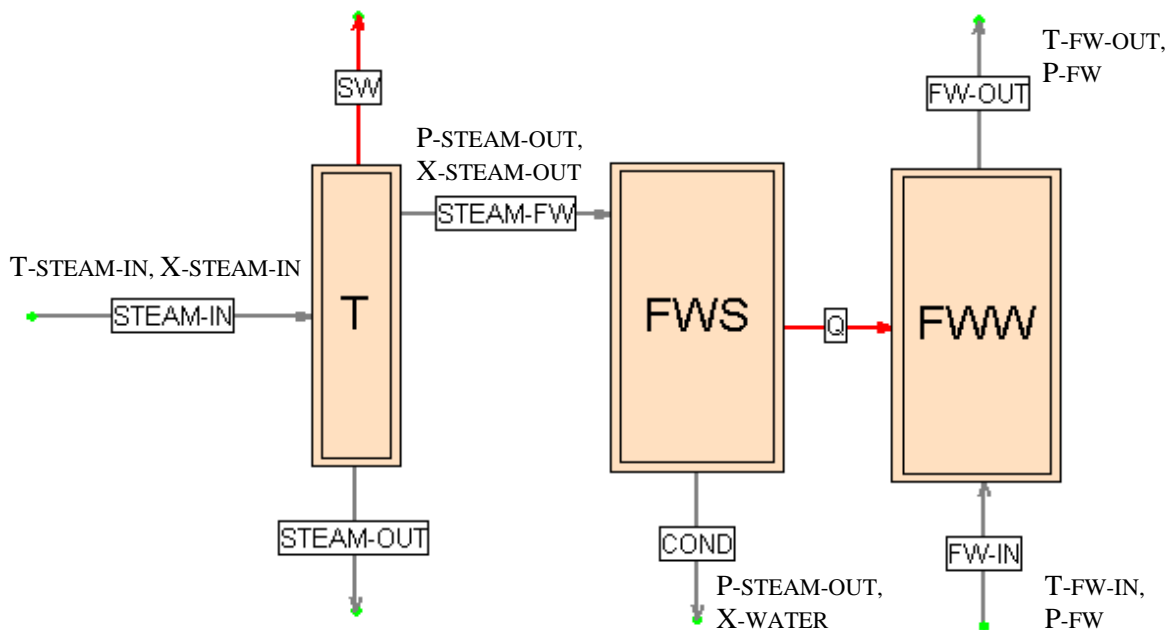


Fig 5.10-1: Balance scheme (demo Example E-10)

The following symbols are to be interpreted:

- T turbine segment
- FWS feed water preheating – steam side of exchanger
- FWW feed water preheating – water side of exchanger
- STEAM-IN saturated steam going into TS, defined by temperature ($X = 0$).

STEAM-OUT steam leaving TS and going into further TS, with wetness 3.6 %
 STEAM-FW extracted steam heating the feed water heater. Its state is the same as
 in stream STEAM-OUT
 SW work (power) on the turbine shaft (Shaft Work)
 COND condensate from STEAM-FW
 Q heat transferred in feed water heater
 FW-IN inlet feed water
 FW-OUT outlet feed water

INPUT DATA

NAME	TYPE	VALUE	MAX.ERROR
M A T E R I A L S T R E A M S [T/H]			
COND	N	100.0000	
FW-IN	M	1350.0000	2.0000 %
FW-OUT	N	1300.0000	
STEAM-FW	N	100.0000	
STEAM-IN	M	1320.0000	3.0000 %
STEAM-OUT	N	1250.0000	
F L O W S O F E N E R G Y [MWH/H]			
Q	N	50.0000	
SW	N	30.0000	
T E M P E R A T U R E S [C]			
FW-IN	M	191.0000	1.0000
FW-OUT	M	223.0000	1.0000
STEAM-IN	M	256.0000	1.0000
P R E S S U R E S [MPAG]			
FW	M	6.6000	5.000E-2
STEAM-OUT	M	2.8400	2.000E-2
W E T N E S S E S [%]			
STEAM-IN	F	0.000E+0	
STEAM-OUT	F	3.6000	
water	F	100.0000	

PANELS OF MODEL PARAMETERS DEFINITION

Node: T

ID: T Description: turbine

Above sea level: Node pres.:

Sort of calculations:

- Balancing
- Hydraulic node
- Heat node
- Reaction node

Non-energetic streams incident with node

Stream	Function	Temperature	Pressure	Wetness
STEAM-FW	H2O(P,X)		STEAM-OUT	STEAM-OUT
STEAM-IN	H2O(T,X)	STEAM-IN		STEAM-IN
STEAM-OUT	H2O(P,X)		STEAM-OUT	STEAM-OUT

Reactions in node: Reaction

Fig. 5.10-2: Turbine segment

Node: FWS

ID: FWS Description: feed water - steam side

Above sea level: Node pres.:

Sort of calculations:

- Balancing
- Hydraulic node
- Heat node
- Reaction node

Non-energetic streams incident with node

Stream	Function	Temperature	Pressure	Wetness
COND	H2O(P,X)		STEAM-OUT	water
STEAM-FW	H2O(P,X)		STEAM-OUT	STEAM-OUT

Reactions in node: Reaction

Fig. 5.10-3: Preheater – steam side

Node: FWW

ID: FWW Description: feed water - water side

Above sea level: Node pres.:

Sort of calculations:

- Balancing
- Hydraulic node
- Heat node
- Reaction node

Non-energetic streams incident with node

Stream	Function	Temperature	Pressure	Wetness
FW-IN	H2O(T,P)	FW-IN	FW	
FW-OUT	H2O(T,P)	FW-OUT	FW	

Reactions in node: Reaction

Fig. 5.10-4: Preheater – water side

RESULTS

GLOBAL DATA

Degree of redundancy	0
Number of equations	6
Qmin	0.000E+00
Qcrit	0.000E+00

V A R I A B L E S

Name	Type	Inp.value	Rec.value	Max.error
S T R E A M S [T/H]				
COND	NO	100.000	112.127	5.448
FW-IN	MN	1350.000	1350.000	27.000
FW-OUT	NO	1300.000	1350.000	27.000
STEAM-FW	NO	100.000	112.127	5.448
STEAM-IN	MN	1320.000	1320.000	39.600
STEAM-OUT	NO	1250.000	1207.873	39.973
F L O W S O F E N E R G Y [MWH/H]				
Q	NO	50.000	54.046	2.625
SW	NO	30.000	22.097	0.685
T E M P E R A T U R E S [C]				
FW-IN	MN	191.000	191.000	1.000
FW-OUT	MN	223.000	223.000	1.000
STEAM-IN	MN	256.000	256.000	1.000
P R E S S U R E S [MPAG]				
FW	MN	6.600	6.600	0.050
STEAM-OUT	MN	2.840	2.840	0.020
W E T N E S S E S [%]				
STEAM-IN	F	0.00E+0	0.00E+0	
STEAM-OUT	F	3.600	3.600	
water	F	100.000	100.000	

5.10 Phase equilibrium water – water vapor

Data reconciliation with the aid of phase equilibrium will be illustrated by the model of a steam generator (SG), which has been already given with details in Section 5.4. The model created in the graphical editor remains unchanged and the reader is referred to the section mentioned above. Further described will be the enlarging of the model by the phase equilibrium equation.

Let us now suppose that besides the temperature, also pressure has been measured in the SG. The relation between temperature and pressure is not given in the graphical editor, but in the editor of user defined equations. There are now two possibilities – either express temperature as function of pressure, or pressure as function of temperature. The result should, however, be practically independent of this choice.

In the original model, altogether 4 balance equations are generated, the phase equilibrium assumption generates the fifth. There are two unmeasured variables in the problem, so three degrees of redundancy are at hand for data reconciliation and validation.

The screenshot displays the 'Editor of user defined equations' interface. It is divided into several functional areas:

- Standard functions:** A grid of buttons for mathematical operations including `exp`, `10^`, `sqr`, `abs`, `ln`, `log`, `^2`, and a numeric keypad (0-9, ., ^, +, -, *, /, (,)).
- Independent variables:** A grid of buttons for variables such as `Mass flow`, `Pressure`, `Concentration`, `Steam wetness`, `Heat flow`, `Pipe roughness`, `Temperature`, and `Auxiliaries`.
- Calculated variables:** A grid of buttons for variables like `Heat function`, `Sat.steam-temp.`, `Molar mass`, `Sat.steam-pres.`, `Reaction extent`, `Water-dens.`, `Material d.viscosity`, and `Material density`.
- User equation:** A large text input field for defining the equation.
- Strap table:** A button for accessing a table of data.
- Equations List:** A table with columns for 'Equation' and 'Description'. It contains one entry: 'EQUIL' with the description 'equilibrium T-P in the SG'. To the right of this table are 'Checking' icons (checkmark, error, warning) and a 'Use in model?' checkbox which is checked.
- Equation Editor:** A section for editing the selected equation. It shows the 'Equation proper' as `[ST<PSG>]-[T<TSG>]` and the 'Resulting value' as `= 0`.

Fig. 5.11-1: Editor of user defined equations (demo Example E-11)

Equation EQUIL represents here the relation between measured temperature T and equilibrium temperature T^* , which is function of pressure.

$$T^* = T(P) \tag{5.11-1}$$

In the editor, the equation is of the form

$$[ST<PSG>-[T<TSG>] \quad , \quad (5.11-2)$$

which is Eq. (5.11-1) rewritten with zero right-hand side. Function ST (Saturated Temperature) called by button „Saturated steam – temp.“ has the argument of measured pressure PSG. The second term in the equation is measured temperature TSG.

INPUT DATA

In addition to the flowrates of streams, we here have 4 temperatures (those of hot water HWIN and HWOUT, the temperature in steam generator TSG valid for outlet steam and blowdown, and temperature of feed water FW). Further, we here have three pressures (pressure in steam generator PSG, for feed water FW and for hot water HW). New are two wetnesses for liquid water (WATER) and steam (STEAM).

NAME	TYPE	VALUE	MAX.ERROR
M A T E R I A L S T R E A M S [KG/S]			
BLOWDOWN	M	6.1200	5.0000 %
FW	M	444.5000	2.0000 %
HWIN	M	5650.0000	5.0000 %
HWOUT	N	5000.0000	
STEAM	M	445.0000	2.0000 %
F L O W S O F E N E R G Y [KJ/S]			
QSG	N	800000.0000	
T E M P E R A T U R E S [C]			
HWIN	M	295.2000	1.0000
HWOUT	M	265.8000	1.0000
SG	M	257.6000	1.0000
FW	M	221.6000	1.0000
P R E S S U R E S [KPA]			
FW	M	10.0000	0.5000 %
HW	M	10.0000	0.5000 %
PSG	M	4.5400	1.0000 %
W E T N E S S E S [%]			
STEAM	F	0.2500	
WATER	F	100.0000	

PAN ELS OF MODEL PARAMETERS DEFINITION

Both panels are the same as in Section 5.4.

RESULTS

G L O B A L D A T A	
Degree of redundancy	3
Number of equations	5
Number of user defined equations	1
Qmin	4.409E+00
Qcrit	7.842E+00

V A R I A B L E S

Name	Type	Inp.values	Rec.values	Max.error
S T R E A M S [KG/S]				
BLOWDOWN	MC	6.120	6.115	0.306
FW	MC	444.500	448.864	6.172
HWIN	MC	5650.000	5471.657	200.269
HWOUT	NO	5000.000	5471.657	200.269
STEAM	MC	445.000	442.749	6.172
F L O W S O F E N E R G Y [KJ/S]				
QSG	NO	800000.000	815954.273	11528.935
T E M P E R A T U R E S [C]				
FW	MC	221.600	221.570	0.999
HWIN	MC	295.200	294.744	0.863
HWOUT	MC	265.800	266.162	0.890
TSG	MC	257.600	257.876	0.522
P R E S S U R E S [MPA]				
FW	MC	10.000	10.000	0.050
HW	MC	10.000	10.000	0.050
PSG	MC	4.540	4.532	0.039
W E T N E S S E S [%]				
STEAM	F	0.250	0.250	
WATER	F	100.000	100.000	

Note: If we created a user defined equation for phase equilibrium and temperature or pressure were unmeasured, this would only serve for computing the unmeasured variable without changing the degree of redundancy. In this case, RECON would only serve as a calculator for equilibrium temperature or pressure. Let us note in addition that in the panel „Node balance“, the unmeasured temperatures resp. pressures under phase equilibrium conditions are available for the user automatically, even without defining any user defined equation. ♦

5.11 User defined variables and equations

This example was taken over from a more extensive problem. It will thus not be completely solved here. Our aim is only to show how the user defined variables and equations are integrated in the whole model.

We have a system with six steam generators, whose thermal powers (flows of energy) have identifiers QSG1, QSG2 to QSG6 [MW]. For the sum of the powers, we have introduced the auxiliary variable QNR [MW] defined by the equation

$$QNR = QSG1 + QSG2 + QSG3 + QSG4 + QSG5 + QSG6 \quad (5.12-1)$$

One must further write the user defined equation. It is important to know that if the equation involves standard variables (flowrates, pressures and the like), the program enters them in units of the SI system.

For better understanding, notice that although the program works, from the viewpoint of the user, with the system of units he has chosen, for the computation proper all values are transformed into the system SI. Only after finishing all calculations, the data are presented to the user in his system of units again. This holds, however, not for the auxiliary variables. It is the user's task to care for the physical dimensions of the latter. ♦

If we thus want, in our case, to have the variable QNR in MW, we must take account of it when writing the user defined equation. The equation then must be of the following form [recall that the equations in the RECON program are written with zero r.h. side – Eq. (5.12-2) is rewritten Eq. (5.12-1)]

$$QNR * 1000000 - QPG1 - QPG2 - QPG3 - QPG4 - QPG5 - QPG6 = 0 \quad (5.12-2)$$

The constant 1000000 represents here transformation of variable QNR (which is in MW) into the expression in system SI with corresponding unit W.

The realization in RECON program is shown in the following figure

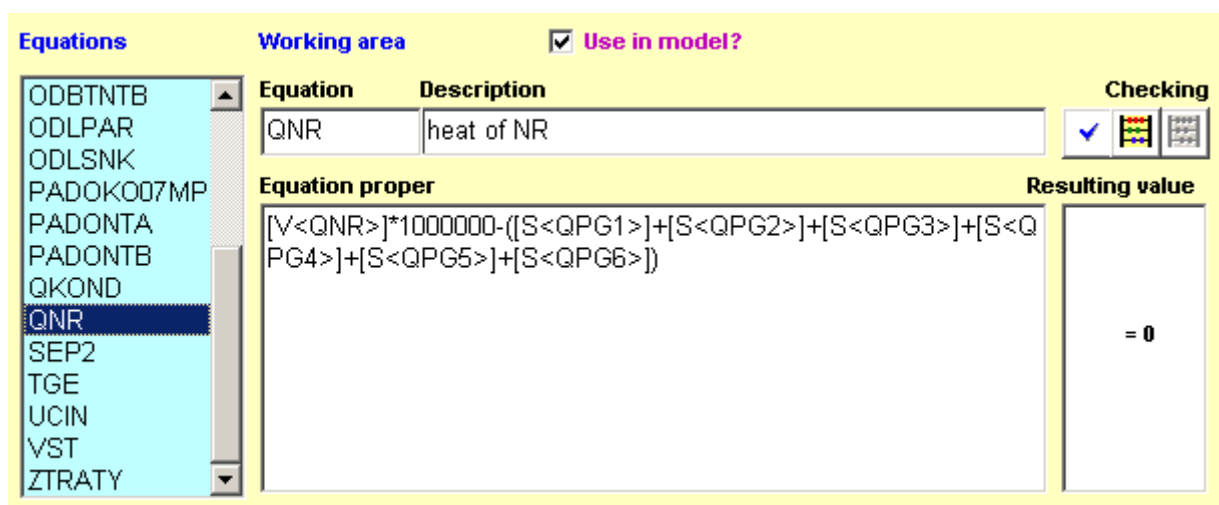


Fig. 5.12-1: Panel for definition of user defined equations

Recall here that the syntax of variables entry is

[kind of the variable<identifier of the variable>].

For example [V<QNR>] means that variable QNR is auxiliary (Various). In contrast to it [S<QPG1>] - variable QPG1 – is the stream flowrate (Stream). Naturally, the user need not worry with this syntax, it is created by the program itself.

5.12 Splitter of streams

Let us finally mention one very frequent thermal unit operation, for which one almost never sets up any energy balance in practice. By *Splitter of streams* is meant dividing one stream into two or more further streams, while the thermodynamic state of all the streams is identical.

Let us further consider a simple splitter with two outlet streams.

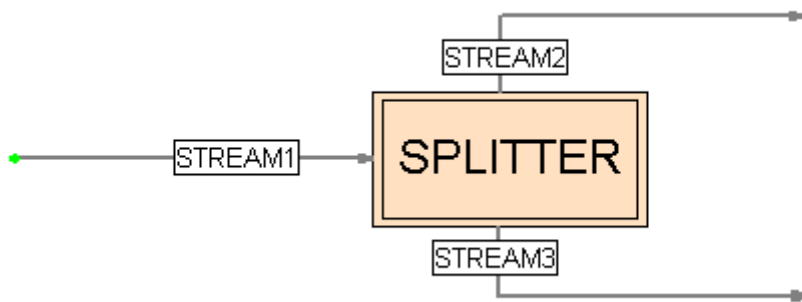


Fig. 5.13-1: Splitter of stream

For this node, we can write the mass balance equation

$$m_1 = m_2 + m_3 \quad (5.13-1)$$

where the m_i are flowrates,
further the energy balance

$$m_1 h_1 = m_2 h_2 + m_3 h_3 \quad (5.13-2)$$

where the h_i are specific enthalpies of the streams.

However, as the specific enthalpies of all streams are equal according to the definition of a splitter (the same thermodynamic state), one can divide Eq. (5.13-2) by the common enthalpy value and we have Eq. (5.13-1). The energy balance thus brings no new information, being linearly dependent on the former. In practice, one thus creates just one, viz. the mass balance equation around any splitter. If setting up also the energy balance, RECON will discard it as dependent.

Note: One can imagine the situation where one measures simultaneously the inlet and outlet state values around a splitter. In this case, it could make sense to set up the energy balance as well. However this variant is not met with in practice. ♦

Case study 1: Identification of steam generator thermal power

In the following case study, we'll describe in more detail the complex processing of data measured on one steam generator of a nuclear power station. Besides the balance proper and data reconciliation, we'll also give further information that can be deduced from the measured data.

Let us consider the steam generator (SG) in the following figure.

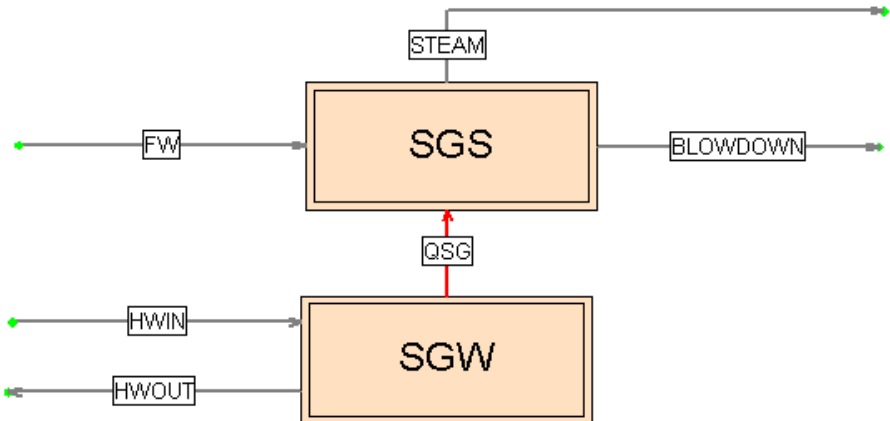


Fig. CS1-1: Steam generator

Hot water (HW) circulates between the nuclear reactor and tube space denoted as SGW. Steam (containing 0.25 % wetness) is generated in the shell space SGS. From the shell space of SG, blowdown is continuously withdrawn in addition. The heat stream QSG represents the heat flow (thermal power) in SG. The following table gives measured values and their uncertainties.

Table CS1-1: Measured variables

Variable	Stream	Meas.unit	Value	Max.error (uncertainty)
Flow	FW	kg/s	444.5	2%
Flow	STEAM	kg/s	445.0	2%
Flow	BLOWDOWN	kg/s	6.12	5%
Flow	HWIN	kg/s	5650	5%
Temperature	FW	deg C	221.6	1
Temperature	STEAM	deg C	257.6	1
Temperature	BLOWDOWN	deg C	257.6	1
Temperature	HWIN	deg C	295.2	1
Temperature	HWOUT	deg C	265.8	1
Pressure	SGS	kPa	4500	0.5%
Pressure	FW	kPa	10000	0.5%

This basic variant of SG was already solved in the frame of the models for individual operations, where the reader can find the results of reconciliation (see Section 5.4). We'll give below further results and analyses.

Propagation of errors in data processing

Information in this respect is provided by the vector of shares introduced in Section 3.9 *Propagation of errors at data processing and....* . Let us recall that this vector contains percentual shares of individual measured variables on the dispersion of the result. The program RECON offers the vector of shares in menu *calculations – Propagation of errors*.

Let us further concentrate on the thermal power of SG – variable QSG.

MESSAGE ON ERROR PROPAGATION

Heat flow QSG

THE VARIANCE OF GIVEN VARIABLE IS CAUSED MAINLY BY:

Type	Variable	Share
MF	FW	47 %
MF	HWIN	2 %
MF	STEAM	47 %
T	FW	3 %
Total		98 %

One can see that the dominant effect on the thermal power precision is due to the flowrates, and in particular those of feed water and steam constituting 94 % from the dispersion of the result. If we wanted to make the result still more precise, this would make sense just with these two variables. The others are of rather negligible impact.

Let us further put the question, why further variables make themselves less valid in the vector of shares. For instance at the blowdown measurement, this measurement itself is very precise. While in the feed water measurement, absolute uncertainty is 8.9 kg/s (2 % from 444.5 kg/s), it is only 0.3 kg/s for the blowdown. At the hot water flowrate measurement, the reason is more complicated and will be explained in the next section of this case study.

Somewhat surprising is the small importance of measured temperatures. Why for example, in the list of relevant variables doesn't occur the temperature in SG, which determines the steam enthalpy, thus the main carrier of energy? One of the reasons is certainly the fact that the assumed uncertainty 1 deg C is relatively small and possible impact on the energy balance is not great in these limits. More essential is however the fact that in the temperature domain typical of SG, the temperature dependency of saturated steam enthalpy is quite flat. For example at 257 deg C, it is ca. 0.4 kJ/(kg deg C), which is only 0.02 % of the evaporation heat at this temperature. For instance an error in steam temperature 10 deg C (hence ten times the assumed uncertainty) results only in a several tenths per cent error in the stream enthalpy, thus substantially smaller than the flowrate measurement error. Somewhat different is the situation at feed water. The specific heat at water pressure 10 MPa and 220 deg C is ca. 4.49 kJ/(kg deg C), which is ca. 10 times more than it was at the steam. Therefore, the feed water temperature occurs in the vector of shares (though as an item of smaller importance).

For the same reason, in the vector of shares absent is the feed or hot water pressure. The pressure dependency of liquid water enthalpy is even less pronounced than in the preceding case of saturated steam, so that even large errors at the water pressure measurement do not cause large errors in the balance. On the other side however, the chances for the detection of these errors are also bad.. Based on the balance models, the detection of these errors is hardly possible (see *threshold value* for the gross errors detection, already described earlier).

Strategy of measurement

One of the important measurement results is heat flow QSG, which plays the main role in the nuclear reactor thermal power identification. One speaks then of a *key variable* of the whole measurement. Usually, there are several different ways for its identification, based on the choice of measured variables and the measured values processing – it is so-called *strategy of measurement* and measured data processing.

Let us further review several variants, in order to show the importance of the strategy for measured data analysis. We here make use of the RECON program, with the aid of which we can give simple form to the complete solution described in the previous section.

The individual variants of heat flow identification are following.

1. From the mass and heat balances of hot water (balance around node SGW): One deals with direct calculation without data reconciliation on the nuclear reactor side.
2. From the mass and heat balances for the steam part of SG: One deals with direct calculation without data reconciliation on the steam generator proper side. Steam flowrate is considered unmeasured and it is calculated from the feed water and blowdown flowrates.
3. From reconciled mass and heat balances for the steam part of SG: Hot water balance is not taken into account.
4. From the reconciled balance of the whole system (model applied in previous section).
5. Strategy No. 4 is made still more perfect on applying new pressure measurement in (the steam part of) SG, and this pressure is reconciled with the temperature in SG according to the phase equilibrium condition.

The results are in the following table.

Tab. CS1-2: Identification of flowrate QSG in different ways

Strategy No	QSG [MW]	Tolerance [MW]	Tolerance [%]	Degree of redundancy
1	866.9	60.4	7.0	0
2	807.9	16.5	2.0	0
3	813.9	11.7	1.4	1
4	816.0	11.5	1.4	2
5	816.0	11.5	1.4	3

The results are in good agreement with simple rules familiar to those which deal systematically with process measurement.

- The choice of the whole measurement strategy is of fundamental importance. Even if strategy No.1 looks very good as for the simplicity of the balance calculation, the result is very bad. The hot water balance suffers from the fact that it is based on the evaluation of temperature difference, which is a difference of two large numbers. In addition, there is a relatively great uncertainty in the hot water flowrate measurement.
- Considerably better is the result of strategy No.2 based on the SG steam side balance. The balance works with a relatively precise knowledge of the feed water flowrate. The measurement of temperatures has only marginal importance for setting up the heat balance, as we have shown above. This strategy based on simple mass balance is often used in practice.
- Strategy No.3 is supported by the reconciliation of mass balance around the SG steam side. Further, the result uncertainty is substantially reduced. In addition, we here have the effect of data validation consisting in the possibility of gross errors detection.
- Strategy No. 4 does not bring any relevant diminishing of the result uncertainty. Two models (No.1 and 3) have been integrated and the degree of redundancy increased by 1. The model according to strategy No.1 brings itself, from the standpoint of thermal power identification, substantially less than model No.3. We here have an asset in enhancing the precision of flowrate measurement on the hot water circuit. More detailed analysis shows that the hot water flowrate uncertainty is lowered by ca. 30 % due to the reconciliation.
- Strategy No. 5 further increases the degree of redundancy, however without any sensible effect on the thermal power uncertainty. One deals with reconciliation and temperature precision enhancement, and as shown above, the temperature of steam in SG is of minor importance for setting-up the energy balance. Still, the chance for the validation of temperature and pressure data in SG is then generally better.

While at strategy No.4, the adjustability of steam temperature in SG was lower than 0.001, after including the phase equilibrium model by strategy No.5 the adjustability has increased to 0.478 (the uncertainty in temperature has decreased almost to one half after the reconciliation). In an analogous way, also the conditions for the detection of a gross error in this temperature have been improved. In the case of strategy No.4, there was practically no chance to detect a gross error in this temperature, even if the error were of the order of magnitude several tens deg C. In contrast to it, at strategy No.5 the threshold value for gross error detection equals 2.3 deg C. This kind of information could be useful in situations where the temperature monitoring is important, for example for the sake of process safety.

Case study 2: Identification of nuclear reactor thermal power

Let us consider the scheme given in Fig. CS2-1; it represents a system of steam generators withdrawing heat from a nuclear reactor. We have here 3 measured streams of condensate (COND1-3), supplied to condensate collector CONDHEAD. From here, the condensate is pumped via two measured streams (FWA and FWB) into feed collector FWHEAD. It is then distributed into 4 steam generators SG1-4. We here further have 4 measured streams of blowdown (PURGE1-4). In each SG, measured stream of steam STEAM1-4 is generated. Steam is led into steam collector STHEAD, from where it goes by measured stream STEAMSUM to the turbine.

Temperatures are measured for all streams of feed water and steam. The blowdown temperatures are assumed to be the same as the steam temperatures in the respective SG. The condensate temperatures are not measured. For this reason, no energy balance around node CONDHEAD is created in the model.

The subsystem of hot water from the nuclear reactor is not included. The heat supply to individual SG is modeled by four heat flows QSG1-4 that come from node NR representing the nuclear reactor. Stream QNR then represents the whole thermal power of the reactor. QNR is the key variable to be identified.

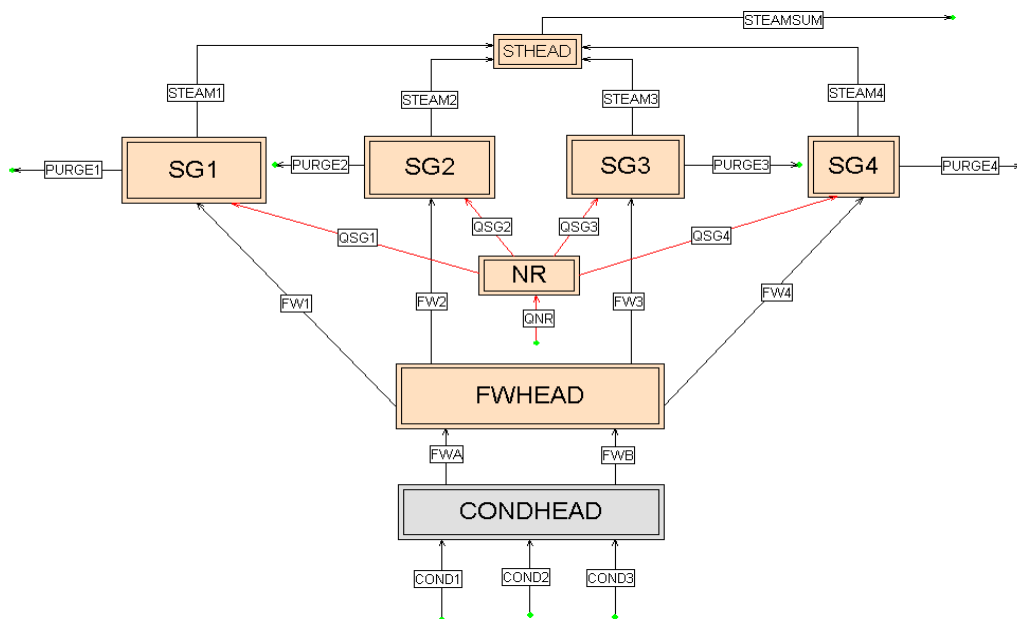


Fig. CS2-1: System of four steam generators

Flowrates and temperatures are measured with the following uncertainties:

Table CS2-1: Measurement tolerances

Type	Stream	Tolerance
Temperature	All	1 °C
Flow	STEAM	4 %
Flow	PURGE	5 %
Flow	Remaining streams	2%

The steam is wet (0.25% liquid phase). The pressure of condensate and feed water is assumed fixed (8 MPag).

For the majority of nodes, the model creates 2 equations – mass and energy balances. The following nodes make an exception:

- CONDHEAD – here, inlet temperatures are not available and only mass balance equation is created
- NR – mass flowrates are absent and only energy balance is created.

Altogether, one creates 14 equations. There are 5 unmeasured variables (flowrates) in the problem. The degree of redundancy is $14 - 5 = 9$.

Main results

The main results read

GLOBAL DATA

Degree of redundancy	9
Number of unmeasured variables	5
Number of observable variables	5
Number of equations	14
Qmin	7.426E+00
Qcrit	1.689E+01

QUANTITIES

Name	Type	Inp.value	Rec.value	Max.error
S T R E A M S [KG/S]				
COND1	MC	780.300	780.076	11.995
COND2	F	0.00E+0	0.00E+0	
COND3	MC	763.350	763.135	11.908
FW1	MC	383.900	384.774	6.430
FW2	MC	387.900	388.206	6.478
FW3	MC	389.500	386.816	6.462
FW4	MC	387.400	383.415	6.409
FWA	MC	773.800	775.213	11.950
FWB	MC	766.700	767.997	11.913
PURGE1	MC	3.520	3.519	0.176
PURGE2	MC	3.890	3.890	0.194
PURGE3	MC	2.810	2.811	0.140
PURGE4	MC	4.120	4.123	0.206
STEAM1	MC	385.100	381.255	6.432
STEAM2	MC	385.800	384.316	6.480
STEAM3	MC	374.400	384.005	6.462
STEAM4	MC	365.400	379.292	6.411
STEAMSUM	MC	1532.200	1528.868	9.747
F L O W S O F E N E R G Y [MJ/S]				
QNR	NO	10000.000	2820.435	18.195
QSG1	NO	2500.000	703.493	11.976
QSG2	NO	2500.000	707.889	12.047
QSG3	NO	2500.000	709.531	12.051
QSG4	NO	2500.000	699.523	11.931
T E M P E R A T U R E S [C]				
FWA	MC	220.500	220.302	0.815
FWB	MC	222.000	221.804	0.818
FWSG1	MC	220.800	220.898	0.958
FWSG2	MC	221.600	221.699	0.957
FWSG3	MC	220.400	220.500	0.957
FWSG4	MC	221.000	221.099	0.958
SG1	MC	259.200	259.185	0.973
SG2	MC	258.600	258.585	0.975
SG3	MC	257.000	256.987	0.978
SG4	MC	259.800	259.785	0.972
steamsum	MC	258.600	258.658	0.447
P R E S S U R E S [MPAG]				
NVPG	F	8.000	8.000	
W E T N E S S E S [%]				
PGsteam	F	0.250	0.250	

The sum of squares of adjustments $Q_{min} = 7.426$, critical value of chi-square distribution with 9 degrees of freedom at the significance level 0.05 [$\chi^2_{0.95}(9)$] = 16.91. Since

$$7.426 < 16.91$$

no gross error presence has been detected.

We have found the value of the reactor thermal power $QNR = 2820.4$ MW with tolerance 18.2 MW (which represents 0.65 % from the computed value).

It can look strange that the result precision is very good. The uncertainty value for the thermal power 0.65 % is substantially smaller than for any of the measured streams. Moreover, also errors at setting-up the heat balance must play a role.

There are several reasons. First, it is the improvement due to reconciliation of the steam generators mass balance (see results of the previous case study). Further, there is considerable redundancy in the measurement of condensate and feed water flowrates (altogether at three levels). The last reason is the fact that the whole thermal power is given as sum of 4 single thermal powers of individual SG, which has generally a positive impact on improving the relative precision of the result.

Further information

By its extent, this case study already approaches real problems met with in practice. It will thus be useful to give further interesting results that can be considered typical of this kind of problems.

For the sake of brevity, let us give the results in abridged form. We'll use the fact that our scheme is symmetric around the vertical axis. The values of parallel variables (e.g. parallel stream flowrates) are nearly equal and the same holds for their further properties). So not single results, but only those for the groups of analogous variables will be given. The results for adjustabilities, threshold values and parametric sensibilities are given in the following table.

Table CS2-2: Further results of data reconciliation

Type	Stream	Adjustability	Threshold value TV	Parametric sensitivity of QNR
Flow	COND	0.23	55.4	0.367
Flow	FW1-4	0.17	31.7	0.729
Flow	FWA,B	0.23	55.3	0.368
Flow	PURGE	0.00	34.6	-1.29
Flow	STEAM	0.57	38.0	0.192
Flow	STEAMSUM	0.68	73.4	0.186
Temperature	FWA,B	0.18	3.9	-0.176
Temperature	FW1-4	0.04	7.9	-0.176
Temperature	STEAM	0.03	10.2	0.014
Temperature	STEAMSUM	0.55	2.5	0.014

Let us further discuss the individual results.

Adjustability gives the decrease of result uncertainty due to the reconciliation. For example at stream COND, this decrease (precision enhancement) is by 23 %. The average value of adjustability is 0.27. This roughly corresponds to the experience from practice, where one gives the result precision enhancement some 30 % on the average. However, one can see here even substantially better adjustability values around 60 %, and also practically nonadjustable variables. The latter are those which are measured with high absolute precision (with respect to the other variables); this is the case of the blowdowns (see also the discussion in the previous case study). Further case is represented by temperatures that are also of relatively high precision and in addition are, as steam temperatures, of minor importance in the heat balance (also already discussed in the previous case study).

Threshold value gives minimum value of gross error that will be detected with probability 90 %. Thus for example the value $TV = 55.4$ for stream COND means that the gross error must be at least 55.4 t/h so as to be detected with probability 90 % (for information, the flowrates of this stream are ca. 770 t/h, so the threshold value represents some 7 % of the nominal stream value).

It follows from the theory that the threshold value is closely connected with the adjustability of the variable. The smaller the adjustability, the higher is the threshold value (hence also the chance for gross error detection is smaller). Thus for example for almost nonadjustable blowdowns (stream PURGE), the threshold value is many times greater than the nominal one.

Parametric sensitivity in the last column of the table gives the sensitivity of the thermal power to the changes of individual variables values. Thus, e.g., the value 0.367 for stream COND means that if the measured value of condensate stream increases by 1 t/h, the thermal power value increases by 0.367 MW.

Detection and identification of gross errors

Let us now give two examples from the domain of gross errors detection and identification. We introduce artificially a gross error into our data and our aim is to find it.

Let us begin with an error in the feed water flowrate FW1. According to Tab. CS2-2, the threshold value for this variable is 31.7 t/h. The gross error will be chosen somewhat greater, say 50 t/h. The measured value 383.9 t/h will be increased to 433.9 t/h and new reconciliation carried out. We have the new result

$$Q_{min} = 47.07 \text{ ,}$$

which exceeds the critical value $Q_{crit} = 16.89$. So the gross error has been correctly detected. We further apply program RECON menu *Results – Gross errors*. As a result, we have the following message.

REPORT ON GROSS MEASUREMENT ERRORS

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

Type	Variable	Norm.adjust.
MF	FW1	-6.309
MF	PURGE1	4.219
MF	STEAM1	4.097
MF	FW4	-3.874
MF	FW3	-3.328
MF	FWA	2.158
MF	FWB	2.139
MF	FW2	-2.027

Key:

Norm.adjust. = normalized adjustment (large value => suspicion of gross error)
 MF Mass flow

We see that the program has found suspected variables and shown correctly the greatest suspect (placed as first, having greatest absolute value of normalized adjustment). There is also a great distance between the first and second variables.

Now, however, a less favorable situation will be shown. We introduce a gross error into the flowrate FWA (feed water into feed water collector), of value +70 t/h (the threshold value is here 55.3 t/h, see Tab. CS2-2). The measured value 773.8 t/h has thus been increased to 843.8 t/h. After reconciliation, one has found the value $Q_{min} = 33.95$, with critical value 16.89. A gross error has thus been again detected.. We have further applied again the method for suspected values identification giving the following result.

MESSAGE ON GROSS MEASUREMENT ERRORS

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

Type	Variable	Norm.adjust.
MF	FWB	-5.174
MF	FWA	-5.158
MF	PURGE4	2.644
MF	STEAM4	2.555
MF	PURGE3	2.008

Key:

Norm.adjust. = normalised adjustment (large value => suspicion of gross error)
 MF Mass flow

In this case, we already have not been that successful as in the preceding case. We have two suspects – both streams of feed water with mutually close values of normalized adjustment. These two suspects cannot be further distinguished by the given method. In practice, only another independent method could be applied (judging whether the increase in flow over the current limit is possible at all, or scrutinize the measurement system for the two streams).

The impossibility of distinguishing the two streams follows from the fact that in the scheme, the streams are parallel and in the balance of the two nodes CONDHEAD and FWHEAD, they make themselves equally valid. In practice, more cases of similar (although not this obvious) situations can occur. One then often doesn't deal with one error only. As a consequence, the results of gross errors identification are not always unambiguous.

Literature

- [1] V.V.Veverka and F. Madron: *Material and Energy Balancing in the Process industries*. Elsevier, Amsterdam 1997
- [2] F. Madron, *Process Plant Performance. Measurement and data processing for optimization and retrofits*. Ellis Horwood New York, 1992
- [3] J. A. Romagnoli, M. C. Sanchez, *Data Processing and Reconciliation for Chemical Process Operations*, Academic Press London, 2000
- [4] S. Narasimhan and C. Jordache, *Data Reconciliation & Gross Error Detection. An Intelligent Use of Process Data*,. Gulf Publishing Company Houston, 2000
- [5] V.V. Veverka, *Balancing and Data Reconciliation Minibook*, www.chemplant.cz
- [6] Bagajewicz M.J.: *Process Plant Instrumentation: Design and Upgrade*. Technomic, Lancaster 2001
- [7] Humphries, M., Hawkins, M.W. and M.C.Dy: *Data Warehousing. Architecture and Implementation*. Prentice Hall Europe, 1999

Appendix 1: Distribution of random variables

We shall limit ourselves to two distributions of random variables, which are important from the viewpoint of measurement results processing.

D1.1 Normal (Gaussian) distribution

The normal distribution is the most important distribution of a continuous random variable by which, under certain conditions, also other distributions can be approximated. The probability density of the normal distribution is given by the function

$$f(x) = \frac{1}{\sigma(2\pi)^{1/2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \quad (\text{D1-1})$$

This function has two parameters, μ and σ , from which μ equals the mean and σ the standard deviation of the random variable. The normal distribution is written, in abridged form, $N(\mu, \sigma^2)$. Examples of probability densities are in Fig. D1-1.

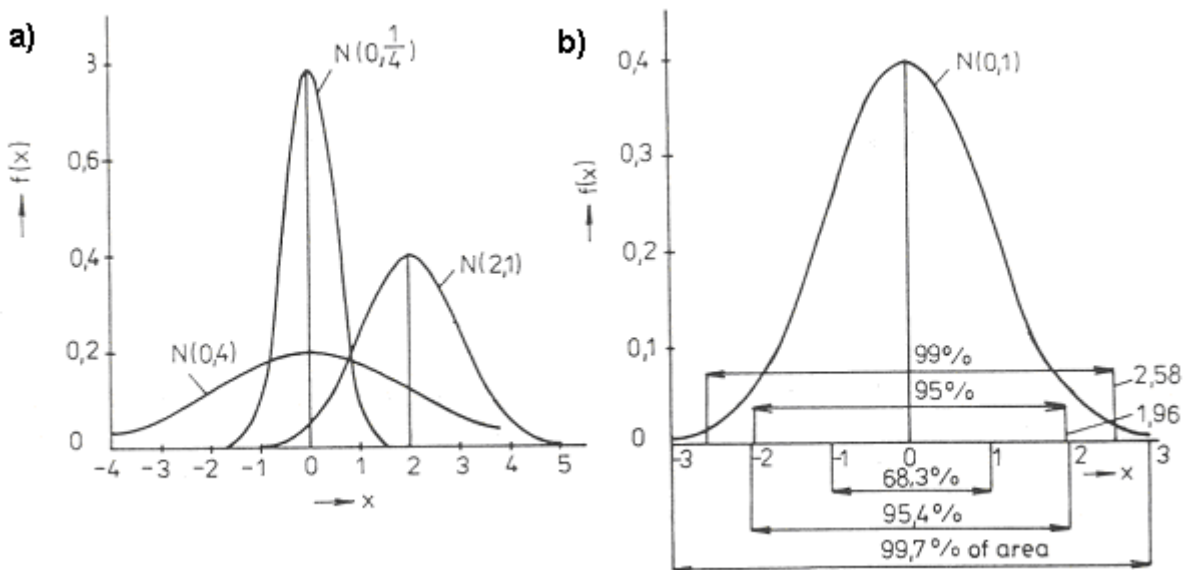


Fig. D1-1: Probability densities of normal distribution

If $\mu = 0$ and $\sigma = 1$, one speaks of so-called *standard (normalized) distribution* $N(0,1)$ and the random variable is denoted by U . The quantiles of the standardized normal distribution denoted by u_p are given in Tab. D1-1 (see also Fig. D1-1).

Tab. D1-1: Quantiles of standardized normal distribution

$$P(U < u_P) = P$$

P	0,500	0,900	0,950	0,975	0,990	0,995	0,999
u_P	0,000	1,282	1,645	1,960	2,326	2,576	3,090

The importance of the standardized normal distribution consists in the universality of its application. We arrive at it from the random variable X , normally distributed $N(\mu, \sigma^2)$, if performing the transformation $U = (X - \mu)/\sigma$. Random variable U then has the distribution $N(0,1)$.

If we need to know the quantile of the distribution $N(\mu, \sigma^2)$, we proceed in the manner that from Tab. D1-1, we take the corresponding quantile of standardized normal distribution u_P and the required quantile x_P is obtained from the relation $x_P = \mu + \sigma u_P$.

The normal distribution is symmetric, from where follow the relations for probability density, distribution function and quantiles (due to the symmetry and also zero mean of the distribution)

$$f(u) = f(-u) \quad (D1-2)$$

$$F(u) = 1 - F(-u) \quad (D1-3)$$

$$u_P = -u_{1-P} \quad (D1-4)$$

Notice in particular that for the values $\alpha < 0.5$ we have

$$P\{|U| < u_{1-\alpha/2}\} = 1 - \alpha \quad (D1-5)$$

D1.2 Distribution χ^2

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

$$\chi^2 = U_1^2 + U_2^2 + \dots + U_\nu^2 \quad (D1-6)$$

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 given, for certain degrees of freedom, in Fig. D1-2. The quantiles are given in Tab. D1-2.

Tab. D1-2: Quantiles of the χ^2 -distribution for ν degrees of freedom

$$P[\chi^2 < \chi_P^2(\nu)] = P$$

ν	P				ν	P			
	0,900	0,950	0,990	0,999		0,900	0,950	0,990	0,999
1	2,706	3,841	6,635	10,827	16	23,542	26,296	32,000	39,252
2	4,605	5,991	9,210	13,815	17	24,769	27,587	33,409	40,790
3	6,251	7,815	11,345	16,268	18	25,989	28,869	34,805	42,312
4	7,779	9,488	13,277	18,465	19	27,204	30,144	36,191	43,820
5	9,236	11,070	15,086	20,517	20	28,412	31,410	37,566	45,315
6	10,645	12,592	16,812	22,457	21	29,615	32,671	38,932	46,797
7	12,017	14,067	18,475	24,322	22	30,813	33,924	40,289	48,268
8	13,362	15,507	20,090	26,125	23	32,007	35,172	41,638	49,728
9	14,684	16,919	21,666	27,877	24	33,196	36,145	42,980	51,179
10	15,987	18,307	23,209	29,588	25	34,382	37,652	44,314	52,620
11	17,275	19,575	24,725	31,264	26	35,563	38,885	45,642	54,052
12	18,548	21,026	26,217	32,909	27	36,741	40,113	46,963	55,476
13	19,812	22,362	27,688	34,528	28	37,916	41,337	48,278	56,893
14	21,064	23,685	29,141	36,123	29	39,087	42,557	49,588	58,302
15	22,307	24,996	30,578	37,697	30	40,256	43,773	50,892	59,703

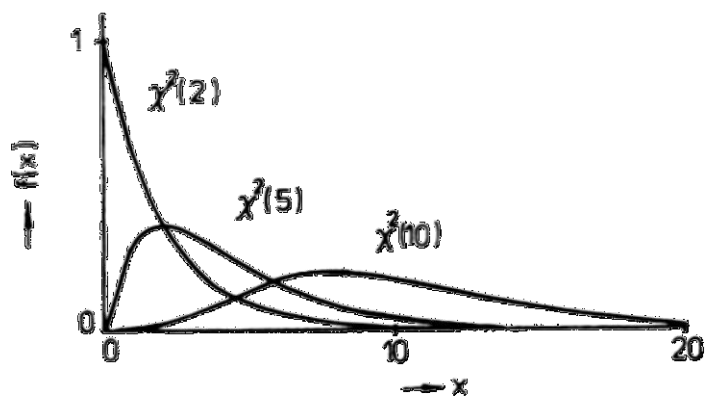


Fig. D1-2: Probability densities of the χ^2 -distribution

For the mean and dispersion we have

$$E[\chi^2(\nu)] = \nu \tag{D1-7}$$

$$D[\chi^2(\nu)] = 2\nu \tag{D1-8}$$

Appendix 2: Measurement errors

Practically every variable obtained by measurement is charged with some error. At the processing of the measurement results and their interpretation, the problems due to measurement errors are thus of considerable importance. They are dealt with in the theory of errors that can be divided into general and special. *General theory of errors* deals with general laws governing their origin and propagation during the measured data processing, and with the methods of acquiring information on the errors based on the measurement proper. *Special theory of errors* concerns errors of individual measuring methods, instruments etc.

This chapter gives results of the theory necessary for applying the methods of measured data processing analyzed in the preceding chapters.

D2.1 Basic concepts and classification of errors

Absolute error of measurement (further only measurement error) e is defined by the relation

$$x^+ = x + e \quad (\text{D2-1})$$

where x^+ is measured and x actual value. Symbol $+$ will be further used for distinguishing the measured and actual values. For judging the correctness of the results, it is convenient to introduce so-called *relative error* e_{rel} :

$$e_{\text{rel}} = \frac{e}{x} \quad (\text{D2-2})$$

expressed often in per cents

$$e_{\text{rel}} = \frac{e}{x} 100 \quad (\text{D2-3})$$

In the majority of cases in practice, one does not deal with one measured variable, but generally with the measurement of I variables (temperatures, flowrates and the like), so that one speaks of a measured vector \mathbf{x} . The measurement of individual variables can be of continuous (temperatures, pressures) or discrete character (identifying concentrations in samples taken periodically). For the needs of the following, computerized processing, continuous measurements are discretized, mostly by reading the continuous signal at a priori determined moments of time.

Let us assume that vector \mathbf{x} was measured at times t_k , $k = 1, 2, \dots, K$. As the result of measurement, one has thus obtained IK values x_{ik}^+ , where index ik means measurement of i -th variable at time t_k . The measured values form a field of two dimensions. One of the dimensions is *time-like*, represented by index k , the other, so-called *space-like*, is represented by index i . In practice, there exist most frequently more complicated data systems. For our purposes, this simple scheme will suffice

and we shall further refer to it. For measured values x_{ik}^+ , we have the relation analogous to Eq. (D2-1):

$$x_{ik}^+ = x_{ik} + e_{ik} \quad (D2-4)$$

According to their character, one usually divides the errors into *random*, *systematic* and *gross*. The importance of error classification consists in the fact that the measured data processing differs mainly according to how the individual kinds of errors participate in the whole error of the measurement result.

Purely random errors

Purely random are such errors that under the same measurement conditions, can attain different magnitudes and have different signs. Their possible values oscillate around zero and their mean is zero. As single ones, they are subject to no laws and independent of one another. Single values are unforeseeable and not substantiated by any reason. They are a typical example of random variables.

These properties of purely random errors can be expressed mathematically in terms of the concepts introduced for random variables :

$$\text{zero mean} \quad \mu_e = E(e) = 0 \quad (D2-5)$$

$$\text{uncorrelatedness of errors} \quad \text{cov}(e_1, e_2) = E(e_1 e_2) = 0 \quad (D2-6)$$

The statistical independence of errors and their uncorrelatedness are equivalent only for errors obeying the normal distribution law (see Appendix 1 for more details). In practice however, the conditions (D2-5) and (D2-6) are currently accepted as the properties of purely random errors, without bothering about the precise interpretation of statistical independence.

Due to the zero mean value of the errors, the mean of the directly measured variable equals the actual value.

$$\mu_x = E(x + e) = x \quad (D2-7)$$

Random errors can be characterized similarly as random variables. Their mean is zero by hypothesis and their most relevant characteristic is the dispersion defined (considering the zero mean value) by the relation

$$D(e) = E(e^2) = \sigma_e^2 \quad (D2-8)$$

The dispersion of a measured variable is here clearly equal to that of its random error and the two concepts are frequently confused. However, this can be done only if the actual value is constant in time. So it is necessary to make a distinction between the dispersion of a measured variable (e.g. with fluctuating actual value) and that of the measurement error, and identify also further possible sources of the measured value variability.

More complete information about random errors is provided by the probability distribution of the random error, expressed e.g. as the probability density. Of key

importance is here the well- known normal distribution law (see Appendix 1). The importance of normal distribution is due to several reasons:

- as found by a number of authors, it approximates generally well the behavior of measurement in different branches of natural sciences, in particular in the domain $\mu \pm 3\sigma$; after all, the results lying outside this interval are usually regarded as unsuitable for the model applied and excluded as outliers,
- the error is often given by the sum of a greater number of single, so-called primary errors. Under fairly acceptable conditions, according to the *central limit theorem* the distribution of this sum approaches the normal distribution,
- if the distribution of errors is unknown (which is a frequent case), the normal distribution hypothesis introduces minimum of (possibly incorrect) information into the problem,
- the normal model based theory of errors is well elaborated and well treatable mathematically. The probability density and standard normal distribution function values are currently tabulated, which makes the solution of practical problems more easy.

Errors obeying the normal distribution are fully characterized by the parameter of this distribution – *standard deviation*, which simultaneously determines the precision of the measurement. The smaller the standard deviation, the more precise is the measurement. In the theory of errors, certain other measures of precision have been established; under the normal distribution hypothesis, they can be expressed as unambiguous functions of the standard deviation.

In older literature (and not quite appropriately), standard deviation is denoted as *mean quadratic error*. Its doubled value determines so-called width of the Gauss curve, i.e. the distance between its inflexion points.

D2.2 Correlated random errors

We can arrive at this concept by the following consideration. Let us suppose that the measurement errors of two variables are formed by sums of a number of uncorrelated random errors. If no elementary error participates simultaneously in giving rise to both of the summary measurement errors, then also the summary errors are uncorrelated. In the opposite case however, the errors present simultaneously in both of the sums make the resulting errors correlated.

Let for example the measurement errors of variables A and B be given as sums of elementary uncorrelated errors e_1 , e_2 and e_3 with zero means and dispersions σ_1^2 , σ_2^2 and σ_3^2 .

$$e_A = e_1 + e_2 \qquad e_B = e_1 + e_3 \qquad (D2-9)$$

The covariance of errors e_A and e_B then is

$$\text{cov}(e_A, e_B) = E(e_A e_B) = E[(e_1 + e_2)(e_1 + e_3)] \qquad (D2-10)$$

After multiplication and rearrangement we have

$$\text{cov}(e_A, e_B) = E(e_1^2) + E(e_1 e_3) + E(e_2 e_1) + E(e_2 e_3) = \sigma_1^2 \quad (\text{D2-11})$$

because the elementary errors are uncorrelated and the last three mean values in Eq. (D2-11) are therefore null. So in this case, the covariance of errors e_A and e_B equals the dispersion of error e_1 present in both summary errors.

Let us now occupy ourselves first with the *space correlatedness of errors*. For directly measured variables, this kind of error correlatedness is relatively rare. The space correlatedness is usually caused by the fact that individual measuring circuits are not independent. Thus for example in temperature measurements by thermocouples, the reference connections of the thermocouples are usually tempered in one thermostat. If the thermostat temperature is not well stabilized, an elementary error thus arisen manifests itself at all measured temperatures simultaneously. In an analogous way, space correlated errors can arise as a consequence of air pressure fluctuations in the net for feeding the control devices. In analytical determinations, correlated errors may arise as a consequence of taking a sample that does not correspond to the bulk stream composition. If for instance some component condenses on the path of taking the sample, the concentrations found for the other components will all be higher than their actual bulk values.

Most frequently however, the space correlatedness of errors is met with at the calculation of secondary variables from the directly measured (primary) ones.

Let us suppose that we have to measure the flowrates of species A and B in a certain liquid stream. One measures the flowrate of the liquid phase, takes the sample and determines the species A and B concentrations therein. The individual species flowrates are then obtained as products of flowrate and individual concentrations. Even if the primary values measurement errors are uncorrelated, the flowrate measurement error occurs simultaneously in calculating both individual species flowrates, so that the resulting errors are already correlated.

The space correlatedness of errors can be quantitatively expressed with the aid of the covariance matrix of errors. Having vector \mathbf{x} of I measured variables, to which corresponds vector \mathbf{e} of errors, the covariance matrix \mathbf{F} of type $I \times I$ is, assuming zero mean of errors, of elements

$$F_{ij} = \text{cov}(e_i, e_j) = E(e_i e_j) \quad \text{for } i \neq j \quad (\text{D2-12})$$

$$F_{ii} = D(e_i) = E(e_i^2) \quad (\text{D2-13})$$

Let us now consider the *time correlatedness of errors*. Any error has its objective cause that is, in the end, of physical character. Due to inertia, one has to assume that this cause will last for a certain time. It depends only on the length of the time interval between two measurements, if the measurement errors in two following time instants will be dependent, or almost independent. The time correlatedness of errors is a very frequent phenomenon, in particular in the context of ever increasing application of automatic measurement systems. It can be time-variable errors of measuring instruments, errors due to fluctuations of technological variables, and the like. The time correlatedness can be quantitatively expressed using the autocorrelation function.

Example D2-1: Autocorrelation function of errors in a time series

Suppose we are measuring one variable x in times t_k , $k = 1, \dots, K$, where $t_k - t_{k-1} = \Delta t$. We thus obtain time series of K measured values x_k^+ , charged with errors e_k . The errors are assumed to be purely random errors with the same dispersion σ_x^2 . For the needs of control we carry out the filtration of data, with the end to decrease the influence of errors on the control actions. We calculate the variables

$$z_i^+ = \frac{x_i^+ + x_{i-1}^+}{2} \quad i = 2, 3, \dots, l \quad (D2-14)$$

and obtain thus a new time series with $(l - 1)$ terms. For the errors in variables z_i^+ it holds obviously

$$e_{z_i} = \frac{e_{x_i} + e_{x_{i-1}}}{2} \quad (D2-15)$$

The error covariances of the neighboring variables z_i and z_{i-1} then equal

$$\text{cov}(e_{z_i}, e_{z_{i-1}}) = E \left[\frac{(e_{x_i} + e_{x_{i-1}})(e_{x_{i-1}} + e_{x_{i-2}})}{4} \right] \quad (D2-16)$$

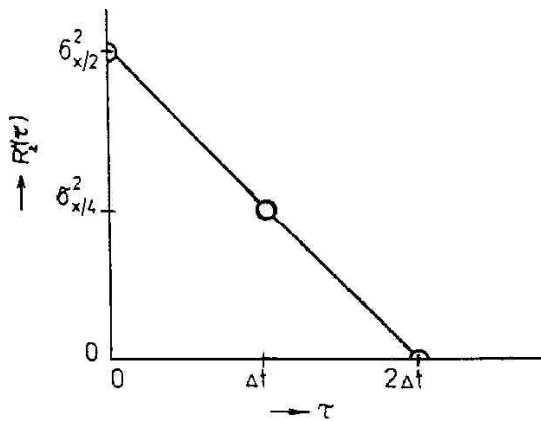


Fig. D2-1: Autocorrelation function of errors from Eq.(D2-1)

After multiplication and considering the uncorrelatedness of errors e_{z_i} we obtain the final expression

$$\text{cov}(e_{z_i}, e_{z_{i-1}}) = E\left(\frac{e_{x_{i-1}}^2}{4}\right) = \frac{\sigma_x^2}{4} \quad (\text{D2-17})$$

In an analogous manner we could find that the errors in variables z_i and z_{i-2} are already uncorrelated. The autocorrelation function of time series z_i^+ , defined only for integer multiples of time interval Δt , is given in Fig. D2-1 ♦

Concluding the section about dependent, resp. correlated errors let us note in addition that one has dealt with a stochastic dependence, i.e. with a dependence valid between large ensembles of errors. The single errors were still random and unforeseeable. One seldom meets with deterministic dependence of errors in the sense of the vector dependence introduced in linear algebra. In such cases, one or several errors can be expressed exactly as function of the remaining ones. The accompanying phenomenon is the fact that the covariance matrix of deterministically dependent errors is singular. This example represents the transition from random to systematic errors.

D2.3 Systematic errors

So far, we have treated the errors as random variables. These errors then either changed quite randomly, or were stochastically bound together. The notion *systematic* error means an error whose value is constant in time, or has a deterministic course. It can be for instance an error caused by imperfectly adjusted measuring device (constant error), an error linearly time dependent due to the shift of the instrument's zero in time, an error with periodic course due to the daily course of ambient temperature, and the like.

Let us examine the relation between systematic and time-correlated errors. In the context of time-correlated errors, it has been stated that it depends only on the length of the time interval between two measurements and on the time variability of the error-causing influence, whether the errors will be time correlated. By this consideration, one can arrive even at the situation where the factors causing an error will, relative to the time of measurement, change only insignificantly. From this point of view, a constant systematic error can be regarded as a limit case of time-correlated error. In reality, the boundary between both kinds of errors is diffuse and often cannot be stated at all.

Similar conclusions hold also for certain space correlated errors. For example in the earlier given example concerning space correlated errors of temperature measurement by thermocouples, the summary errors in measuring individual temperatures were random because one assumed that besides imperfect temperation, also other elementary (random) factors were in action. If these other influences were negligible, knowing the measurement error for one temperature we could exactly determine also the errors for the other ones. The resulting error could then be regarded as systematic.

Let us further mention the random character of systematic errors. For example, let us consider a set of measuring instruments of the same type and produced in a certain period. If these instruments were used for measuring certain constant variable, one probably would find that the instruments don't measure quite equally. The set of individual instruments' errors forms a sampling space. If selecting randomly one

instrument (e.g. on buying it) and use it for the measurement, we put a constant systematic error into the measurement. However, the value of this error can be regarded as a sample of range 1 from the given sampling space. In other cases, the sampling space can be hypothetical – it can be for instance the set of all admissible ways of flowrate measurement by an orifice, where the individual ways are given mainly by the pipeline and orifice opening diameters. By selecting certain variant, we have taken a random sample from the hypothetical sampling space.

The actual value of a constant systematic error is, in the given cases, unknown and can be regarded as realization of a random variable. Understanding the mutual relations between random and systematic errors is important in the statistical measured data processing. As will be shown later, the solution of practical problems requires to take into account simultaneously and in unified manner all available information about random and systematic errors.

Concerning the systematic error problems, let us finally stress their importance at the process measurements. In the literature about the theory of measurement, these problems are usually solved by the remark that it is necessary to avoid systematic errors. One recommends a careful calibration of the instruments, using standards and the like. This approach, which is legitimate in a physical or chemical laboratory, doesn't however solve the problem of systematic errors at industrial measurements. Even if we take all feasible measures, the systematic errors will essentially influence the final results of measurement. It is thus necessary to take into account their existence and employ the information about them in the mathematical measured data processing, as is common at purely random and correlated errors.

D2.4 Gross errors

As a *gross error* is regarded a single large error arisen as a consequence of inattention, measurement system failure, erroneous calculation, unforeseen event and the like. By its magnitude, it exceeds the frame of the other errors. Its origin, if regarded as realization of a random error, would be quite unlikely. If its cause is not removed in due time, it occurs repeatedly and we have a gross systematic error.

While random and systematic errors are inseparable part of measured values and measurements charged with them are used in the formulation of measurement results, the measurement charged with a gross error must be eliminated from further treatment. It is thus an important task of acquiring correct results, to analyze the measured data from the viewpoint of finding measurements subject to gross errors and their elimination.

D2.5 Measurement precision and accuracy

Concerning the occurrence of random and systematic errors, two important concepts are introduced – *measurement accuracy* and *measurement precision*.

The *accuracy of measurement* means the agreement between the measured and true values. The *precision of measurement* then represents the agreement among a series of repeated measurements of the same value. The difference between the two concepts consists in the fact that unlike the accuracy, the precision of measurement takes no account of systematic measurement errors. The precision and accuracy are depicted graphically in Fig. D2-2.

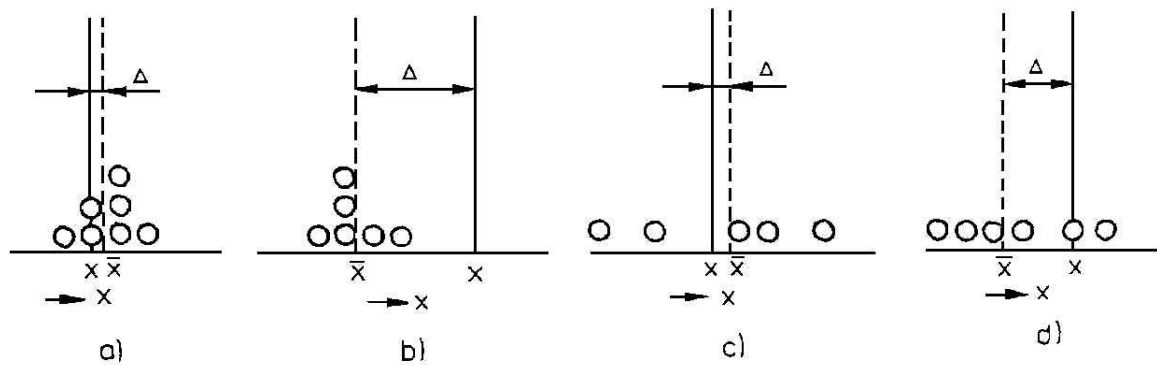


Fig. D2-2: Accuracy and precision of measurement
a) precise and accurate measurement **b)** precise and inaccurate measurement (incorrect mean value) **c)** imprecise measurement with correct (accurate) mean **d)** imprecise and incorrect measurement

Suppose we repeatedly measure one variable x , with true (correct) value x . If each measurement result is marked by drawing a circle at the respective value on the horizontal axis, we can obtain the following characteristic cases.

In Fig. D2-2 a), we have a well reproducible measurement and the mean \bar{x} is close to the true value (precise and correct measurement is sometimes called *reliable measurement*). In Fig. D2-2 b), the measurement is again well reproducible, but the measured values deviate systematically from the true value. We have to do with a precise, but inaccurate measurement (the mean is incorrect). In Fig. D2-2 c) is depicted the case where the measurement is badly reproducible, but with a great number of measurements, the mean is close to the correct (true) value. One then speaks of an imprecise measurement with correct mean (for a sufficiently great number of repeated measurements). The last, least desirable variant is an imprecise and incorrect measurement, where the measurement is badly reproducible and deviates systematically from the true value. It is of course obvious that the boundaries of the given classification depend, to a great extent, on convention.

A good insight into the problems of measurement precision and accuracy is important for measurement planning. The best thing is naturally a precise and correct (thus in the mean accurate) measurement; however sometimes, it is difficult to attain this goal. Sometimes, we can put up with a precise, but less correct measurement (e.g. when comparing the values measured under different technological conditions, where a constant systematic error need not make a serious trouble). In other cases, we require correct measurement, while small precision of the method applied can be eliminated by calculating the average from a greater number of repeated measurements.

Appendix 3: Testing of statistical hypotheses

We are sometimes interested in whether certain assertion on data properties holds true (e.g. the assertion that a gross error is present in the data). Such assertions are called *statistical hypotheses*.

The procedure, by which we decide on the truthfulness of the postulated hypothesis, is so-called *testing of statistical hypothesis*. The hypothesis to be verified is called the *null hypothesis*, the opposite one *alternative hypothesis*. The test of the null hypothesis H_0 against the alternative one H_1 leads, on the basis of examining a random selection (sampling), either to the rejection of H_0 , or to non-rejection. (i.e. rejection of H_1). Here, the non-rejection of hypothesis H_0 must not be identified with its acceptance, although the practical consequences of the two concepts use to be the same. The non-rejection of a hypothesis simply means that on the basis of available information, one has no reason to doubt of it.

The testing procedure: We have postulated certain hypothesis about the basic ensemble. We carry out a selection (sampling) of range n and have vector $\mathbf{x} = (X_1, \dots, X_n)$. We choose an appropriate statistic $T = T(X_1, \dots, X_n)$, called in this case *testing criterion*. One finds the distribution of random variable T under the assumption that the null hypothesis (to be tested) holds true. The interval where the statistic can lie [e.g. $(-\infty, +\infty)$], is divided into two intervals, interval R and its complement R^+ . R is chosen in the manner that under the validity of H_0 , the statistic T takes values from this interval with probability $(1 - \alpha)$. Most frequently, the choice of this interval is such that given α , its length is minimal (near to zero). The value of α is chosen sufficiently small (e.g. 0.05); it is called the *level of significance*. The region R^+ is called *critical region*. Now, from the random selection (sample), one calculates the statistic T . If T assumes a value from the critical region R^+ , hypothesis H_0 is rejected.

When testing the hypotheses, one can commit basically two kinds of errors.

Error of the first kind consists in rejecting the hypothesis on the basis of the random selection while in reality, it holds true. The probability of this kind of error equals the level of significance α .

If hypothesis H_0 does not hold, but is not rejected on the basis of our random selection, we speak of an *error of the second kind*. The probability of a second kind error is denoted as β and the value of $(1 - \beta)$ is called the *power of the test*. While the probability of a first kind error was, for a given test, one single number (equal to the significance level), the power of the test depends on how much the null hypothesis deviates from the reality. If we are able to measure somehow this deviation, the dependence of variable β on the deviation is called *operating characteristic of the test*. The corresponding *power of the test* thus also depends on the deviation of the hypothesis from reality; the latter dependence is called *power characteristic* of the test.

Let us now illustrate this important topic by a simple example.

Example D3-1: Testing a statistical hypothesis

A liquid flows through a pipe with volume flowrate Q_V . On the pipe, two flowmeters are installed: a common flowmeter A and control flowmeter B. For the values measured by the respective flowmeters $Q_{V,A}$ and $Q_{V,B}$ we have

$$Q_{V,A} = Q_V + e_A ; Q_{V,B} = Q_V + e_B \quad (D3-1)$$

where e_A and e_B are random errors of measurement for the respective flowmeters.

The errors are assumed to behave as normally distributed random variables with zero mean and dispersions σ_A^2 and σ_B^2 . Flowmeter A is prone to give a systematic error; our aim is to judge, on the basis of the measured values, whether such an error actually arises in flowmeter A.

Let us formalize the problem mathematically.

The above equations are rewritten as

$$Q_{V,A} = Q_V + e_s + e_A ; Q_{V,B} = Q_V + e_B \quad (D3-2)$$

where e_s is systematic error of flowmeter A. The hypothesis that the flowmeter is not charged with a systematic error will be written in the form $e_s = 0$ (hypothesis H_0), the alternative hypothesis H_1 will be written as $e_s \neq 0$.

Let us now carry out the sampling, represented in our case by measuring the values of $Q_{V,A}$ and $Q_{V,B}$. As a statistic, we take the measured values difference

$$T = Q_{V,A} - Q_{V,B} = e_s + e_A - e_B \quad (D3-3)$$

Assuming the validity of H_0 (i.e. $e_s = 0$), T equals the sum of two random variables with zero means and normal distribution, and has itself also normal distribution $N(0, \sigma_A^2 + \sigma_B^2)$. It can thus take the values from the interval $(-\infty, +\infty)$.

Let us further choose the significance level α and the corresponding interval R . In Fig. D3-1 a), we have the probability density of statistic T . Since T has normal distribution with zero mean, the interval R will be chosen symmetric around 0

$$R = \langle -\sigma_T u_{1-\alpha/2} ; \sigma_T u_{1-\alpha/2} \rangle \quad (D3-4)$$

where $\sigma_T = (\sigma_A^2 + \sigma_B^2)^{1/2}$ and $u_{1-\alpha/2}$ is the quantile of distribution $N(0,1)$.

Let us now assume the validity of H_0 thus $e_s = 0$. In 100α per cent of cases the value of T falls into the interval R^+ and the hypothesis will be unduly rejected. The probability of this error of the first kind is depicted by the hatched area in Fig. D3-1 a).

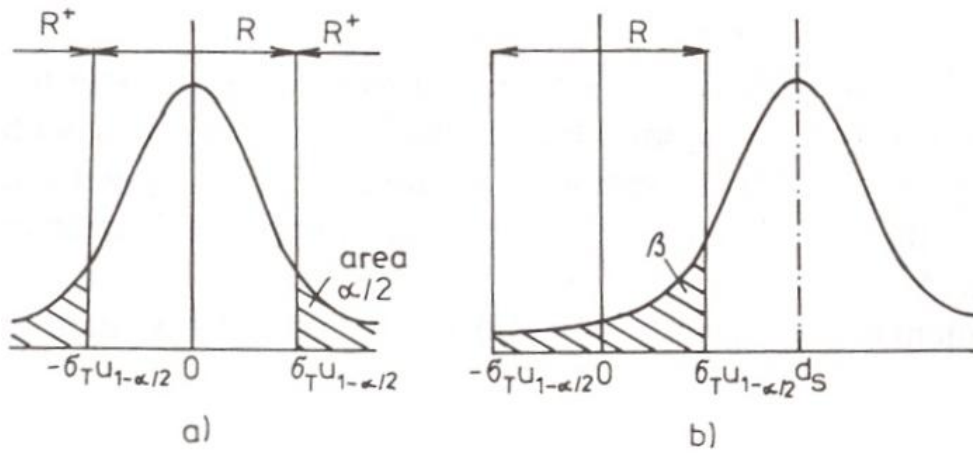


Fig. D3-1: Magnitude of errors of Ist kind (α) and IInd kind (β)

a) $e_s = 0$ **b)** $e_s \neq 0$

Let us conversely assume that $e_s \neq 0$. The statistic T has, in this case, the mean value e_s and the distribution illustrated in Fig. D3-1 b). The hatched area here represents the probability of an error of the second kind (T falls into the interval R and the hypothesis H_0 is not rejected). Clearly, the probability of a IInd kind error decreases rapidly with increasing value of e_s ♦

Appendix 4: Effectiveness of gross error detection

The gross error detection effectiveness cannot be judged generally, but only for individual concrete measured variables. It is thus not a property of the whole model.

One has to begin with testing the gross error presence hypothesis (see the preceding Appendix 3).

As any statistical test, also the χ^2 test has its *power characteristic* depicted in Fig. D4-1.

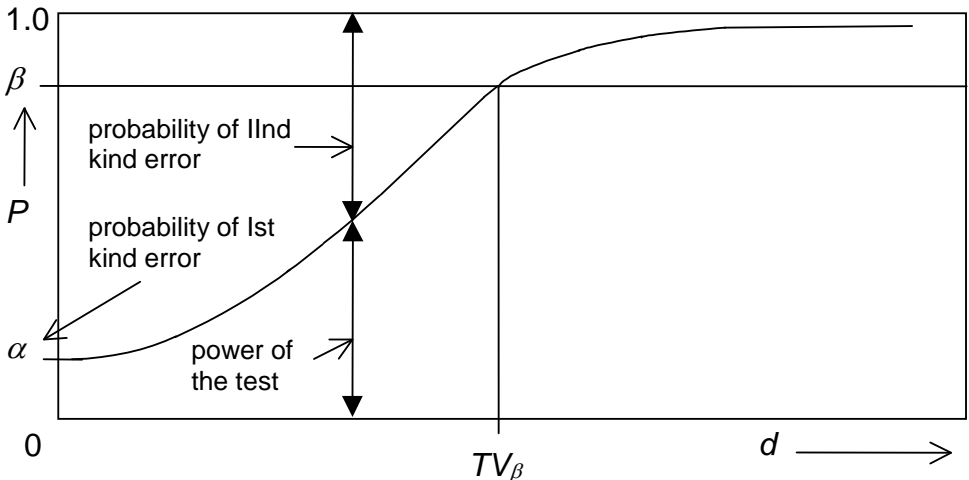


Fig. D4-1: The *power characteristic* of the χ^2 test

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 an adjustable measured variable equals the significance level α of the test assuming the absence of gross error ($d=0$), and it approaches 1 for high values of the gross error ($d \rightarrow \infty$).

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

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

The threshold value is computed from the equation

$$q_i = \delta_\beta(v, \alpha) / [a_i(2-a_i)]^{1/2} \tag{D4-1}$$

where q_i is physically dimensionless threshold value TV/σ

$$q_i = TV/\sigma_i \tag{D4-2}$$

and $\delta\beta(\nu, \alpha)$ is a constant, characteristic of 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 [2].

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

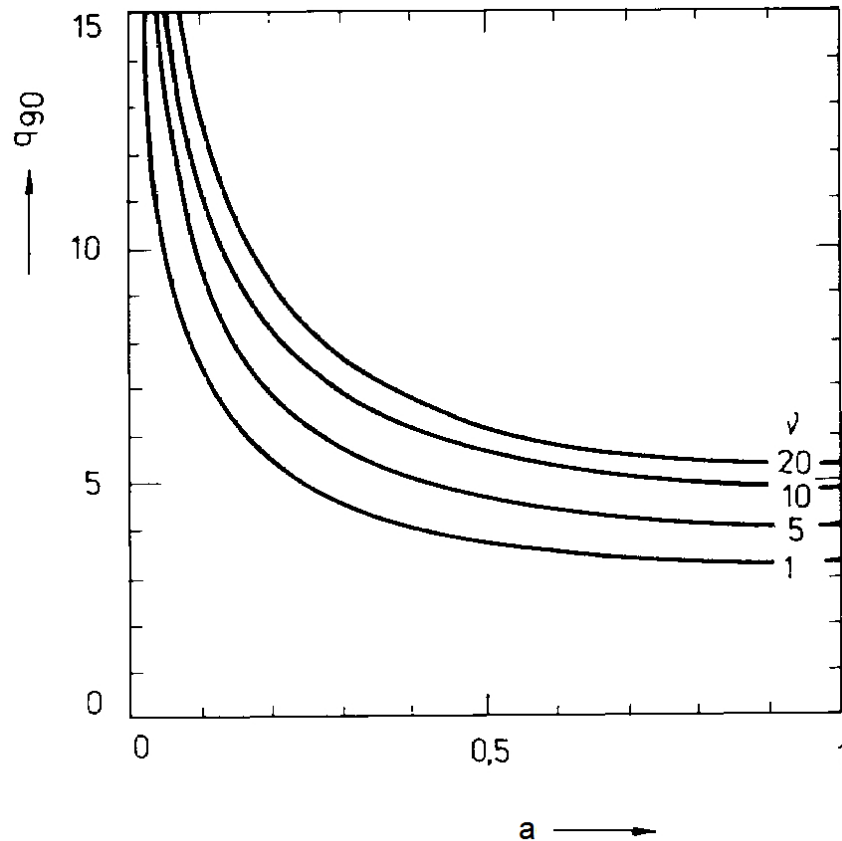


Fig. D4-2: Dimensionless threshold value q as function of the degree of redundancy ν and adjustability a (for $\alpha=0.05$ and $\beta=0.9$)

From this diagram, one can derive certain simple conclusions:

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

Appendix 5: Momentum balance in a pipeline

The detailed balance is quite complicated. It will be given only in simplified form obtained by integrating the general differential balances for the conservation of mass and momentum (product of density and flow velocity vector at a given point); the integration is over the points of a given region in three-dimensional space.

We consider stationary streaming in a segment of circular tube with diameter D (m), cross section $S = \frac{\pi D^2}{4}$ (m²) and length L (m). (D5-1)

The pipeline need not be horizontal; it generally comprises ascending (or descending) parts according to the scheme

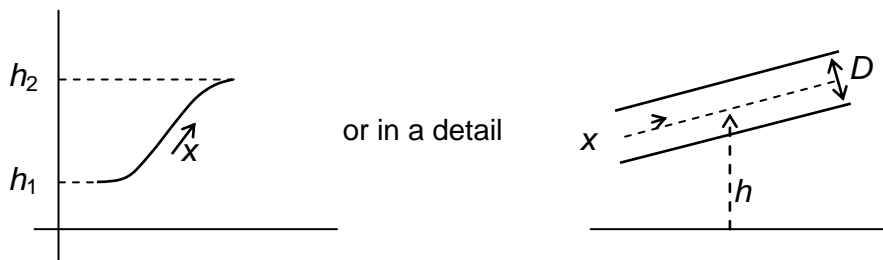


Fig. D5-1: Pipeline segment

Here, x is general coordinate on the tube axis and h the (variable) height above the chosen zero level. The gravitational acceleration will be denoted by g (m/s²). Further

m (kg/s) is constant value of the flowrate. (D5-2)

The pressure P (Pa) changes along the coordinate x . The fluid density ρ (kg/m³) is in our case function of temperature and pressure. Temperature T will be for simplicity considered constant (independent of x). The course of pressure along x is then described by the differential equation

$$\frac{dP}{dx} + \frac{1}{2} \frac{\lambda}{DS^2} m^2 \frac{1}{\rho} + \rho g \frac{dh}{dx} + \frac{m^2}{S^2} \frac{d(1/\rho)}{dx} = 0 \quad (x = 0 \text{ to } x = L) \quad (\text{D5-3})$$

where we have in addition the parameter λ (friction factor of the pipeline). The latter is given in engineering literature as function of the Reynolds number and tube wall roughness. Here, the Reynolds number equals

$$\text{Re} = \frac{mD}{S\mu} \quad (\text{D5-4})$$

where μ is dynamic viscosity (kg/(ms)). The viscosity is generally function of pressure and temperature and can be calculated, as well as density ρ , using available methods (see below). Then, assuming that also the roughness is known, the coefficient λ becomes also function of temperature and pressure; after all in the expression for the Reynolds number, we can estimate the viscosity by a certain mean value and take coefficient λ as a constant calculated (given m) from well-known relations.

Under all the simplifying hypotheses introduced, in Eq. (D5-3) we then have independent variable x , and dependent variable P as itself and implicitly via function ρ (given T); the slope dh/dx is also considered known. Eq. (D5-3) can be generally solved by *numeric methods*.

An *approximate formula* is obtained, if we a priori estimate the corresponding region of pressures and for the variable value of ρ , substitute its mean value $\bar{\rho}$. The last (by the way frequently quite small) term on the left hand side sum in Eq. (D5-3) can be neglected and we have the pressure difference (pressure drop)

$$P_1 - P_2 \cong \frac{1}{2} \frac{\lambda}{DS^2} \frac{m^2}{\bar{\rho}} L + \bar{\rho} g(h_2 - h_1). \quad (\text{D5-5})$$

Density ρ and viscosity μ for liquid water and steam can be computed using database IAPWS – IF 97.

Appendix 6: Classification of variables

The starting point for the following solution is the solvability analysis of a set of linear equations in variables representing measured and unmeasured variables. The so-called *General linear model* is of the form

$$Ax + By + a = 0$$

where

- x is vector of measured variables
- y vector of unmeasured variables
- a vector of constants
- A and B matrices of constants

Let us set up so-called macro-matrix C formed by matrices B and A .

$$C = (B, A) \tag{D6-1}$$

For the sake of simplicity, let us suppose that all the rows of matrix C are linearly independent (which should be the case of correctly set-up balance equations).

In the first step, we carry out the Gauss-Jordan elimination (transformation) of matrix C (see Fig. D6-1, where the void fields represent zeros and the hatched ones certain general elements). In the left upper part of transformed matrix C , we thus have unit matrix, in the lower part of (transformed) submatrix B , there can arise a band of zero rows. This operation may require interchanging rows of matrix C and also columns of submatrix B .



Fig. D6-1: Matrix C after elimination rearrangements of submatrix B

If the horizontal band 2 in submatrix **B** does not arise, the system is not redundant and there is no need of data reconciliation.

Let us now assume that after the elimination, the latter band has arisen, and continue in further elimination in band 2 of the (pre-transformed) submatrix **A**. We thus obtain the matrix schematically depicted in the following figure.

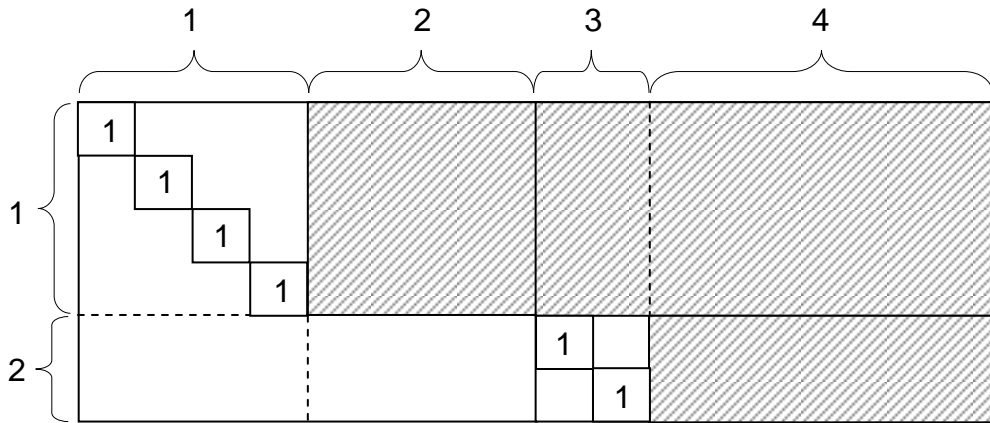


Fig. D6-2: Matrix **C** after elimination rearrangements of submatrix **A**

We further regroup the rows of horizontal band 1 (contingently also the columns of submatrix **B**). Our aim is here to form a zero submatrix – intersection of horizontal band 1a and vertical band 2. We further aim at creating a further zero submatrix – the intersection of horizontal band 2 and vertical band 4b. After all these rearrangements, the resulting form of matrix **C** is called *canonical*.

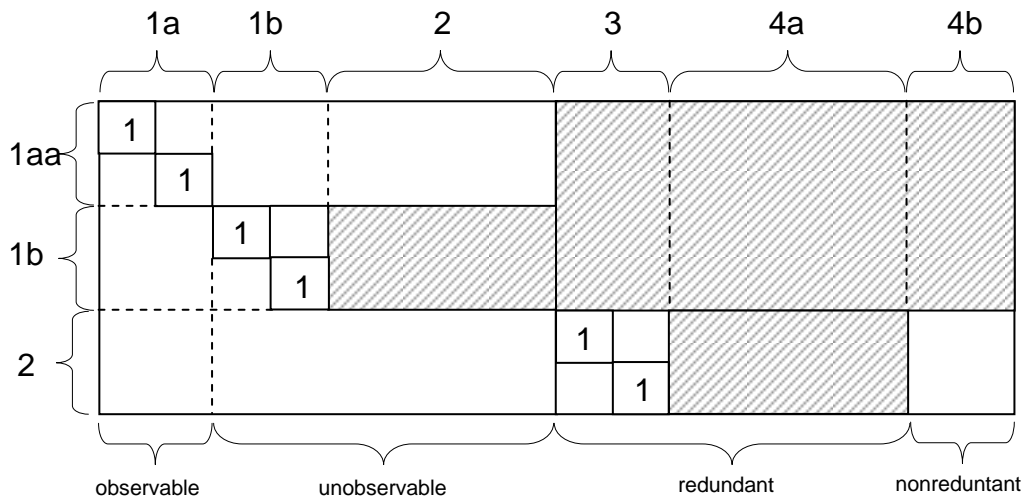


Fig. D6-3: Canonical form of matrix **C**

The canonical form of matrix \mathbf{C} provides all necessary information about the solvability of the system. One only needs to realize that the horizontal bands of the matrix represent groups of equations, and the vertical bands represent groups of variables. The conclusions are following:

- The unmeasured variables in vertical band 1 are uniquely determined by the values of measured variables in vertical bands 3 and 4 – they are thus observable.
- The unmeasured variables in vertical bands 1b and 2 are not uniquely determined and are thus unobservable.
- the reconciliation of measured values concerns only the measured values in vertical bands 3 and 4a. These bands comprise redundant variables to be reconciled.
- If the canonic form of matrix \mathbf{C} contains vertical band 4b, this band represents nonredundant (just determined) measured variables.

Having transformed the system to the canonical form, one can see that the reconciliation proper does not concern the whole set of the (transformed) equations, but most frequently only a small part of it, i.e. horizontal band 2. In the first step, one carries out the reconciliation and the reconciled (adjusted) values are substituted for the measured ones into the equations in horizontal band 1. In the second step, the remaining observable unmeasured variables are computed using the equations in horizontal band 1a.

The so far described procedure of solution can be directly applied to linear models. In the nonlinear case, the procedure can be applied to models linearized in a region sufficiently close to the solution point.

In this Appendix, we have committed certain simplifications (for example we have assumed linear independence of the rows of matrix \mathbf{C}). A complete solution can be found in literature [2]. The above conclusions hold for the case where the covariance matrix of measurement errors is diagonal (uncorrelated errors). If the errors are correlated, it can happen that also nonredundant data are reconciled. This interesting, though not very frequent case is described, e.g., in literature [5].

Appendix 7: About program RECON

ChemPlant Technology s.r.o.

RECON



Mass, heat and momentum balancing software with data reconciliation

Description

RECON in version 7 is a comprehensive interactive software for mass, heat and momentum balancing of complex chemical plants on the basis of measured or otherwise fixed data.

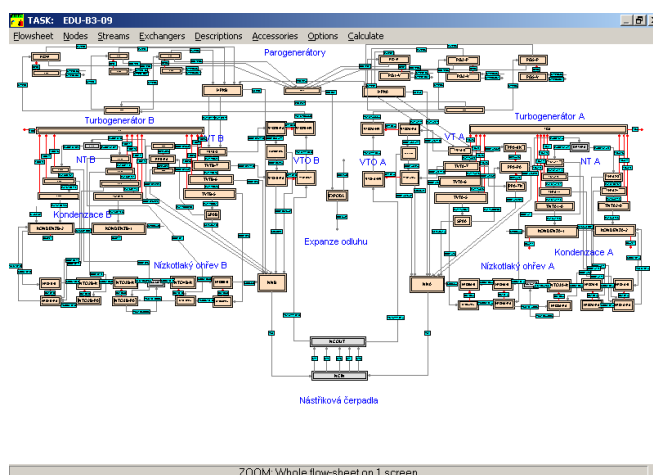
It is designed primarily for the validation of data, which has been obtained from operating processes. RECON can also be used for classical balancing in the stage of the process design.

What is data reconciliation?

Reconciliation is a method for extracting all information present in plant data. Reconciliation is based on statistical adjustment of redundant process data to obey laws of nature (mass and energy conservation laws). As a result, new consistent set of data is obtained. Moreover, data reconciliation serves as a basis for other important activities:

- ◆ Finding confidence intervals for results (error propagation analysis)
- ◆ Detection and elimination of gross measurement errors
- ◆ Measurement planning and optimization.

RECON can be used to treat measured data before its further usage for other purposes (simulation, optimization, control, ...)



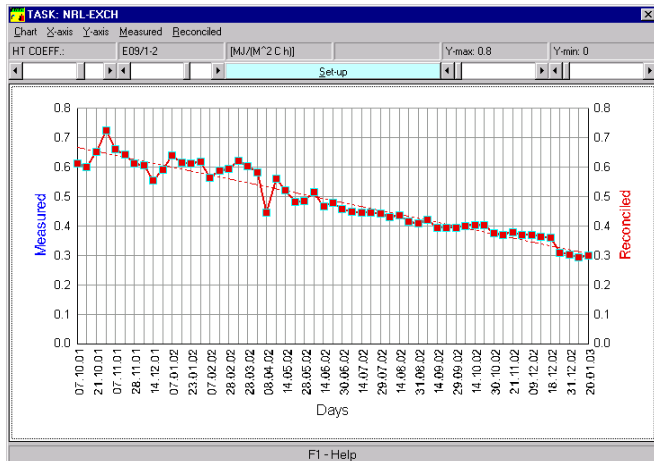
Graphical User Interface of RECON

FEATURES

RECON is PC oriented software with user friendly facilities. Problems (tasks) are defined interactively in the graphical user interface. User manual contains brief survey of reconciliation theory.

RECON is aimed at single or multi-component material and heat balancing of complex systems. It is also capable to perform momentum balancing based on hydraulic calculations. RECON reconciles measured flow rates, concentrations, temperatures and other process variables and calculates unmeasured variables.

Problem (task) is commonly defined by creating process flowsheet and defining process variables like flow rates, temperatures, pressures, etc. The flowsheet comprises nodes, mass and energy streams, and heat exchangers. Users are also allowed to complete (or even replace) balancing model with their own equations.



Monitoring of a heat transfer coefficient in RECON

VARIABLES

The following information on task variables must be specified:

- ◆ Classification of variables (measured, unmeasured, fixed)
- ◆ Values of measured and fixed variables
- ◆ Estimates (guesses) of unmeasured variables (for nonlinear problems only)
- ◆ Maximum errors or standard deviations of measured variables

CAPABILITIES

- ◆ Calculation of unmeasured variables
- ◆ Reconciliation of redundant measured variables
- ◆ Analysis of input data as concerns its consistency
- ◆ Confidence intervals of results
- ◆ Statistical analysis focused on detection and identification of gross measurement errors
- ◆ Detailed classification of variables (redundant / non-redundant, observable / non-observable).
- ◆ Measurement placement optimization
- ◆ Parametric sensitivity

- ◆ Database of historical data useful for monitoring of operating plants

RECON's connectivity

RECON can be connected to process data information systems based on the Oracle, MS SQL, PI System, Industrial SQL server and MS Access databases as well as to MS Excel, .DBF or .TXT files.

RECON's physical property database

RECON contains three databases of physical properties:

- ◆ The IAPWS IF 97 database of properties of steam and water
- ◆ Parameters of BWR equation for density calculations
- ◆ Critical parameters of components for viscosity calculations of gaseous mixtures

Running RECON

RECON can be run in the following modes

- ◆ Interactive solving of one task
- ◆ On-line monitoring of industrial processes
- ◆ Automatic processing of historical data
- ◆ As ActiveX DLL from other applications

HARDWARE REQUIREMENT

RECON is the 32 bit MS Windows application which can be installed on any PC or server operating under MS Windows 95 or higher (98, NT, 2000, XP). The minimum recommended processor is Pentium with RAM 128 MB.

Other software provided by ChemPlant

IBIS - A comprehensive reconciliation plant balancing system (yield accounting in large plants and companies)

PDIS – Process Data Information System. A process database of data like flow rates, temperatures, laboratory data, etc. on the Oracle platform.

ChemPlant Technology, s.r.o.

Vinařská 1015/4
400 01 Ústí nad Labem
Czech Republic
Tel: +420 474 527 221
E-mail: support@chemplant.cz
WebSite: www.chemplant.cz