ChemPlant Technology s.r.o.

Report CPT-475-24

Date of issue: 10.3.2024

This report has 47 pages

Digital Twins in the Process Industries

Author: Vit Madron

Ústí nad Labem March 2024

ChemPlant Technology, s.r.o.
Vinařská 1015/4
400 01 Ústí nad Labem
Czech Republic

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

CONTENTS

LIST OF SYMBOLS AND TERMS

DM	Digital Model
DS	Digital Shadow
DT	Digital Twins
DVR	Data Validation and Reconciliation
EH&S	Environment, Health, and Safety
GED	Gross Errors Detection
HTC	Heat Transfer Coefficient
KPI	Key Performance Indicator
PI	Process Industries
Recon	SW for modeling of industrial processes by ChemPlant
RTO	Real Time Optimization
RS	Real Space
SPP	Smart Process Plants
SW	software
VS	Virtual Space

SUMMARY

1. Chapter 1 is the introduction to the concept of *Digital Twins* (DT). It also mentions the similar concept of *Smart Process Plants* in Process Industries (PI) which was developed earlier than *Digital Twins.*
2. Chapter 2 analyses typical features of data flow and processing in PI. Discussed is the issue of mathematical modeling of physical and chemical processes in PI. The main problem of the mirroring the Real Space into the Virtual Space are measuring errors of physical variables in the Real Space. This issue was thoroughly analyzed in the past century by thousands of papers and by several books about Data Reconciliation and Validation.
3. Chapter 3 is about the Simulation of industrial systems in the Virtual Space. The proposed solution is based on the *Process Data Driven Simulation*. This solution Is based on two steps: (1) Process Data Validation and Reconciliation. In this step the model parameters are identified and (2) the model parameters are used as the inputs for the second step - the Simulation. This approach can be applied automatically or on request by operators in the manual mode (What if Queries?). This Chapter also mentions the possibility of improving the fidelity of models by using empirical models enhanced by historical data.
4. Chapter 4 summarizes areas where Digital Twins in PI can be useful
5. Chapter 5 presents two examples of Digital Twins: (1) the supercritical coal fired power station and (2) the vacuum distillation of the heavy crude oil.
6. Appendix 1 summarizes briefly the process of Data Validation and Reconciliation
7. Appendix 2 describes abilities of the SW **Recon** which was several times mentioned throughout this Report

1 INTRODUCTION

Digital Twins (DT) is nowadays a very hot subject. It is generally acknowledged as the key part of the Industry 4.0 (the fourth industrial revolution) as the main tool for simulation of the real world in the Virtual Space. DT started in aeronautics and space industries. Reported applications from other industries are mostly from automotive industry and robotics. Let's try to analyze the place of DT in so called Process Industries (PI). By PI we mean mostly general chemical processing (bulk and fine chemicals, pharmacy), Oil & Gas processing and Power generation in classical and nuclear power stations. Everybody can imagine other similar PI sectors like minerals processing, food production, or so. Now follows very short description of basic terms.

According to Grieves [1] DT contain three main parts: a) physical objects in Real Space, b) virtual objects in Virtual Space, and c) the connections of data and information that ties the virtual and real objects together. DT concept model is shown in Figure 1.1.

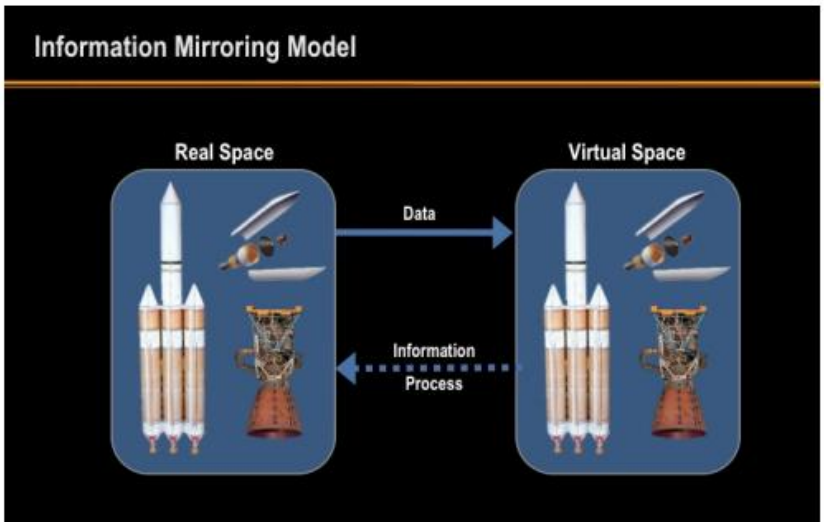


Fig. 1.1: Digital Twins concept (example from aeronautics of a rocket with its elements) [1]

Note: There are two *classes* of DT: (1) DT of *products,* like a rocket in Fig.1.1 and (2) DT of *systems*, for example producing plants. In this report is sometimes used the term *object*, which covers both classes of DT.

Generally, there can be more sophisticated definitions of DT. In [2] are presented some other definitions by major global companies shown below in Table 1. The main expectation from DT is the on-line prediction and optimization of a real object behavior.

At present it is generally accepted that the Real (Physical) Space consists of (1) Devices and of (2) Sensors, which provide the information about values of the physical variables in the Real Space. Calculations in the Virtual Space are based on mathematical models which are of two kinds: (1) models based on raw data collected

in the Real Space which can be further processed by classical statistical methods like correlation and regression or by Neural Networks and (2) models based on natural Laws, namely mass, energy and momentum conservation (balances) and thermodynamic Laws. Such models are sometimes called Equation models. The Equation models represent constraints on measured data which must be fulfilled, as will be seen later in this report. In practice we can also meet hybrid models comprising raw data models and Equation models.

Table 1.1: According to [2]

The Definition of Digital Twin According to Seven Different Companies

Company	Definition
IBM	"A Digital twin is a virtual representation of an object or system that spans its lifecycle, is updated from real-time data, and uses simulation, machine learning, and reasoning to help decision making."
Siemens	"Based on the consistent data model across all aspects of the product life cycle, some of the actual operations are accurately and veritably simulated."
General Electric	"Through the virtual models of devices and products, the actual complexities of physical entities are simulated, and insights are projected into applications."
NASA	"The application of interdisciplinary modeling and simulation across the product lifecycle."
ANSYS	"Combined outstanding simulation capabilities with powerful data analysis capabilities, it is to help enterprises gain strategic insights."
PTC	"PLM process is extended into the next design cycle to create a closed-loop product design process and help achieve predictive maintenance of the product."
FANUC	"A digital twin is the concept of creating a digital replica of the physical machines, production processes or shop floor layouts in order to generate a number of competitive advantages."

Important is that the DT concept should be applied during the whole object's Lifecycle, from the design proper, and should include also historical data, object's tests, P&I drawings, etc. In other words, DT should integrate all available information about the subject (from the cradle to the grave).

The *Fidelity* of DT [17]: High fidelity means that the model in the Virtual space is very close to reality. The 1:1 fidelity means identical results of modeling compared with the Real Space. We will see later that in PI such fidelity is not possible due to inevitable measurement errors of state variables in industrial systems.

Important is also the level of *Integration* (automatic connection) of Real and Virtual spaces. If both links between spaces are manual, we talk about the Digital Model. If the link from Real to Virtual spaces is automatic (on-line), this is called the Digital Shadow. Only the case with both links among spaces are automatic such system can be called Digital Twin [16]. See the next figure:

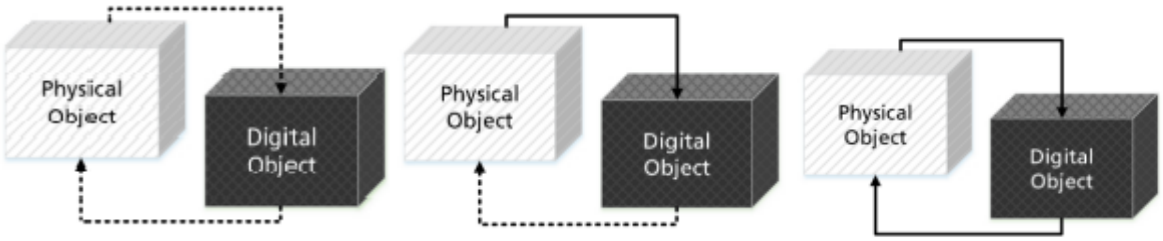


Fig.1.2: Level of integration between Real and Virtual Spaces. ------ manual data link ——— automatic data link. (a)Digital Model, (b) Digital Shadow, (c) DT

The DT notions described above are not new for the Chemical and Power Engineering and are used in some way in practice at least for the past five decades. DT resemble the concept of *Smart Process Plants* developed since the first decade of this century (see the book *Smart Process Plants. Software and Hardware Solutions for Accurate Data and Profitable Operations* [11]). The long term vision of smart plants presented there is:

- *They should "Drive" themselves, with minimal human intervention. In other words, full automation that is fault tolerant and capable of decision making.*
- *Are capable of gathering information from the environment as well as from markets, to predict problems (faults) and act proactively to prevent them.*
- *Reduce emissions and meet EH&S factors.*
- *Operate in such a way that maximum profit is achieved.*

The author of [11] admits that this vision is far from being materialized so far.

The purpose of this report is to summarize the ChemPlant's view on the DT issue in PI and to present our experience with using some DT elements in this area. We will also pinpoint some special features of data flow and processing which are typical for application of DT in Process Industries and in which PI differs from other areas of DT use.

This report contains several examples and solutions based on SW Recon. This means mainly the Data Validation and Reconciliation modeling which can be switched into the on-line Simulation based on validated process data. More info about Recon can be found in the Appendix 2 of this report.

2 MATHEMATICAL MODELING IN PROCESS INDUSTRIES

This Chapter analyzes special features of data flow and processing which are typical for application of DT in Process Industries and in which PI differs from other areas of DT use. The main concern here is the simulation of PI systems which is conditioned by the existence of a mathematical model.

2.1 History of computerized mathematical modeling in PI

The history of mathematical modelling of objects in PI with the aid of computers dates back to late fifties of the past century. The first attempts were connected with the application of so called First Laws of nature which were mass, component and energy conservation (balancing SW). In the beginning these programs were used mostly in the design stage of processes.

After mastering mass and energy balancing more complicated nonlinear models were developed for every important unit operation used in process industries (distillation, absorption, hydraulics of liquid and gas systems, heat operations, steam and gas turbines, etc.).

Later, in early sixties, there were first attempts of “almost on-line” processing of data gathered from operating plants. At this moment it was discovered that measured redundant data are not consistent – they don’t obey conservation laws. There were two main reasons for that: (1) all measured data are inevitably burdened by Random Measurement Errors, and (2) there can be present so called Gross Errors which are caused by defects in sensors. This has lead to development of methods for reduction of random errors by Data Reconciliation and also by methods of Data Validation targeted at detection and elimination of Gross Errors. The method of Data Validation and Reconciliation (DVR) is now the standard method used in the framework of the process control and data evaluation systems for plants in PI.

Probably the most DT resembling application in PI is so called *Real time Optimization (RTO)* which means the automatic control of PI processes in the closed control loop.

Note: The mention of “almost on-line processing of data gathered from operating plants” above concerns the case of one US refinery where process data in the form of punched cards were daily transferred by car from the plant to the IBM computing center ([3], published in 1961). This project was targeted at Yield Accounting which is the main KPI of any crude oil refinery.

2.2 Special features of Data Flow and Processing in PI

If someone produces cars, he works with countable components from which a car is composed. He can know exactly stock of individual components in individual places of the plant (unless some components are stolen). In the opposite case, if someone produces gasoline, kerosene and similar stuffs from the crude oil, he never knows

exactly how much oil he has purchased and how much products was produced. This stems from the fact that all physical continuous measurements (mass, volume, temperature, pressure, …) are burdened by measurement errors. In other words, **every piece of information has its uncertainty**. Neglecting this can devalue efforts invested into running a plant in PI in an optimum way. In this Section we will look at special features of data flow and processing in PI.

2.2.1 Data flow in PI

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

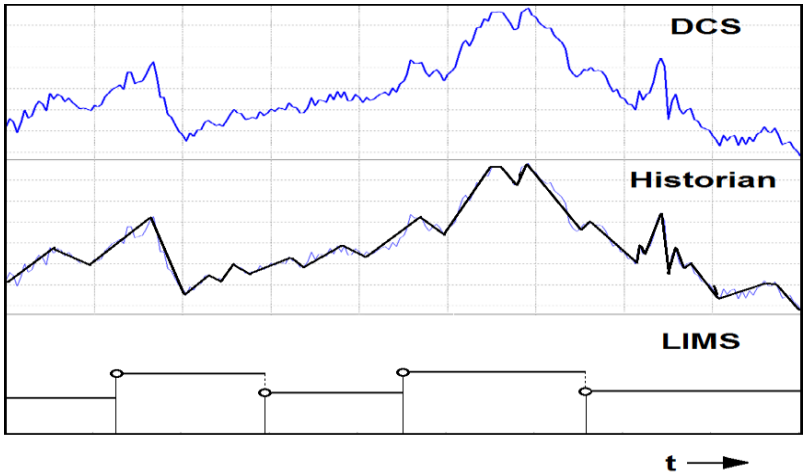


Fig. 2.1: Process data

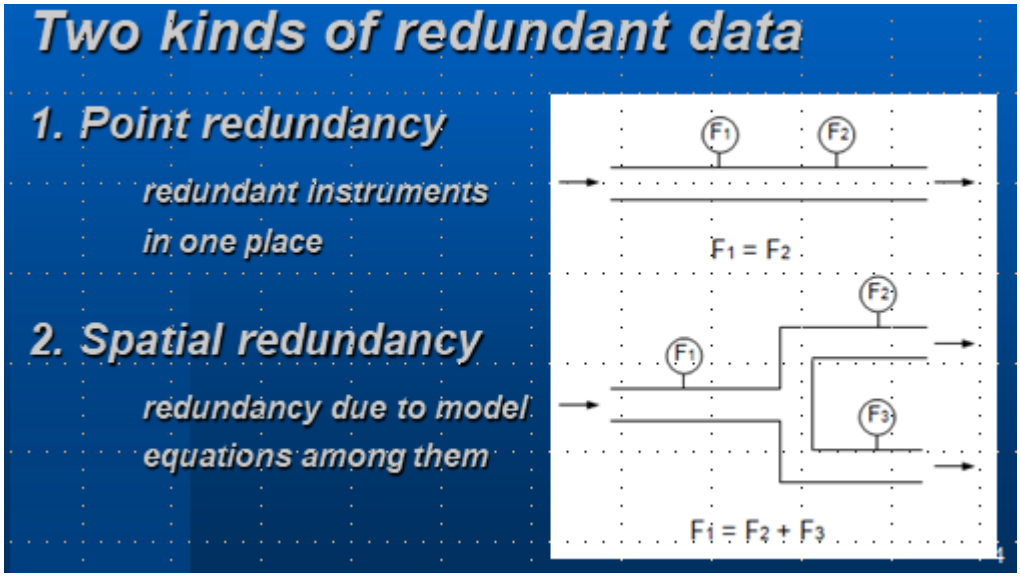
In the case of laboratory data the situation is different. The frequency of lab data nowadays is not so high. Typical is one analysis per day or shift. It is supposed that the analyzed value is constant until the new analysis is available. Important is that there is some delay between the sampling and the knowledge of results.

2.2.2 Data redundancy

For data in PI are typical two kinds of redundancy:

1. The so called *point redundancy* is caused by measurement of one process variable by two or more instruments
2. The so called *spatial redundancy* is caused by model equations valid among process variables.

See the next picture.



2.2.3 Data preprocessing

The next steps of data pre-processing can be

- compensation of instrument readings for non-standard measurement conditions if they are not done by DCS (for example compensation of orifices for temperature and pressure)
- transfer of measured values to one base (for example measuring pressure in kPa, kPag or MPa)
- calculating inventories from level measurement if this is not done by DCS.

To ease the following modeling, it is recommended to have for one variable one value we can trust. The first problem here is the existence of multiple measurements of one variable. Such redundancy is common in the last decade, especially in critical processes like are supercritical or nuclear power stations.

2.2.4 Multiple instruments

In the next figure is the example of multiple instruments measuring the same variable. This case is typical for some critical applications in the last decade.

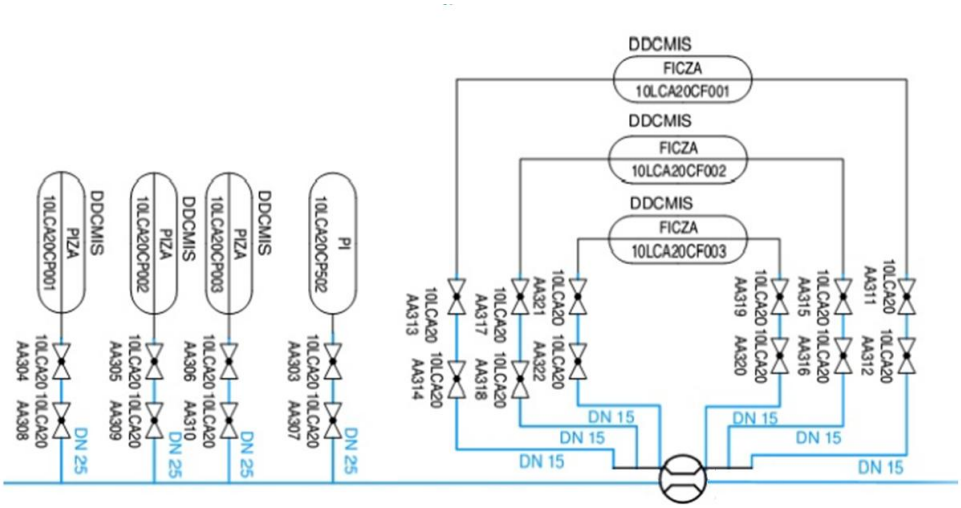


Fig. 2.2: Example of multiple instruments (3 flowmeters and 4 barometers) [14]

In [14] were analyzed 3 possibilities of treating multiple measurements (multitags) which are groups of values obtained by measuring one variable by several instruments. The main criterion was the ability to detect as small as possible gross errors. In the theory of GED this means to minimize Threshold Values of Gross errors' detection.

1. The first possibility is to incorporate all measured values in the main Recon model. This solution enlarges the whole model (number of equations), increases the calculation time and increases also the critical value which is used in Chi-square testing for a gross errors presence.

2. The second possibility is to use the basic Recon model without multitags. This means "feeding" Recon model by values selected from multitags for example by averages or medians. This is the classical functioning used in Recon so far.

3. The third possibility is to make some preprocessing of multitags' data. In general, it is possible to use Recon for reconciliation of values in individual data groups. It was shown that on the assumption that all data in one group has the same uncertainty (precision), this solution leads to the arithmetic average of values in the group. In this case is possible to find simple analytic solution including more advanced statistical methods like testing hypotheses of gross errors presence and finding threshold values for the GED power. The last possibility has also shown its superiority over the two preceding methods (minimum threshold values). The method is so simple that there is a chance to create software which will not only inform users about some issues with multitag data but also to make some automatic data correction until the problem with data is resolved by the system administrator.

In large systems we recommend to use the method 3. This means to have a special software which preprocess multitag data and creates preliminarily validated data which can be used by Recon for the next data processing.

2.3 Mathematical models in PI

2.3.1 The main mathematical model

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

$$x^+ = x + e \qquad (2\text{-}1)$$

where x^+ is the measured value

x is the true (unknown) value

e is the measurement error

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

The general mathematical model of a PI system has the form of the system of generally nonlinear implicit equations

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

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

- ***x*** is the vector of directly measured variables
- ***y*** is the vector of directly unmeasured variables
- ***c*** is the vector of precisely known constants

Such model is sufficient for mass and energy balancing of macroscopic objects in the sense of [12] and also thermodynamic modeling needed for performance evaluation in the sense of [7,12]. Models based on differential equations are not supposed in this Report. For details of modeling see the Appendix 1 of this report.

2.3.2 Data reconciliation and validation

Typically some measured data are redundant which means that due to measurement errors the data system is not consistent (data contradicts with Laws of nature). This means that with measured values the system of equations (2-2) is not zeroed regardless of values of unmeasured variables. We look for the reconciled values with which the system of equations is zeroed:

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

where x' vector of reconciled measured variables

y' vector of calculated unmeasured variables

The vector of adjustments $x' - x^+$ (where x^+ are measured values) is minimized by the Least squares method. For details see the Appendix 1. It can also happen that some unmeasured variables can't be calculated uniquely, we say that they are not observable.

The classification of variables in a general case is shown in the next figure:

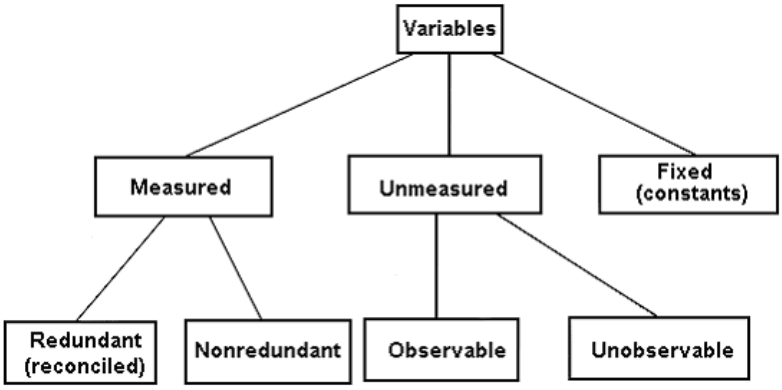


Fig. 2.3: Classification of variables

The process of data reconciliation provides also the information about data quality. If discrepancies among measured data are significant (can't be explained by small random measurement errors) we say that one or more Gross measurement errors are present. In this case Gross errors should be identified and removed from the data set.

All this process which is called DVR (Data Validation and Reconciliation) should be the integral part of data processing. For details see the Appendix 1.

2.3.3 Data postprocessing

Results of DVR described in the preceding subsection 2.3.2 contain all available information about the state of the real system. In practice it is useful to exclude some unmeasured variables from the main data set and from the model (2-2). These variables can be calculated easily after the main model (2-2) is solved from other calculated variables. Usually this is the case of variables which don't take part in the physical modeling (economy variables etc.). This makes the main model significantly smaller and calculation faster and more stable one.

2.3.4 Data processing summary

The individual steps in data processing described in this Chapter are summarized in the figure below.

Fig. 2.4: Data processing train

The whole process consisting of obtaining validated values of measured variables and also values of directly unmeasured variables and model parameters can be called the **system identification** (the term commonly used in the control theory). Such system can be now used for simulation which will be treated in the next Chapter.

3 SIMULATION

The next step in using DT is the simulation. The simulation itself has in DT significant place. It should guarantee the mirroring of RS into VS. You can see the discussion about "*When is a Simulation a Digital Twin*?" in [5].

Our opinion is that the model for DVR should be almost the same as the model for the simulation. The main difference between both models is in the classification of individual variables: Some input variables will become outputs and vice versa. This problem is solved in Recon via Model Variants.

3.1 Direction of calculation

What follows now is not a scientific problem analysis but a common sense view on calculations in practice. Calculation **variants** are characterized by "direction" of calculation or, in other words, by selection of inputs and outputs of a calculation. In this sense variables belong to one of two groups:

- Known variables, which can serve as **calculation inputs**. In Recon they belong to types M (Measured) or F (Fixed). In practice they are measured or otherwise set (qualified estimates, literature data, etc.). Fixed variables can be viewed as Measured ones with zero uncertainty.
- Unknown variables (type N), which are supposed to be calculated when solving model equations (**calculation outputs**). They belong to two classes: (1) state (physical) variables like flowrates, temperatures, pressures etc., and (2) model parameters which can be physical (heat transfer coefficients, turbine efficiencies, etc.) or economical like KPIs, efficiencies, etc.

Examples of calculation variants are shown in the next figure. Here it is supposed that process inputs (raw materials, fuel, etc.) enter in a flowsheet from left and products step out on the right.

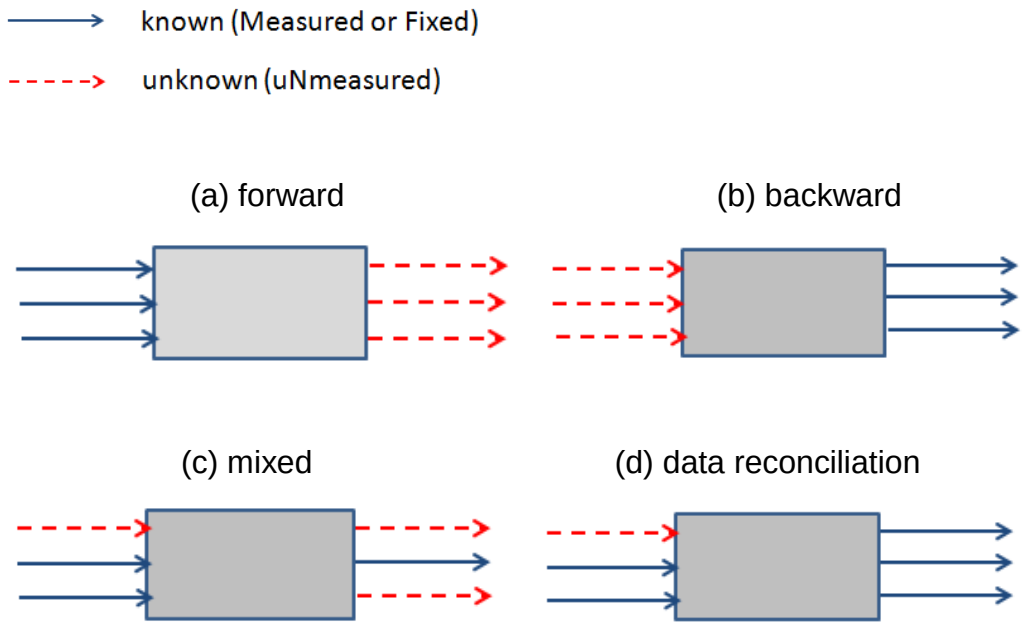


Fig. 3.1: Calculation direction

Sometimes it is required to calculate production from feeds (forward calculation), in another situation we need to calculate inputs for required production (backward calculation) or to start calculation with some combination of inputs and outputs in a general case. In the first 3 cases it is supposed that we just solve a system of equations for a given selection of unknown variables (direct calculation). In this case the number of unknowns equals the number of (independent) equations.

$$N_{eq} = N_{nv} \qquad (2\text{-}2)$$

where N_{eq} = No of equations

N_{uv} = No of unknown variables.

The “direct calculation” variant is frequently used for “simulation” calculations. This case is sometimes called the **Nonredundant** or **Just determined model.**

3.2 Simulation variant of a Base Case model

The data processing program Recon was originally designed for calculation of mass and energy balances of processes in the Chemical Industry. Later were added possibilities of thermodynamic modeling needed for power plants. The Base Case model is therefore the model which process measured data obtained from industrial plants. Typically data are redundant and the result contains validated process data plus calculated unmeasured process variables and model parameters (DVR model).

It could be possible to take such model and to create the simulation model by reversing the direction of calculation described in the previous section. But in this

way there stands up the problem of further development of 2 separate models, for example adding new streams, etc. To resolve this quite significant issue, in Recon exists the possibility to create so called model Variants. Variant of the model differs from the Base Case model in several things which are stored in the special database. The long term experience has shown that the main differences needed for using variants are:

1. Information about the direction of calculation, namely what are inputs and what are outputs of calculations
2. Using of user defined equations (active/inactive).

3.3 Process data driven simulation in Recon

The term "**Process data driven simulation**" means the two step calculation:

1. In the first step the real process data are imported from the process and are reconciled and validated (DVR). Model parameters are identified.
2. The second step is the simulation: Model parameters are used as inputs and results are some selected process variables.

Important is that the Simulation Variant model must be "just solvable", there must be no redundancy.

Below will be described the SW solution proper. More details can be found in [15].

3.3.1 Simulation variant

The simulation variant can be configured in the Recon via the special panel. The simulation variant can use 3 kinds of inputs from the Base Case model:

Tab. 3.1: Values used in individual Variants

	Measured value	**Non-measured value**
Kind A	Measured	Guess
Kind B	Measured	Calculated
Kind C	Reconciled	Calculated

For the simulation purposes the Kind C is used.

After that the user can enter the Variant and to edit the task. He can test his actions by calculating the task. This stage of using the Variant is mostly used for the simulation model development and tuning. There can be usually two ways of solution:

- from the DVR model which is usually redundant to put some inputs (measured variables) among unmeasured ones, until the zero redundancy is achieved. Then interchange the inputs for outputs.
- Make the interchange of inputs for outputs and then remove redundancy.

The selection of the methods described above depends upon the user.

3.3.2 What if? Queries

After the simulation model is tuned it is possible to use it. Recall that the use consists of two steps: (1) importing real plant data and making the DVR step, and (2) the simulation step proper.

The natural solution is the Recon module for so called *What if? Queries (WiQ).* WiQ is a simple Recon module which can be used by plant operators to simulate the plant behavior in some vicinity of a real plant state. WiQ is used in the following way:

- Process data are imported automatically on operator's request. The time can be either the actual time or some time available in the process data historian.
- Then the operator enters a change of one or more values of process variables and runs the calculation. WiQ calculates automatically the Base Case model with the original values of process variables in the DVR mode. Parameters of the model are thus identified. Then the calculation is automatically repeated in the Simulation mode with changed values of selected variables.
- The operator can see results of selected outputs of calculation and differences between the original plant state and the state after perturbation introduced by the operator.

To make the use of WiQ easier, there is the possibility of WiQ configuration, to select lists of the most frequent inputs and outputs of the whole solution.

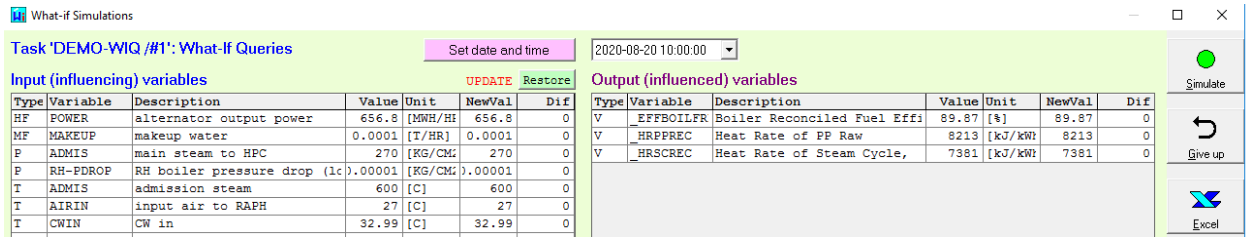


Fig. 3.2: The main panel for WiQ

Here the user can select the date and time (the last available data is the default). Then the user can edit the *UPDATE* (*NewVal*) column in the left table. It is possible to change one or more values. The extent of the change can be limited by the Recon Administrator. After that it is possible to press the *Start* button.

WiQ program reads validated data from the MS SQL server, identifies model parameters and simulates the plant with new values of input variables. Results can be seen in the right table (*NewVal* and *Dif* columns).

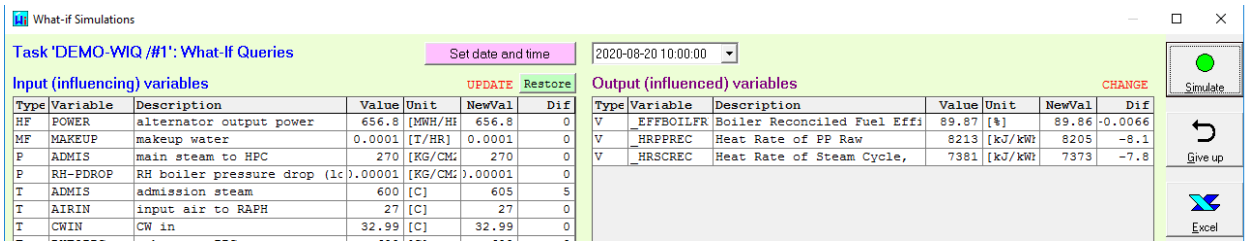
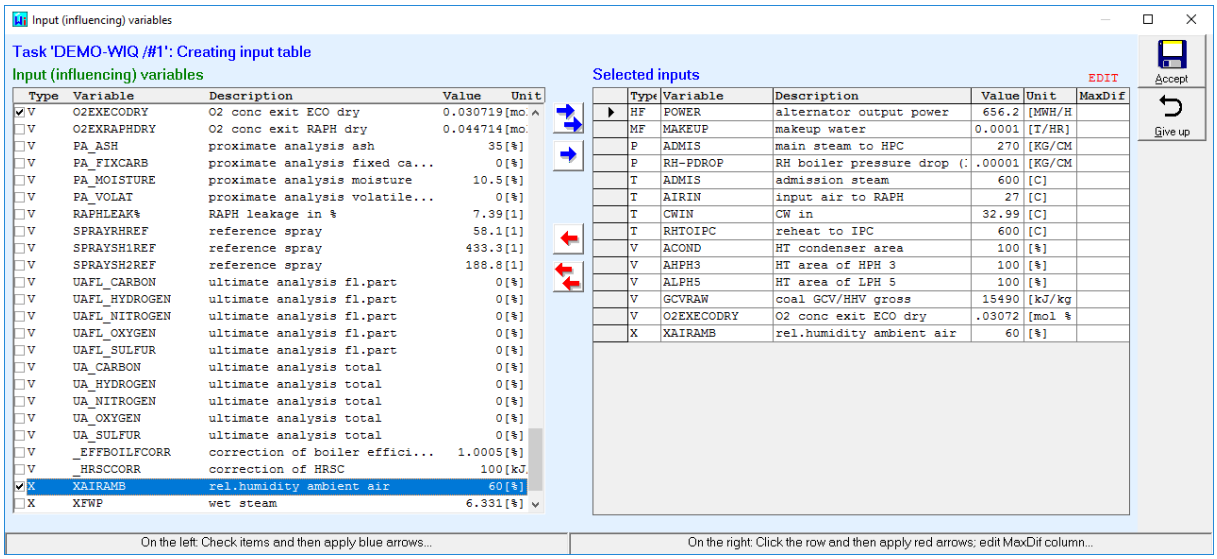


Fig. 3.3: The main panel for WiQ after running WiQ

The calculation can last several seconds. In this simple example the temperature of the admission steam was increased from 600 to 605 ^{o}C. In the right end table can be seen a negligible decrease of boiler's efficiency but a significant decrease in Heat Rates. It is also possible to export result to MS Excel.



Obr. 3.4: Example of configuration – creation of the list of inputs

3.3.3 Automatic simulation

Simulation variants of a model can be run also automatically. Also in such cases calculations are done in two steps – DVR and Simulation. Such solution is useful for example for preparing data for dashboards. In this way can be obtained information needed for special tasks, like the prediction of influence of different raw materials, fuel quality, etc.

3.3.4 Parametric sensitivities

Important is that the integral part of every Recon's calculations are so called *Parametric Sensitivities*. They are based on the linearization of the simulation model and can be used for the fast prediction of influence of individual input variables on outputs. Parametric sensitivities can be exported as the standard variables to other applications and databases.

3.3.5 Frequency of automatic calculations

One of prerequisites of DT is the on-line mirroring from the Real Space to the Virtual Space. Besides some delay caused by data processing there is another problem. Balancing of dynamic processes must evaluate changes of inventories (stock) in process equipment. Due to the process noise some time is required for averaging measured data. In other words, to set up balance in process industries requires some

time. For typical processes in PI the shortest time of completing the information about the plant is in the order of several minutes.

3.4 Improving fidelity of DT

High fidelity [17] means that the model in the Virtual space is very close to reality. The 1:1 fidelity means identical results of modeling compared with the Real Space. In Chapter 2 it was shown that in PI this is not possible due to measurement errors. The other issue is the limited accuracy of mathematical models of process equipment.

Theory of typical unit operations used in PI (heat exchangers, steam turbines, etc.) described in books is not directly applicable to real equipment provided by equipment vendors. Even if we have some information from vendors (drawings, charts), the modeling based on such information is not errorless. The great help can be statistical analysis of plant's historical data. In this way the quality of models can be improved

Example: Steam condenser of the power plant

The functioning of the steam condenser influences the power plant efficiency significantly. In the basic simulation solution described in Section 3.2 is supposed that the heat transfer coefficient (HTC) is constant. In such solution the validity of the model is good only in the small vicinity of the operational state of the power plant.

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

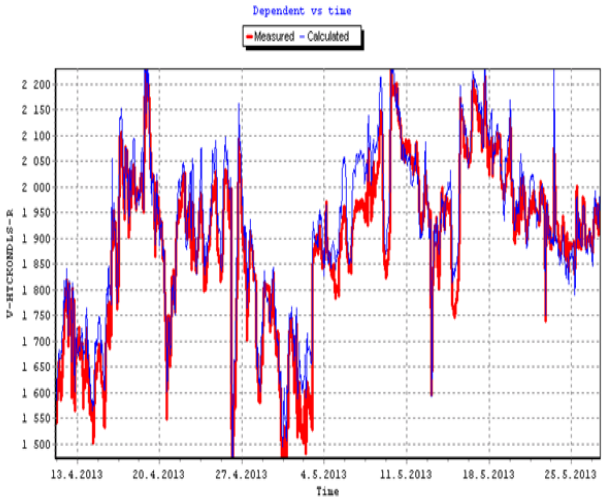


Fig. 3.5: Prediction of HTC in time (one year data)

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

4 TYPICAL TASKS WHERE DT CAN BE USEFUL

On the basis of extensive literature research authors in [4] identified 35 benefits which can result from DT application. Below we have selected those 17 which are relevant to PI:

1. Simulation of complex processes
2. Providing better knowledge about the process
3. Provide visibility in manufacturing
4. Enables real-time monitoring
5. Improve product quality
6. Provide new ways to save costs
7. Improves maintenance
8. Faults identification
9. Better data collection
10. Determine plant performance level
11. Design cost reduction
12. Detect possible problems
13. Enable system for prediction
14. Simulate reality
15. Reduce errors during manufacture
16. Monitor equipment and systems state
17. Minimize human decision making.

It should be noted that all items above are well known and acknowledged in PI. They have been addressed in thousands of projects, papers, etc.

5 EXAMPLES

5.1 The fossil fuel power plant

The electricity generation in fossil and nuclear power plants can be a good example of using DT (see also for example literature [19,20]). To get some feeling about simulation in the Power sector next follows the (a little bit simplified) example of the industrial model. It is the coal fired supercritical 660 MW power plant. The model consists of:

- furnace
- steam generator consisting of 7 heat exchanging zones
- air preheater
- deaerator
- condenser
- 5 condensate heaters
- 5 feed water heaters
- 3 turbines consisting of 9 turbine segments

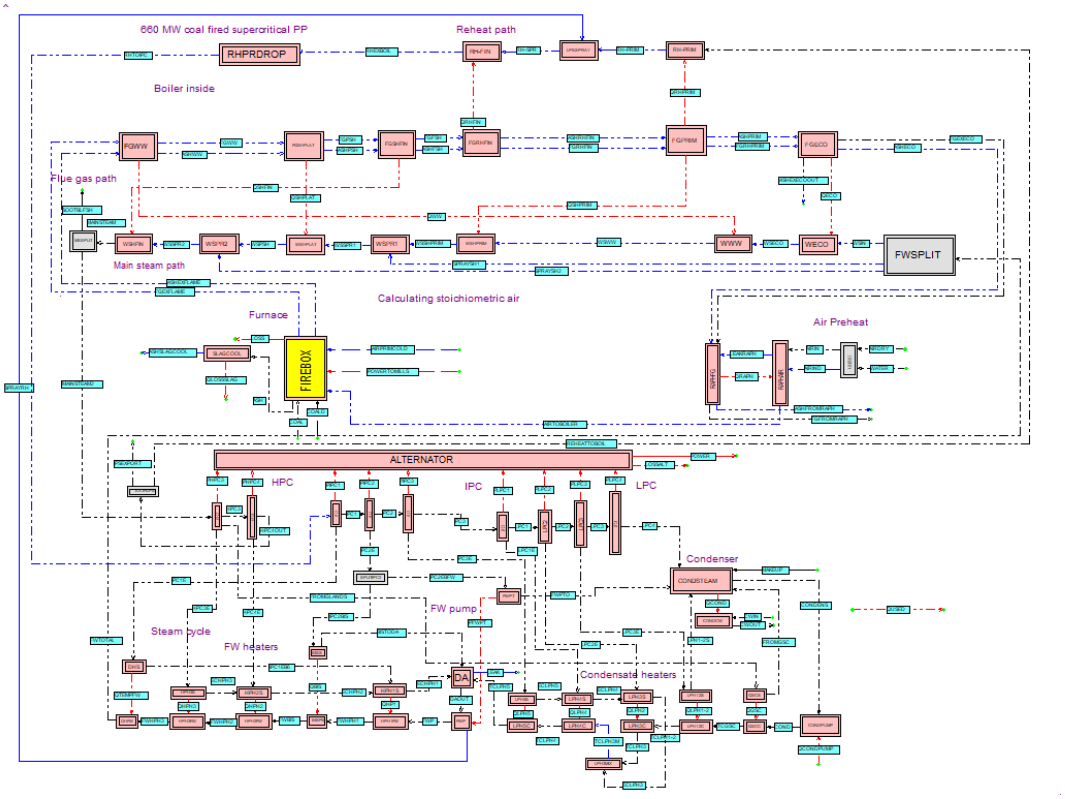


Fig. 5.1: The power plant flowsheet

Main parameters of the model:

Number of streams	133
Number of temperatures	71

```
Number of pressures                          39
Number of measured variables                 89
Number of non-measured variables            209
Number of equations                         209
```

The DVR evaluation of the model lasts several seconds on a standard server and the same is approximately needed for the simulation. This means that the time of mirroring of RS into VS is in the order of fractions of one minute.

The difference between the DVR model and the Simulation model is in inputs and outputs of the calculation. 21of model parameters (turbine segments efficiencies, heat transfer coefficients and some other parameters) were exchanged for other process variables.

The main purpose of all these calculations is in:

- finding the full information about the state of the power plant
- comparison of plant's state with the optimum condition
- identification of opportunities for process improvement
- localization of possible equipment degradation.

In practice, the difference between the actual plant operation and the optimum state is caused by three groups of factors:

1. different external conditions (ambient temperature and air humidity, cooling water temperature, quality of the fuel, etc.)
2. degradation of the plant's equipment (heat exchange areas fouling, etc.)
3. improper control of the plant by operators.

It is important to separate the influence of these three groups on the overall economy of the plant and to quantify these influences in details. In the next figure is an example of the dashboard used for analysis of the real power plant operation.

The main KPI of a power plant is the Heat Rate (the fuel energy needed for production of the unit of electric energy produced). Below is the example of the dashboard helping operators to analyze the actual situation in the power plant:

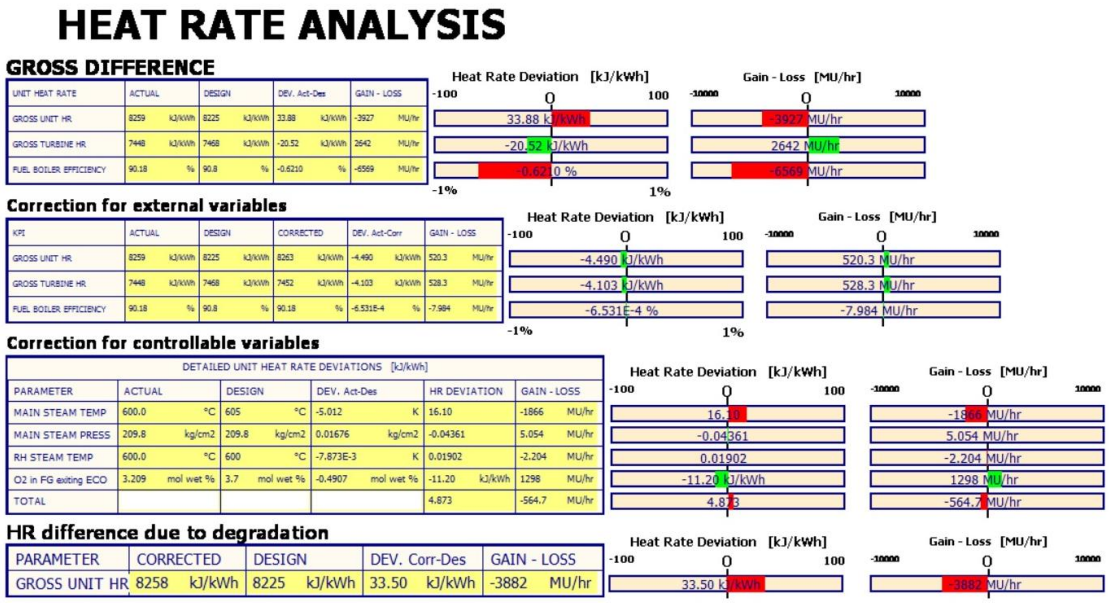


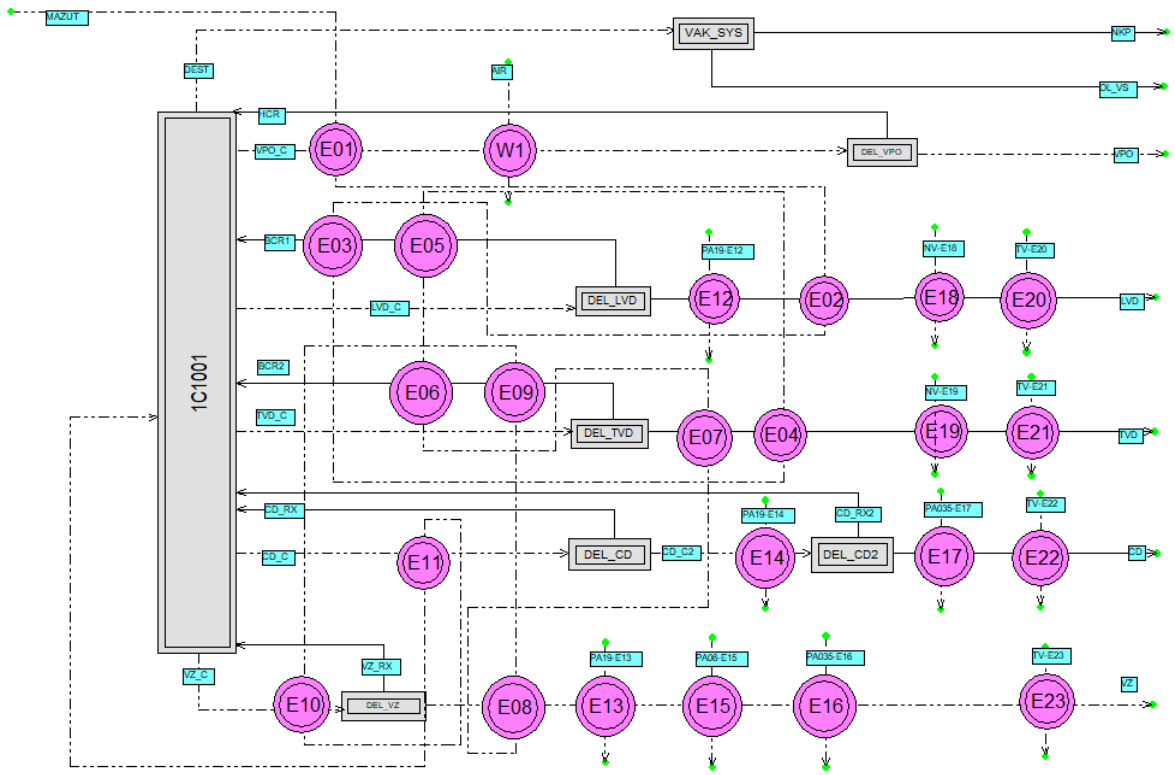
Fig. 5.2: The Heat Rate analysis dashboard.

Operators can see the main reasons why the power plant's Heat Rate is different from the expected value. They can also see the individual control variables which are out of the prescribed values.

There are also other dashboards analyzing the individual parts of the power plant (diagnostics of individual power plant's parts).

5.2 Vacuum distillation of the heavy fuel oil

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



This model has

58 measured variables (flowrates, pressures and temperatures)
32 model equations
24 heat exchangers (coolers, heaters, steam generators)

The DVR evaluation of the model lasts only 0.1 seconds on a standard server and the same is approximately needed for the simulation. This means that the time of mirroring of RS into VS is in the order of fractions of one second.

The main KPI of this system is the heat removed from the system by the cooling water (the lost heat).

6 DISCUSSION AND CONCLUSIONS

As it was mentioned in the Chapter 1, the concept of *Digital Twins* is the logical continuation of the concept of *Smart Process Plants* coined in the first decade of the present century. While the origin of DT is connected with aeronautical and space industries, Smart Process Plants are linked with Process Industries (mostly plants producing chemicals and energy). The way of this idea in the Process Industries started in the mid of the past century with the advent of computers applied in plants designed for huge industrial production.

There were tremendous efforts targeted at solution of problems connected with data archiving and further with the simulation of industrial processes. There were also strong efforts for getting a good reasonable data (removing the influence of random and gross errors). There exist thousand of papers and many books solving the problem of data processing in Process Industries.

Main ideas connected with Digital Twins are:

1. Mirroring of objects from the Real Space into the Virtual Space. This requirement is identical with the notion of the Simulation
2. The simulation should be of high fidelity
3. DT concept should be applied during the whole object's Lifecycle, from the design proper, and should include also historical data, object's tests, P&I drawings, etc.
4. Important is also the level of *Integration* (automatic connection) of Real and Virtual Spaces. Only a system in which both links between spaces are automatic can be called DT.

The concept of DT in PI starts with the problem of modeling in PI. Models are based on physical laws which are well known in the process industry community. The second issue is the problem of data consistency which is connected with measurement errors. This issue was solved in last decades with sufficient results. Details of this issue are described in Section 2.1

Simulation (mirroring of RS in VS) of a real process is the backbone of any DT. In Chapter 2 was shown that exact mirroring (1:1) is not possible due to the presence of measurement errors of continuous variable which are the inevitable part of data provided by sensors. The way of the whole train of data processing was described in the Subsection 2.3.4.

In Chapter 3 was presented so called *Process Data Driven Simulation.* This means the two step simulation concept. In the first step the process system is identified (the model parameters are calculated on the basis of measured data). In the next step these model parameters are used for the system simulation.

Important is the simulation model *Fidelity*. Models of DT can't be based on theory only. Models published in chemical and power engineering literature are not reliable enough to describe behavior of the real plant equipment. The great help in this area can be the statistical processing of historical data gathered during the long term running of industrial processes (as was shown in the Section 3.4).

DT concept should be applied also during the whole object's Lifecycle, from the design proper, and should include also historical data, object's tests, P&I drawings, etc. This is very strong requirement for PI. SW used for the design of plants in PI is quite different from SW for the simulation. As an example let's consider a classical power station. There are usually at least three different vendors (steam generator, steam cycle and cooling towers). The combining of the design SW from the individual vendors does not seem to be realistic in practice. On the other hand side, the use of historical process data including plant tests can be very useful.

The level of integration of the Real and Virtual Spaces in PI is also disputable. For modern plants it is quite common the automatic link from the RS to VS. Previously this concept was in PI called *On-line modeling.* The time delay in this direction is from the practical point of view negligible. In this case we can talk about the Digital Shadow. In the opposite direction this on-line link is applied in cases of so-called RTO (*Real Time Optimization*) systems which mean the on-line control in the closed control loop. Such systems are used only in critical operations which warrant reasonable payback of such, usually quite expensive, solutions. There is also the possibility of the semi-automatic link in direction from VS to RS. Such link can be realized via so called *Active Dashboards* where the process operators can see results of on-line simulation and can act from the dashboard directly.

In the literature survey [16] was found that in manufacturing applications of DT there were 35 % of DS, DM 28 % and DT 18 % (the remaining cases had the indeterminate level of integration). On the opposite, in the newer literature research [6] was 59 % of papers categorized as DT, DM 22 % and DS 19 %.

The authors [16] also state that: *The development of the DT is still at its infancy as literature mainly consists of concept papers without concrete case-studies. However, some applied case-studies already exist – especially at the lower levels of integration (DM and DS).*

We can conclude this report by finding that the level of implementation of DT in PI is (in comparison with other industries) not so bad. It is very similar to the concept of *Smart process plants* coined in the Process Industries about 20 years ago.

LITERATURE CITED

[1] A Whitepaper by Dr. Michael Grieves: *Digital Twin: Manufacturing Excellence through Virtual Factory Replication.* 2014. See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/275211047

[2] Shahid Aziz et all: *Digital Twins in Smart Manufacturing*. Chapter 4 in *Handbook of Manufacturing Systems and Design. An Industry 4.0 Perspective*, CRC Press 2024.

[3] Kuehn,D.R., Davidson, H.: *Computer Control II.* Chem.Eng.Progr. **57**(6),44 (1961)

[4] Javaid,M., Haleem,A.,Suman,R.: *Digital Twin applications toward Industry 4.0. A Review*. Cognitive Robotics, **3** (2023) 71 – 92

[5] Wooley,A.,Silva.D.F.,Bitencourt,J.: *When is a Simulation a Digital Twin? A Systematic Litewrature Review.* NAMRC 51,2023

[6] Juárez-Juárez, MG.; Botti, V.; Giret Boggino, AS. (2021). Digital Twins: Review and Challenges. Journal of Computing and Information Science in Engineering. 21(3):1-23. https://doi.org/10.1115/1.4050244

[7] Madron, F., 1992, Process Plant Performance: Measurement and Data Processing for Optimization and Retrofits, Ellis Horwood, New York.

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

[9] Romagnoli, J. A., and Sanchez, M. C., 2000, Data Processing and Reconciliation for Chemical Process Operations, Academic Press, London

[10] Bagajewicz,M., *Process Plant Instrumentation. Design and Upgrade,* Technomic Publishing Company,2000

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

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

[13] Gay, R. R., Palmer, C. A., and Erbes, M. R., 2004, Power Plant Performance Monitoring, R-Squared, Woodland, CA.

[14] Report CPT 445-21: *Processing Multiple Instruments Data.* Usti nad Labem 2021

[15] Report CPT 441-21: *Calculation Variants in Recon*. Usti nad Labem2021

[16] Kritzinger, W. et all: *Digital Twin in Manufacturing: A categorical literature review and classification.* IFAC PapersOnline 51-11 (2018) 1016-1022

[17] Kober, C.: *Digital Twin Fidelity Requirements Model For Manufacturing.* CPSL 2022, 595 – 611.

[18] V.V. Veverka, *Balancing and Data Reconciliation Minibook*, www.chemplant.cz

[19] Z. Lei, H. Zhou, W. Hu, G.-P. Liu, S. Guan, X. Feng, “Towards a Web-Based Digital Twin Thermal Power Plant”, IEEE Transactions on Industrial Informatics, vol. 18, issue 3, 2022, pp. 1716-1725. DOI: 10.1109/TII.2021.3086149

[20] Milič,S.D., Kožicic,M., Šiljkut,V.M.: *Concept of Digital Twins in the Powerstations.* CIGRE D216, Srbija

APPENDIX 1: MODELS IN PROCESS INDUSTRIES

This Appendix summarizes theory of data processing including some more advanced methods like measurement errors propagation and the Power of testing hypotheses about gross errors. The theory is taken over from the book [7].

A1.1 Models

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

$$x^+ = x + e \qquad \text{(A1-1)}$$

where x^+ is the measured value

x is the true (unknown) value

e is the measurement error

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

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

Let us start from the mathematical model

$$\boldsymbol{F(x,y,c) = 0} \qquad \text{(A1-2)}$$

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

- ***x*** is the vector of directly measured variables
- ***y*** is the vector of directly unmeasured variables
- ***c*** is the vector of precisely known constants

The starting point for the following solution is the solvability analysis of a set of linear equations in variables representing measured and unmeasured variables. The important simplification of the nonlinear model (A1-2) is so-called ***General Linear model***

$$\boldsymbol{A'x + By + a = 0} \qquad \text{(A1-3)}$$

where

x is vector of measured variables
y vector of unmeasured variables
a vector of constants

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

The General Linear model can be further simplified by elimination [2] of unmeasured variables to the form (note that matrices ***A*** and ***A'*** are different):

$$\boldsymbol{Ax} + \boldsymbol{a} = \boldsymbol{0} \qquad \text{(A1-4)}$$

A1.2 Data reconciliation

Eq. (A1-2) holds for the true (unknown) values of the variables. If we replace them by the measured values $\boldsymbol{x}^+$, the equations need not (and most likely will not) be exactly satisfied:

$$\boldsymbol{F(x^+,y,c) \neq 0} \qquad \text{(A1-5)}$$

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

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

$$x_i' = x_i^+ + v_i \quad , \qquad \text{(A1-6)}$$

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

The adjustments must satisfy two fundamental conditions:

1) The reconciled values obey Eq. (A1-2) – we say that they are consistent with the model

$$\boldsymbol{F(x',y',c) = 0} \qquad \text{(A1-7)}$$

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

minimize $\sum (v_i/\sigma_i)^2 = \sum ((x_i' - x_i^+)/\sigma_i)^2$. (A1-8)

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

The inverse values of the standard deviations σ – 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 (A1-8) is used in the case of uncorrelated (statistically independent) errors. In the case of correlated errors a more general criterion is minimized:

minimize $\boldsymbol{v^T F^{-1} v}$ (A1-9)

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

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

The mathematics of the solution itself was in the last decades many times described in the literature (e.g. [7-11]) and will not be mentioned in the sequel. See also the *Balancing and Data Reconciliation Minibook* [18], 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.

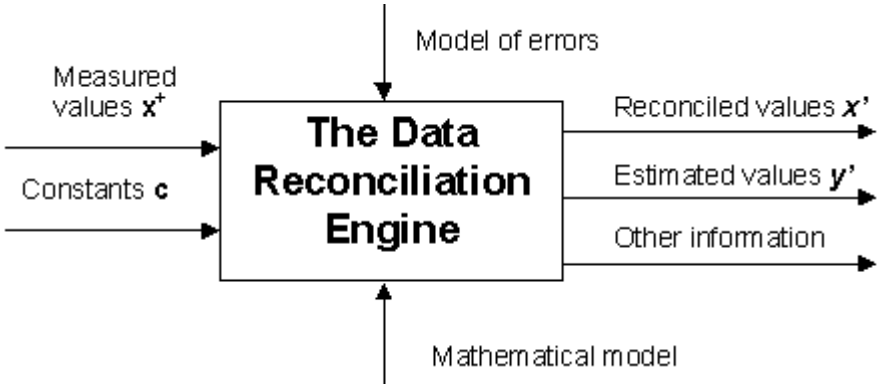


Fig. A1-1: The Data Reconciliation Engine

The model (A1-4) is used for DR proper. In the first step the adjustments $\boldsymbol{v}$ are calculated according to the equation

$$\boldsymbol{v = - F_X A^T (A F_x A^T)^{-1} (a + A x^+)} \quad \text{(A1-10)}$$

Reconciled values ***x'*** are then calculated from the equation

$$\boldsymbol{x' = x^{+} + v} \qquad \text{(A1-11)}$$

by substitution from Eq. (A1-10).

A1.3 Statistical properties of results

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

$$\boldsymbol{F_v = F_x A^T (A F_x A^T)^{-1} A F_x} \qquad \text{(A1-12)}$$

The Quadratic form of adjustments (A1-8) or (A1-9) is the random variable with $\chi^2_{(1-\alpha)}(\nu)$ distribution with ν degrees of freedom. Values of $\chi^2_{(1-\alpha)}(\nu)$ for probability $(1-\alpha)$ are tabulated in statistical tables.

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

$$F = F_v + F_{x'} \qquad \text{(A1-13)}$$

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

$$\sigma_i^2 = \sigma_{vi}^2 + \sigma_{x'i}^2 \qquad \text{(A1-14)}$$

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

$$\langle x_i' - 1.96\ \sigma_{x'i}^2\ ;\ x_i' + 1.96\ \sigma_{x'i}^2\ \rangle \qquad \text{(A1-15)}$$

cover the (unknown) true values of individual variables.

Reconciled data are more precise in the statistical sense, if compared with the measured ones. The enhanced precision can be quantified with the aid of the

standard deviation of the reconciled value, which is always smaller than the standard deviation of the measurement error.

$$\sigma_{x'} < \sigma \qquad \text{(A1-16)}$$

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

$$a = 1 - \sigma_{x'} / \sigma \qquad \text{(A1-17)}$$

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

A1.4 Detection of gross errors

The most frequently used method for Gross Errors Detection (GED) is the test based on the value the least squares function (A1-8) or (A1-9). The Quadratic form of adjustments (A1-8) or (A1-9) is the random variable with $\chi^2_{(1-\alpha)}(\nu)$ distribution with ν degrees of freedom. Values of $\chi^2_{(1-\alpha)}(\nu)$ for probability $(1-\alpha)$ are tabulated in statistical tables.

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

$$Q_{min} = \boldsymbol{v^T F^{-1} v}, \qquad \text{(A1-18)}$$

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

$$Q_{min} < \chi^2_{(1-\alpha)}(\nu) = Q_{minCrit} \qquad \text{(A1-19)}$$

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

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

Probability level $(1-\alpha)$ is usually supposed to be 0.95 (95 %)). All this holds on assumptions that only random errors with the Normal distribution are present.

A1.5 Power of the GED test

Let's recall the Eq. (A1-1) which defines a random measurement error *e*. This model will be now enhanced by introducing the constant gross error (bias) *d*

$$x^+ = x + d + e \qquad \text{(A1-20)}$$

where x^+ is the measured value

- x is the true (unknown) value
- e is the random measurement error
- d is the constant gross error (bias).

It is clear that the presence of the bias will change values of adjustments and reconciled values. We feel that increasing *d* will also increase Q_{min}. Further on we suppose that the reader is a little bit familiar with **testing of statistical hypotheses** which will be further used.

The null hypothesis H_0 is: No bias is present. This means

$$d = 0 \qquad \text{(A1-21)}$$

The alternative hypothesis H_1 is

$$d \neq 0 \qquad \text{(A1-22)}$$

If it holds that

$$Q_{min} > \chi^2_{(1-\alpha)}(\nu) \quad , \qquad \text{(A1-23)}$$

The hypothesis H_0 **is rejected**. In other words, the presence of a gross errors is confirmed (a gross error is detected).

It is well known that during testing hypotheses we can commit errors of two kinds:

Error of the Ist kind: A gross error has been detected though actually this is not present. The probability of this case equals the probability α of the test (0.05 in our case)

Error of the IInd kind: An existing gross error has not been detected. This probability depends on the magnitude of the gross error.

As every statistical test, also the $\chi 2$ test has its power characteristics shown in Fig. A1-2.

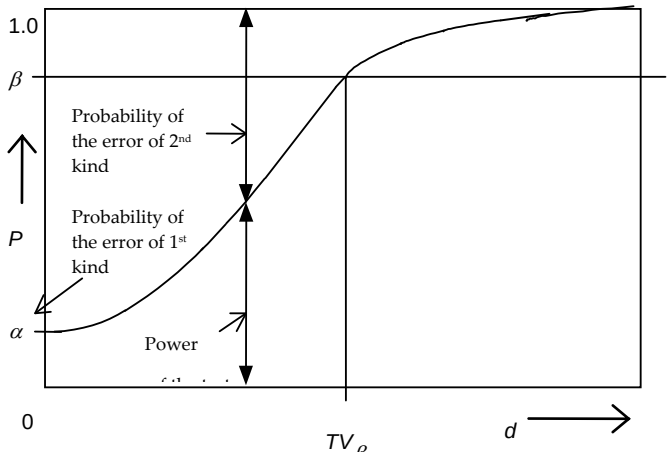


FIG A1-2: POWER CHARACTERISTIC OF THE χ^2 TEST

On the *x*-axis there is the magnitude of a gross error *d*. On the *y*-axis is the probability *P* that the gross error will be detected. The power characteristic for a measured variable equals the confidence level α in the absence of the gross error (d=0) and for redundant variables approaches 1 for high values of the gross error ($d\rightarrow\infty$). TV_β is the value of a gross error which will be detected with probability β (β = 0.90 further in this document). TV_β is characteristic for every measured variable. The lower is TV_β, the better. It is clear that gross errors can be detected only for redundant measured variables.

Threshold values can be calculated from equation

$$q_i = \delta_\beta(\nu,\alpha)/[a_i(\text{A1-}a_i)]^{1/2} \qquad \text{(A1-24)}$$

where q_i is a dimensionless threshold value TV_i/σ, which means

$$q_i = TV_i/\sigma_i \quad \text{or} \quad TV_i = q_i\,\sigma_i \qquad \text{(A1-25)}$$

and $\delta_\beta(\nu,\alpha)$ is a constant characteristic for the confidence level of the *chi*-square test α, number of degrees of freedom ν and the probability that a gross error will be detected β.

Equation (A1-24) is slightly re-arranged equation (4.143) from literature [2]. Values of $\delta_\beta(\nu,\alpha)$ are not available in standard statistical tables. Details about calculating threshold values and constants δ (for α=0.05, β=0.9 and ν = 1,2, … ,20) can be found in literature [2]. In this report will be used the new equation (A1-26) for the β = 0.90. This equation approximates δ (for α=0.05) in the range of ν = 1,2, … ,400.

$$\delta_{0.90}(\nu,0.95) = 3.23176 + 0.456899\ln(\nu) + 0.014449\ln(\nu)^2 + 0.014124\ln(\nu)^3 \qquad \text{(A1-26)}$$

It is worth mentioning that threshold values are simple functions of adjustabilities defined by Eq. (A1-17), see also the next graphical presentation of Eq. (A1-3).

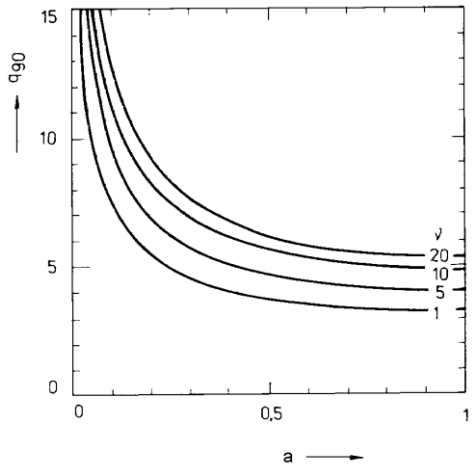


FIG.A1-3. EXAMPLE OF THE DIMENSIONLESS THRESHOLD VALUE q AS A FUNCTION OF ADJUSTABILITY a AND ν, $\alpha = 0.05$ AND $\beta = 0.90$)

Some simple conclusions can be deduced from this graph:

- the higher is the adjustability, the higher is the probability to detect a gross error (lower value of the threshold value)
- the higher is ν, the lower is the probability to detect a single gross error (higher value of the threshold value)
- for adjustabilities less than 0.1 the chance for detecting gross errors diminishes steeply.
- the minimum possible value of q_{90} for $\nu = 1$ and $a = 1$ is 3.24. This means for example for a temperature with tolerance 1 K, the sigma is 0.51 and the gross error must be at least 0.51 * 3.24 = 1.65 K to be detected with probability 90 %. But in practice the conditions (ν and a) are usually not so favorable. You can see that the gross errors detection is not omnipotent. In practice we are happy when a gross error of the magnitude of 2 or 3 times greater than the measurement tolerance is detected.

APPENDIX 2: ABOUT RECON

ChemPlant Technology, s.r.o.

RECON

Mass, heat and momentum balancing software with data reconciliation

Description

RECON is a comprehensive interactive software for modeling (mass, energy and momentum balancing, thermodynamic calculations, etc.) of complex chemical and power 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 but can be used also for simulation of plant's behavior under changed conditions. RECON can be also used for classical balancing in the stage of the process design.

RECON can be used in two operating modes:

- Data Reconciliation and Validation (DRV) mode. In this case models are solved on the basis of measured or otherwise set process data (flows, temperatures, etc.) to reconcile redundant data, calculate unmeasured process variables and identify model parameters (heat transfer coefficients, efficiencies, etc.).
- Simulation mode predicting plant's behavior on the basis of models identified in the preceding DRV step.

In practice, RECON's operation can be switched between these two modes of operation for updating models and using them for predicting plant's behavior, optimization or answering "What if?" queries.

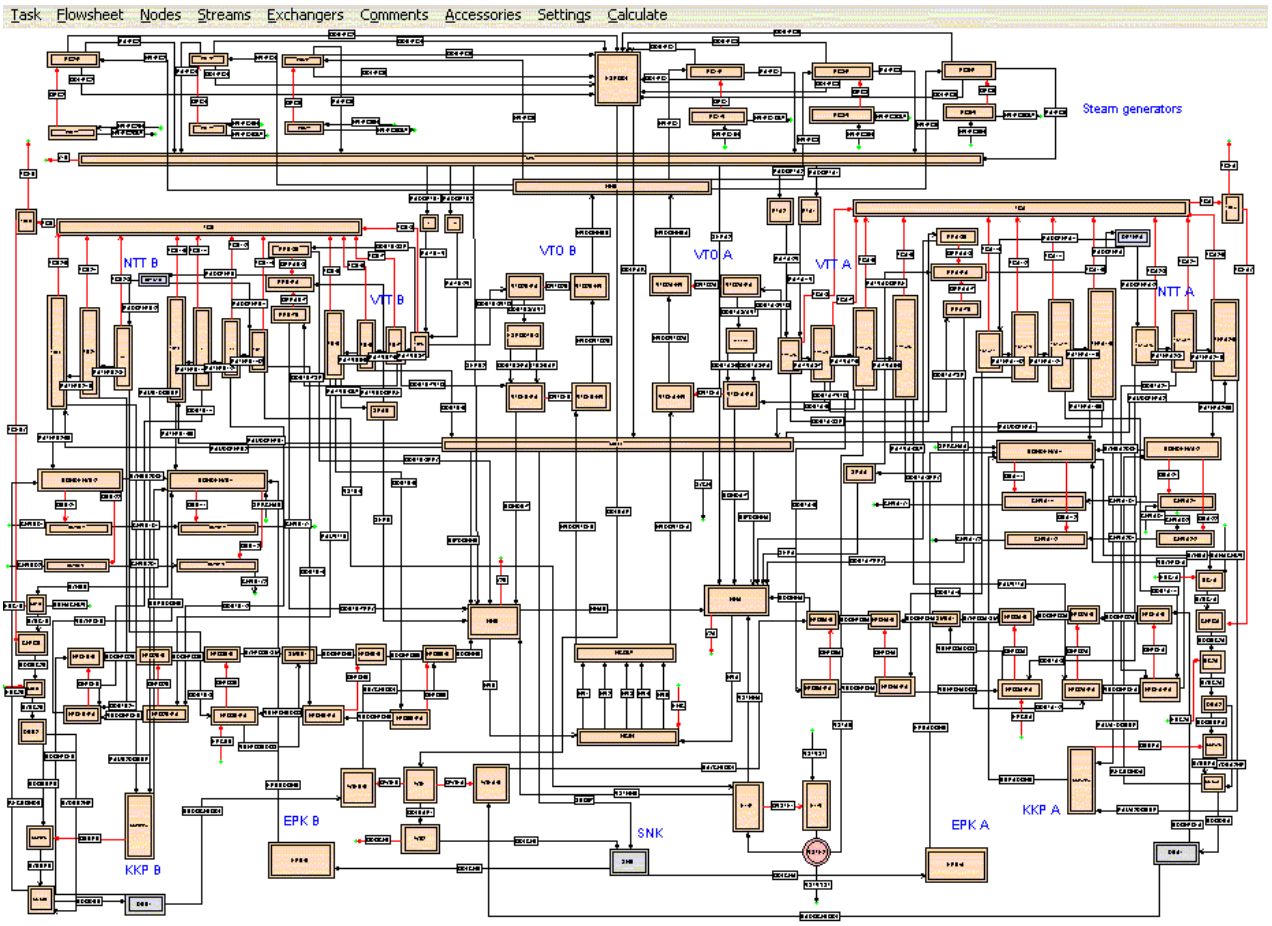
What is data reconciliation?

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). Results are enhanced by calculated unmeasured variables. As a result, a new consistent more precise data set 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, ...). Brief survey of reconciliation theory is contained in Recon's User Manual of the application.



Graphical User Interface of RECON

Main Features

RECON is a PC oriented software with user friendly facilities. Problems (tasks) are defined interactively in the graphical user interface. RECON is aimed at single or multi-component material and energy balancing of complex systems at steady or unsteady (dynamic) state, without or with chemical reactions (reactor balancing). It is also capable to perform momentum balancing based on hydraulic calculations of flow in pipeline systems. RECON reconciles measured flow rates, concentrations, temperatures and other process variables and calculates unmeasured variables. Problem (task) is commonly defined by creating a 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 by their own equations.

Monitoring of a heat transfer coefficient in RECON

Typical tasks solved by RECON

- mass and energy balances of chemical and refinery plants
- natural gas processing
- distribution systems of utilities
- detailed balancing of chemical plants (multi-component systems with chemical reactions)
- monitoring industrial heat exchange systems (crude preheat trains and similar systems)
- balancing steam systems
- monitoring and on-line thermodynamic analysis of power generation systems (turbine efficiencies, low and high pressure preheats, condenser heat transfer, etc.)
- combustion processes (industrial furnaces, coal and gas fired boilers)
- balancing water treatment in power plants (softening, demineralization ...)
- power and heat cogeneration systems including sea water desalination plants
- flow of liquids and gases in pipes and pipeline systems including their hydraulics
- detailed mass and energy balance of nuclear plants (primary and secondary circuits, advanced assessment of nuclear reactor power)

Variables

Typical process variables are: Flowrates, concentrations, temperatures, pressures, steam wetnesses, etc. The following information about 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)
- Uncertainty (Maximum errors or standard deviations) of measured variables (and their covariances, if inevitable).

Capabilities

- Data import and pre-processing
- 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 operating plants
- Reporting module based on user-defined templates in MS Excel
- Monte Carlo simulation
- The module for the Process Data driven simulation a What if? Queries
- Data mining.

RECON consist of 5 parts:

- Configuration of models in GUI
- Calculation engine which is optimized as concerns speed and memory requirements. 2 methods of solution are available: Successive Linearization and Sequential Quadratic Programming. A special feature of RECON is the existence of the calculation engine also in the form of an ActiveX object which can be called from customer's proprietary software. In this way a customer can fully manage data processing
- Recon Manager for configuring and coordinating data processing
- Database and files containing Metadata (data about models, users,)
- Database of operational data. This can be either standalone for RECON (Access, Oracle or MS SQL Server) or fully integrated with customer's databases or historians (PI, PHD, InSQL, AIM*, Oracle, MS SQL Server,) - this solution has important advantages.

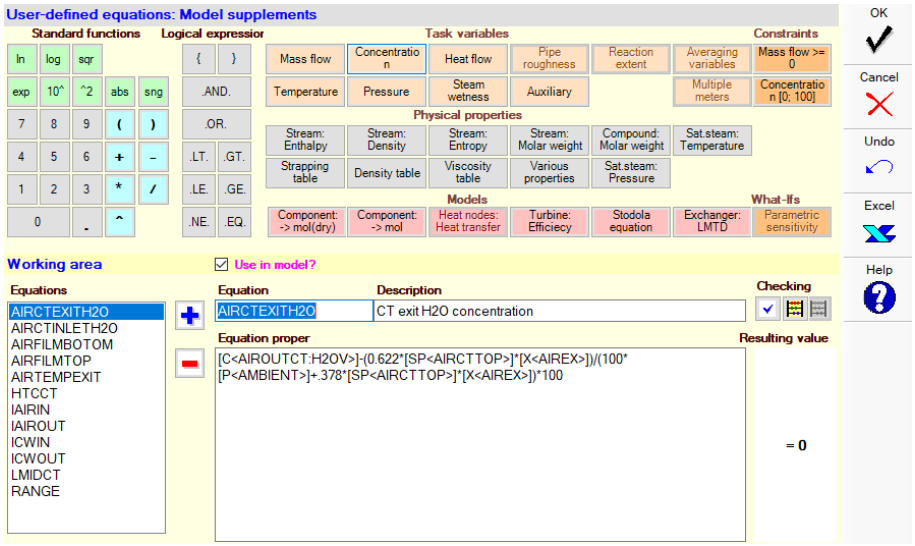
Operational data can be entered manually or imported from 1 or more other sources (process data historians, databases, Excel or text files).

Data pre-processing

Imported data can be pre-processed on the basis of several techniques (limits on input variables, limits on control variables). There is also a Basic-like editor for general data pre-processing calculations.

User defined equations

Aside of problem configuration in the Recon's GUI the user can define his own model equations. There is a Basic-like editor for writing new model equations. User can use conditional programming with logical variables. It is possible to use in equations complex thermodynamic functions, physical properties, etc. User defined equations can be used also for defining model inequalities.



Editor of User Defined Equations

Chemical reactors

Chemical reactors can be modeled by 2 ways: The standard method uses chemical reactions (stoichiometric coefficients) stored in Recon's reaction bank. Sometimes reactions in the system are not well known (e.g. burning of coal). In this case user can use the second method based on the so-called reaction invariant chemical reactor. In this case only the knowledge of atomic composition of components is required.

Gross errors treatment

There are 3 steps of Gross Errors (GE) treatment:

1. detection
2. identification
3. elimination

GE detection - finding the GE presence. The data quality is continuously monitored via the Status of data quality (SDQ). This variable should be under normal situation below 1. Values of SDQ above 1 signal the presence of some Gross Error (either an instrument or model error). SDQ is based on so-called Global chi-square test testing the value of the least square function minimized during DR.

There are several methods available for GE identification:

1. Values of normalized adjustments (NA). Suspect measurements are those with the highest NA. Such flows are marked in color on the flow-sheet.
2. Successive elimination of suspect measurements. Suspect variables from the step 1 are automatically set one by one as unmeasured. Variables with the largest decrease of SDQ are suspect.
3. Measurement credibility. Values calculated in the step 2 are compared with limits set on variables. If the calculated value is not probable or feasible, this suspect is excluded from the set of suspects
4. Nodal test. In the case of mass balance the imbalances around individual nodes or their combinations can be statistically tested. Streams around suspect nodes are suspect. This method is suitable also for finding the frequent model errors - leaks or neglected mass accumulation.
5. Covariance matrices of adjustments are available for detailed analysis of the gross error identification problem.

The steps in GE identification are semi-automatic. All this serves as a Decision Support System for the user.

GE elimination. We do not recommend and support the automatic elimination of gross errors. This function offered by some providers of DR software frequently leads to wrong decisions. Good data can be deleted and the wrong ones remain with formally good balances. This opinion is supported also by balancing theory. GE elimination should be done by the DR system administrator with further actions (instrumentation repair etc.).

Data mining

A long term use of RECON enables one to create a historical database of validated process data. Such database represents an invaluable source of information which can be utilized in many areas (creation of empirical models, optimization, etc.). Such activities are usually denoted as data mining.

The new data mining module of RECON enables one to analyze easily large data sets and to find hidden relationships among data. By a few key strokes a user can easily create multiple regression models among process variables. Attention is also paid to the statistical analysis of process data and to removal of outliers.

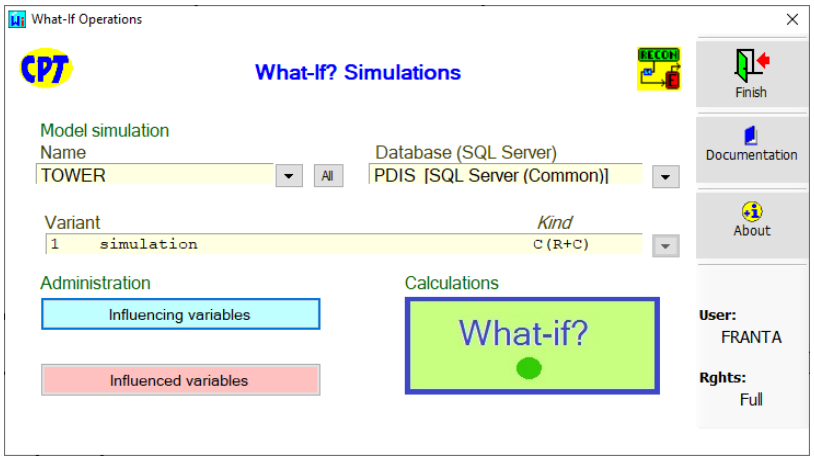
Process data driven simulation

Besides the main Recon model (the Base Case) there can be one or more model variants. The individual variants differs in inputs and outputs of the calculation This feature enables one to switch Recon between the data reconciliation and validation mode to the simulation mode. This can be done either manually or automatically.

What if? Queries

The Process data driven simulation is the basis of the What if? Queries module of Recon. In this way operators can easily see what will happen if they will change some control variable. Data from the process are imported and parameters of the

process model are identified. Then the operator enters the change of one or more process variables and runs the simulation. After that he can see in few seconds the influence of changes of inputs on outputs.



The opening screen for What if? Queries

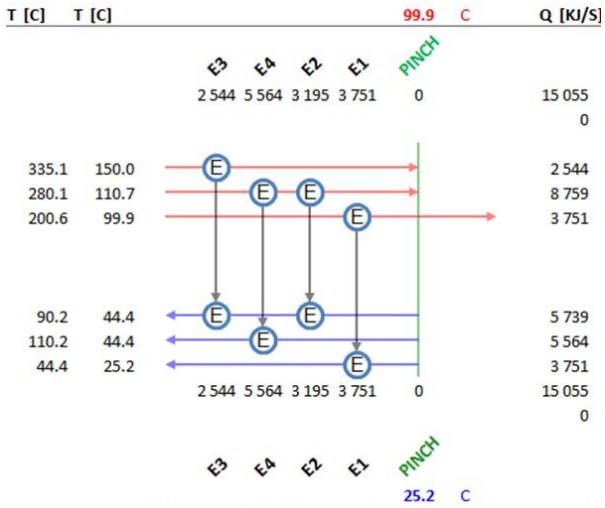
Monte Carlo simulation

Statistical theory of data reconciliation is exactly valid for linear models only. Any use of statistical theory to general nonlinear models is limited to a small neighborhood of the final solution. In practice this means that standard deviations of measurement errors should be "small" enough to justify acceptable validity of results (reconciled values, confidence intervals, gross errors detection, etc.). Analytical solution of this problem is not feasible in practice.

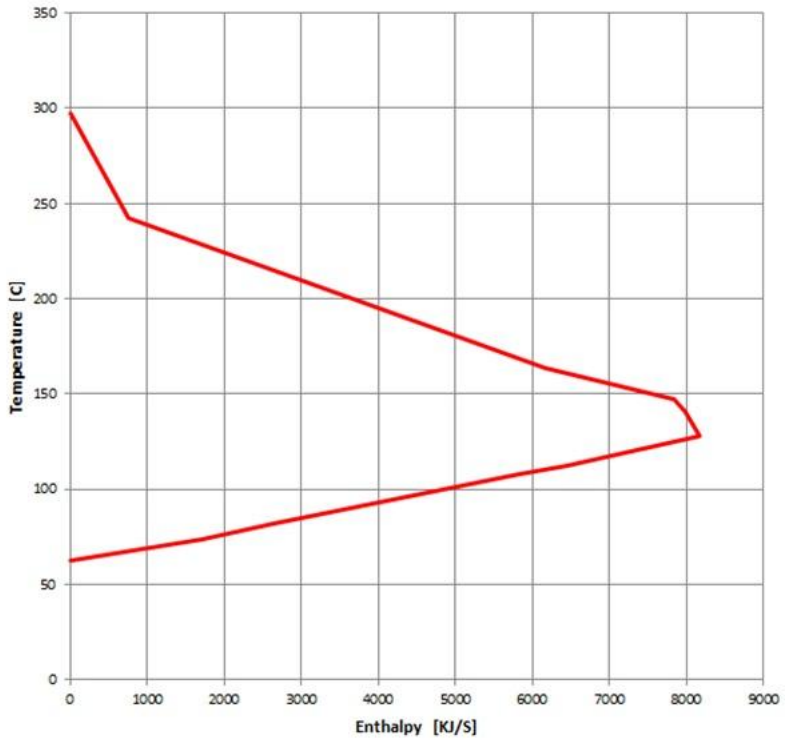
The only way how to tackle this problem is the Monte Carlo simulation. This solution is based on repeated simulation of the measurement process. The starting point is an errorless data set (a base case). Individual "measured" data sets are generated by adding random errors to the base case data. In this way a real measurement is simulated many times with the following data reconciliation. Results can be statistically evaluated and compared with theoretical values.

Pinch technology and heat integration

Recon heat balance models can provide also information directly applicable for deeper analysis of such systems by the Pinch Technology, Heat Integration, using Heat Pumps, etc. Necessary validated data for such activities are available in the form of MS Excel workbooks.



GRAND COMPOSITE CURVE (threshold: min dT = 74.68624 C)



Reporting engine

RECON enables one to create reports in the MS Excel environment. The solution is based on Excel templates prepared by users. These templates contain links to RECON's database. In this way any user can define unlimited number of report templates which can be later used for generation of reports. The complexity of these reports is limited by Excel's capabilities only. Reports can be generated either in the interactive way by a user or automatically at a pre-defined time.

RECON's connectivity

RECON can be connected to process data information systems based on the Oracle, MS SQL, PI System, Industrial SQL server, PHD, AIM*, Exaquantum, OPH and MS Access databases as well as to MS Excel, .DBF or .TXT files.

How to set up data sources is shown in Recon Data connectivity video.

RECON's physical property database

RECON contains the following databases of physical properties:

- The IAPWS IF 97 database of properties of steam ad water
- Properties of hydrocarbons and other chemicals according to API procedures
- SRK equation of state to estimate gas compressibilities
- Critical parameters of components for density and viscosity calculations of gaseous mixtures
- User defined physical property models

RECON languages

- Czech
- English
- German
- Spanish

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
- Simulation and “What if?” Queries via Web services
- As ActiveX DLL called from other programs.

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 (XP, Vista, Windows 7: 32-bit or 64-bit processor, Windows 10: 32-bit or 64-bit processor). The minimum recommended processor is Pentium with RAM 128 MB.

Available versions

Version	Mass balance	Energy balance	Momentum balance	User-defined equations	DB connectivity	Active X object (DLL)
Lite*	x	x	x	x		
Academic**	x	x	x	x		
Full	x	x	x	x		
Professional	x	x	x	x	x	x
Trial***	x	x	x	x	x	

* max 50 nodes, 100 streams, 10 heat exchangers, 10 components and 20 user-defined equations

** for non-commercial purposes only

*** valid for 1 month