2
Data Validation: a Technology for Intelligent Manufacturing

Boris Kalitventzeff, Georges Heyen, and Miguel Mateus

2.1
Introduction

This document is intended to progressively demonstrate the technical assets of the data validation technology. Most of the technical features of the technology will be enlightened by specific process systems. However, validation technology can be and is implemented in various industrial sectors. Namely, it covers chemical, petrochemical and refining process plants, thermal and nuclear power plants, upstream oil and gas exploitation fields. Data validation is an extension of data reconciliation. Before demonstrating the technical assets of the validation, the reconciliation concept will be reviewed.

2.2
Basic Aspects of Validation: Data Reconciliation

Data reconciliation (DR) is the first mathematical method that addressed the concept of data validation for linear problems. It exploits information redundancy and (linear) conservation laws to extract accurate and reliable information from measurement data and from the process knowledge. It allows for the production of a single consistent set of data representing actual process operations, assuming the plant is operated in a steady state.

To understand the basic principles of data reconciliation, one must first recognize that plant measurements (including lab analyses) are not 100% error free. When using these measurements without correction to generate plant balances, one usually gets incoherence in these balances.

Some sources of errors in the balances directly depend on sensors themselves:

- intrinsic sensor accuracy,
- sensor calibration,
- sensor location.

Computer Aided Process and Product Engineering. Edited by Luis Puigjaner and Georges Heyen

ISBN: 3-527-30804-0

A second source of error when calculating plant balances is the small variations in the plant operating conditions and the fact that samples and measurements are not exactly taken at the same time. Using time averages for plant data partly reduces this problem. However, lab analyses are usually carried out at a low frequency, and thus can seldom be averaged.

Finally, one must also realize that in some parts of a plant too many measurements are available, whereas in other parts some measurements are missing and must be back-calculated from other measurements.

As shown in detail in Section 3, Chapter 3 of this book, data reconciliation can be expressed mathematically as:

$$\text{Min} \sum_i \left(\frac{y_i^* - y_i}{\sigma_i} \right)^2 \tag{1}$$

subject to $F(x, y^*) = 0$

$G(x, y^*) \geq 0$

where y^*_i is the reconciled value of measurement i,

where y_i is the measured value of measurement i,

x_j is the unmeasured variable j,

σ_i is the standard deviation of measurement i defining its confidence interval.

$F(x,y^*) = 0$ corresponds to the process equality constraints.

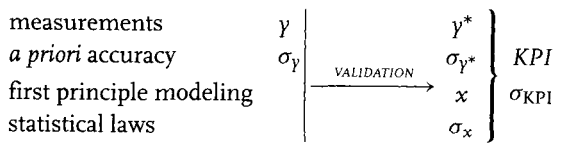
$G(x,y^*) \geq 0$ corresponds to the process inequality constraints.

The term $\left(\frac{y_i^* - y_i}{\sigma_i} \right)^2$ is called the *penalty* of measurement i.

In early publications on DR, equality constraints were considered linear. Thus, one obtains a quadratic formulation, where the Jacobian matrix of F is constant. It is a Gaussian regression problem: given a set (y, σ_y), the algorithm provides x and y^* vectors together with their standard deviation σ_{y^*} (when computed).

When inequality constraints were not considered, some values y or x could be negative, what had no physical meaning in chemical or mechanical processes, where most variables must be positive (e.g., pressure, flow rate, mole fraction). It was considered as a source of information because one had to find which measurement was responsible for that negative value. Later on, simple inequalities ($y \geq 0$, $x \geq 0$) were considered.

When F or G is nonlinear, the DR problem can be solved by sequential linearization. The minimization problem is solved iteratively, using algorithms such as SQP (sequential quadratic programming). It is now possible in some commercial codes to calculate not only the reconciled values of measurements (y^*, σ_{y^*}) but also unmeasured state variables (x, σ_x) and some key performance indicators (KPIs) related to measured and unmeasured state variables (y^*, x), as well as their uncertainty σ_{KPI}:



Another problem is the identification and elimination of gross errors in the measurement data. This is a key asset of modern validation tools. We use the term data validation instead of DR, when all these features are exploited in the corresponding software tool, together with the use of thermodynamic conservation and equilibrium laws besides the mass balance equations.

The validation problem is solved using an interior point SQP solver in the VALI software that will be considered as an example in the subsequent applications [1].

2.2.1 Redundancy Analyses: Local/Overall

The level of redundancy is the number of measurements, which are available beyond the absolute minimum needed to calculate the system. Three different cases can be encountered:

- If a system's redundancy is negative then there is not enough information to determine the state of the system. Additional measurements need to be introduced.
- A redundancy equal to zero means that the system is globally just calculable.
- And finally, if a system has a positive redundancy, DR can use it as a source of information to correct the measurements and increase their accuracy. In fact, each measurement is corrected as slightly as possible but in such a way that the reconciled measurements match all the constraints of the process model.

However, overall redundancy is not enough. It must also be achieved at the local scale. Indeed, redundancy can be positive at the global scale, but negative locally; consequently, information is lacking to completely describe the whole process.

This point is illustrated with Fig. 2.1, based on a typical synthesis loop. Components A and B are introduced into the process feed, and converted into component C in the reactor unit SYNTHES (2C = 3A + B). Afterwards, the product ABC is separated in three distinct streams. One is recycled upstream in the process, another represents a purge, and finally an outlet stream contains only the compound C.

Let us consider a process model restricted to mass balances. Measured variables are shown on Fig. 2.1. This simple process model presents a global redundancy level of 2 (20 equations for 18 unmeasured variables). However, local redundancy of unit SEP-2 is equal to zero. If one of the measurements around this unit was missing then global redundancy of the model would still be 1 but local redundancy of unit SEP-2 would be −1. Therefore, the system would not be reconcilable until a supplementary measurement around the mentioned unit has been provided.

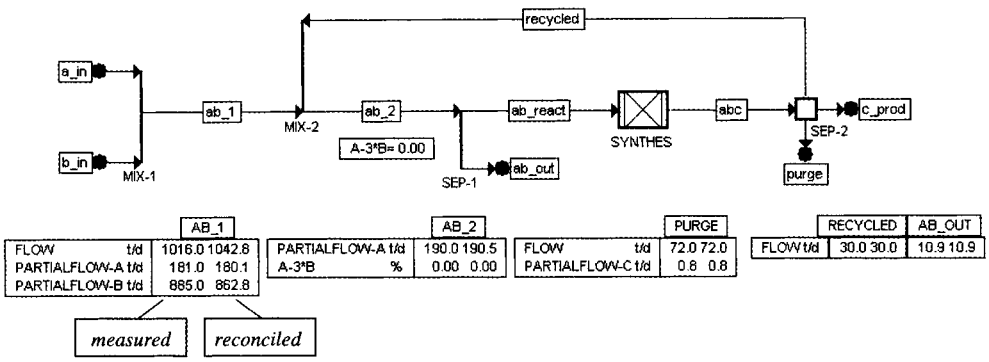


Figure 2.1 Process flow diagram (PFD) of a synthesis loop

2.2.2
If Complementary Measurement(s) are Needed: Which One(s)?

If the available measurement set is not enough to calculate all required process performance parameters, how do you propose an extra set from which complementary measurements can be chosen? Thus, the system becomes either just calculable or locally redundant, but necessarily globally redundant, as illustrated before.

Consider the previous example, but here we would remove the total flow rate measurement of stream "purge". Reconciliation software would then propose a set of variables from which possible complementary measurements ought to be chosen. Namely, the software would purpose in this case a choice between partial flow rates of compounds A and B in either stream "abc" or "purge", or compound C partial flow rates in either stream "abc" or "c_prod".

If it is not possible to add any measurement to the system (because of economical constraints for example), another way of avoiding negative redundancy is to aggregate some units in the model as a more global "black box" (that simply ensures global balances to be satisfied). Less information will be obtained locally, but this may allow estimating the required KPIs.

2.2.3
Increased Accuracy on Measured Data: Why?

As explained before, data reconciliation is based on measurement redundancy. This concept is not limited to replicate measurements of the same variable by separate sensors; it includes the concept of topological redundancy, where a single variable can be estimated in several independent ways, from separate sets of measurements. Therefore, *a posteriori* accuracy of validated data will be better than *a priori* accuracy of measured data. *A priori* and *a posteriori* means before and after consistency treatment, or in other words before and after validation and reconciliation.

Table 2.1 DR back corrects measurements and increases their accuracy

			Meas.	Meas. Acc.	Reconc.	Reconc. Acc.
AB_1	Flowrate	ton/d	1016,0	3,00 %	1042,8	1,64 %
	Partial Flowrate (A)	ton/d	181,0	3,00 %	180,1	2,98 %
	Partial Flowrate (B)	ton/d	885,0	3,00 %	862,8	2,00 %
RECYCLED	Flowrate		30,0	3,00 %	30,0	3,00 %
AB_2	Partial Flowrate (A)		190,0	3,00 %	190,5	1,60 %
	A−3*B		0,0	0,00 %	0,0	0,00 %

In the previous example, unit MIX-2 presented a level 2 redundancy. Indeed, for 5 equations and 9 variables (and thus 4 degrees of freedom) we have 6 measurements (6 – 4 = 2). Table 2.1 shows the *a priori* and the *a posteriori* accuracy of those measurements around unit MIX-2.

Reconciled measurements are more accurate than raw data when measurement redundancy is available. But when no redundancy is available locally, no improvement can be expected. This is the case for the estimation of the recycled flow rate: the measured value is not corrected, and its accuracy is not improved. When some measurement is not corrected that does not imply it can be trusted; this would only be the case if the standard deviation would decrease.

2.2.4 DR Avoids Error Propagation

Progress in automatic data collection has presented plant operators with a flood of data. Tools are needed to extract and fully exploit the relevant information it contains. Furthermore, most performance parameters are often not directly measured, but calculated from measured values. Thus, random errors on measurements also propagate in the estimation of KPIs. Data reconciliation, on the contrary, allows state estimation and measurement correction problems to be addressed in a global way. As a result, validation technology avoids error propagation, and provides the most likely estimate of the actual operating point of the process. Thus, the plant can be safely operated closer to its limits. Illustration of error propagation is addressed in Table 2.2 for the example considered in Fig. 2.1. The goal is to estimate the flow of component C in the process output.

Because raw measurements are not error free, mass balance equation around mixer MIX-1 is not respected (fourth row of Table 2.2). Cases 1 to 3 show what happen when each of the three (process inlet) flow rates are manually corrected to close the mass balance, the flow rate of stream C being computed afterwards. In the last case DR is used to provide a consistent and accurate set of reconciled measurements. Indeed, Table 2.2 shows a balance value equal to zero. Note that measurements may be considered as correct since reconciled values are inside their confidence limits.

Table 2.2 Error propagation

		Measured	Accuracy	Case 1	Case 2	Case 3	Reconcilied	Accuracy
A in	ton/d	181.0	3.00 %	181	181	131	180.1	2.98 %
B in	ton/d	885.0	3.00 %	885	835	885	862.8	2.00 %
AB in	ton/d	1016.0	3.00 %	1066	1016	1016	1042.8	1.64 %
Balance in	*ton/d*	*−50.0*	/	*0*	*0*	*0*	*0*	/
ABC Purge	ton/d	72.0	3.00 %	72	72	72	72.0	3.00 %
C out	ton/d	/	/	994	994	944	959.9	1.80 %

Knowing that the standard deviation of flow measurements is 3 % of the measured value, one obtains for outlet compound C the flow rate:

- with DR: a standard deviation equal to 1.80 % with an estimate of 960 ton/d;
- with manual correction: a spread of estimates equal to 5.03 % (from 944 to 994 ton/d).

Thus, DR avoids error propagation and so provides more accurate computed parameters than those calculated by less rigorous or *ad hoc* correction modes. Plant engineers have to solve that type of problem regardless of if they have the appropriate tools or not.

2.2.5 Process Measurements to be Exploited

Key performance indicators (KPIs) can be determined accurately by validation of process measurement data. They are very useful for many purposes, e.g., revamping, energy integration, improved follow-up of the plant, possibility of working closer to specifications, detecting degradation of equipment performance, etc.

A hydrogen plant process is used to illustrate the determination of accurate and reliable KPI. Namely, this example concerns the steam to carbon ratio (S/C) in the steam reformer feed, that is, one of the key control parameters in such plants. It allows controlling the conversion of methane to carbon oxide and hydrogen while avoiding carbon deposition on the catalyst. Two different cases were studied to compute this ratio:

- First, DR was not considered. Ratio S/C was calculated from raw measurements of flow rates and compositions of process inlets (steam and natural gas) and reforming gas recycled.
- Afterwards, the same KPI was determined by means of DR.

Each of these two cases were reassessed, considering a measurement error on the steam flow rate (e.g., due to a leak). Namely, the steam flow rate is measured at either 72 ton/h or 78 ton/h.

Results shown in Table 2.3 demonstrate that the uncertainty on the S/C ratio is reduced when data reconciliation is performed. Also, reconciled S/C ratio is less sensitive to the flow rate measurement error, which is detected and corrected by data

reconciliation. Thus, reconciliation detects errors in available measurements and yields accurately consistent and complete estimates of measured as well as unmeasured process parameters. Furthermore, in industrial practice one must take a safety margin for the S/C ratio to avoid carbon deposition in the catalyst. With DR, safety margins can be thinner, steam consumption is reduced and therefore plant operation costs less.

Table 2.3 KPI computation

	Without meas. errors		With meas. errors	
	S/C ratio	**rel. error**	**S/C ratio**	**rel. error**
without DR	3.545	4.24 %	3.840	4.24 %
with DR	3.514	3.52 %	3.673	3.53 %

Here a real industrial case encountered in a hydrogen plant is described, for which validation technology was applied. In a hydrogen plant (operated by ERE company), the feed gas composition was not monitored accurately; measurement errors were leading to an approximate knowledge of the steam/carbon ratio [2], uncertainty being on the order of 30 %. However, the hydrogen production efficiency and cost are strongly related to this ratio. Indeed a low S/C ratio decreases energy consumption. Therefore, a potential return of 500,000 euro per year had been identified. On the other side, a low S/C ratio could lead to carbon deposition (see Fig. 2.2) entailing a risk of catalyst damage (shut down for replacement costs five million euro).

With on-line validation software the steam/carbon ratio is determined nowadays with a precision of 1 %. This allows operating at the optimal point where energy costs are mastered and carbon deposition is avoided. This example shows how validation software allows for operation closer to the limits, taking care of safety constraints.

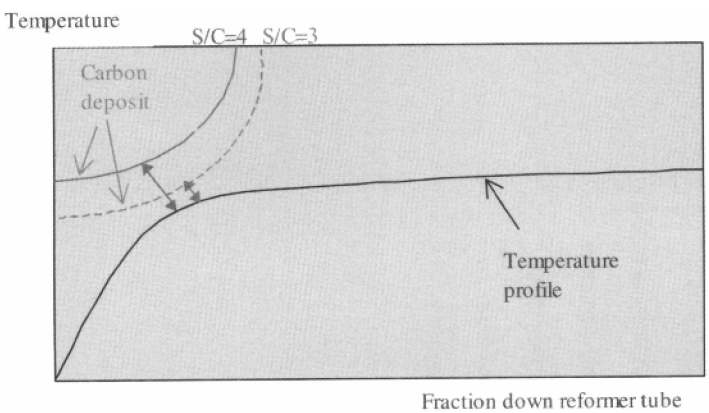


Figure 2.2 Profile of reformer reactor (courtesy of BP-ERE [3])

2.3
Specific Assets of Information Validation

Data validation is an extension of DR. In that case the set of corrected measurements and other calculated data respect linear and nonlinear constraints (mass, components and energy balances, reaction constraints as well as physical and chemical thermodynamic equilibrium constraints). Furthermore the technology includes data filtering, gross error detection/elimination, and it also provides the *a posteriori* accuracy of all the calculated data. Therefore, accurate and reliable KPIs are determined, as well as their accuracy. Moreover, validation software detects faulty sensors and pinpoints degradation of equipment performance (heat rate, compressor efficiency, etc.).

2.3.1
Accuracy of Nonmeasured but Calculated Data

Unmeasured variables of the system are calculated and their accuracy is quantified on basis of the measurements that are related to them. Therefore, in addition to providing substitution values for failed instruments, data validation software also calculates values that are not directly measured. Validation acts as a set of "soft sensors" that are robust and accurate because they are based on the reconciled values of all the measurements. Typically, validation technology provides three times more calculated data (and their accuracy), than the number of effectively measured data.

Benefits are undeniable, costly lab analyses can be avoided. For instance, on the chemical site of Wacker Chemie (Germany) an on-line implementation of validation software reduced the number of routine analyses up to 40 % (see Fig. 2.3) [3].

Wacker considered validation as a revolutionary way for quality follow-up of their plants: f_{obj}, the sum of weighted squares of measurement corrections were checked for three years (see Fig. 2.4) [3]. They showed a reduction of the objective function (f_{obj}) from 30,000 to 1000, demonstrating a better quality of sensors tuning. Any increase of that validation criterion alerts operators on possible plant upset.

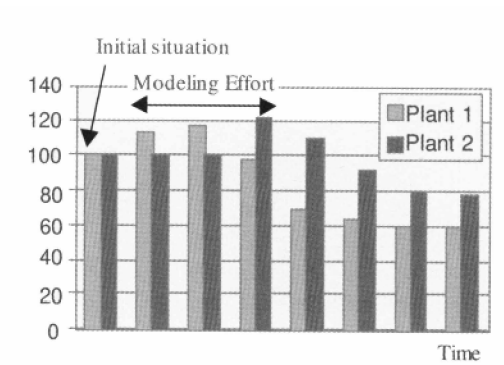


Figure 2.3 Reduction of lab analysis cost (courtesy of Wacker Chemie [3])

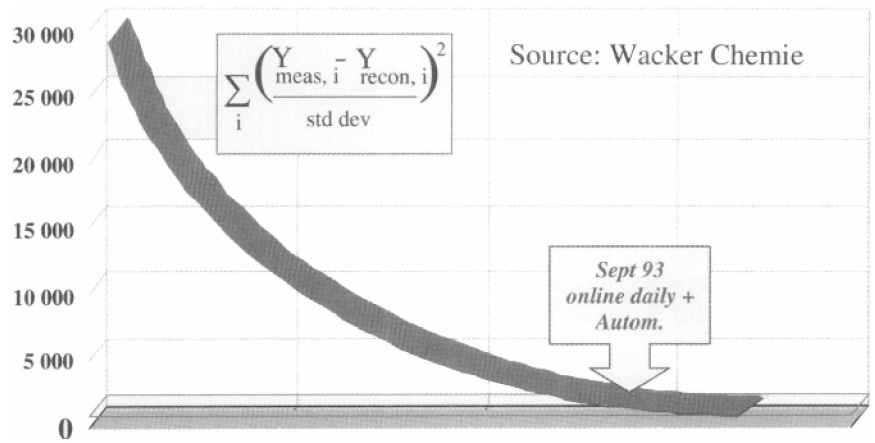


Figure 2.4 Sum of weighted squares of measurement corrections (courtesy of Wacker Chemie [3])

Furthermore, Wacker also follows the ratio $\frac{\chi^2}{f_{obj}}$, based on the chi-square statistical test (χ^2). The chi-square test value depends on the number of redundancies of the system and on the statistical threshold of the test, typically 95 %. Active bounds are considered as adding new levels of redundancy.

Two different cases are possible, whether the ratio is higher or lower than 1:

- If $\frac{\chi^2}{f_{obj}} > 1$: no presence of gross errors in the set of measurements can be expected.
- If $\frac{\chi^2}{f_{obj}} \leq 1$: presence of at least one gross error in the set of measurements is expected.

A data reconciliation result can only be exploited if the chi-square test is satisfied.

Gross error detection and elimination is a feature of validation software that will be detailed next.

2.3.2 Key Performance Indicators and Their Accuracy

Key performance indicators (KPIs) are identified in the same way as nonmeasured state variables. Because measurement errors have been withdrawn from the set of reconciled data, the best possible estimate of the plant performance is delivered. Thus, KPIs can be accurately determined.

Typical KPIs include:

- global plant efficiency
- yields

- steam/carbon ratio, oxygen/carbon ratio, H_2/N_2 ratio, etc.
- specific energy consumption
- specific energy cost
- equipment duty and efficiency
- catalyst activity, etc.

Table 2.4 shows S/C ratio values and accuracy using data validation technology. In the third case, thermodynamic constraints were taken into account. KPI accuracy is more improved with data validation than with data reconciliation. This is due to the fact that data validation considers all available process information (temperatures, pressures, chemical reactions, equilibrium constraints, etc.), the redundancy level being thus higher. Moreover, S/C ratio is much less sensitive to measurement bias, as demonstrated with the introduction of a measurement error on the steam flow rate entering the reformer (see Table 2.4). The additional assets of data validation are described here after.

Table 2.4 S/C ratio

	Without meas. errors		With meas. errors	
	S/C ratio	rel. error	S/C ratio	rel. error
without DR	3.545	4.24 %	3.840	4.24 %
with DR	3.514	3.52 %	3.673	3.53 %
with data validation	3.423	0.63 %	3.432	0.63 %

2.3.3 Nonlinear Thermodynamic-based Data Validation

2.3.3.1 The Limitation of (Linear) Mass Balance-based Reconciliation

Most commercial data reconciliation packages are based on a linear solver and reconcile measurements on the basis of overall mass balances. Moreover, bounds on variables are seldom considered, meaning that negative flow rates or negative inventories can appear in the results. Additionally, mass balance-based systems only offer a low level of redundancy: at the most one gets one level of redundancy around each node where all incoming and outgoing rates are measured. As a consequence, the improvement in data quality is low and the results are very sensitive to gross errors in the measurements.

On the contrary, thermodynamic-based data validation software provides additional equations increasing consequently the redundancy of the system, making it more accurate and less sensitive to measurement errors. At the same time, key performance indicators can be directly derived with a high level of accuracy and reliability. Of course, using thermodynamic properties has its drawback: most of the equations become nonlinear making linear solvers useless. Therefore, one must then use a nonlinear algorithm as large scale SQP-IP (sequential quadratic programming-

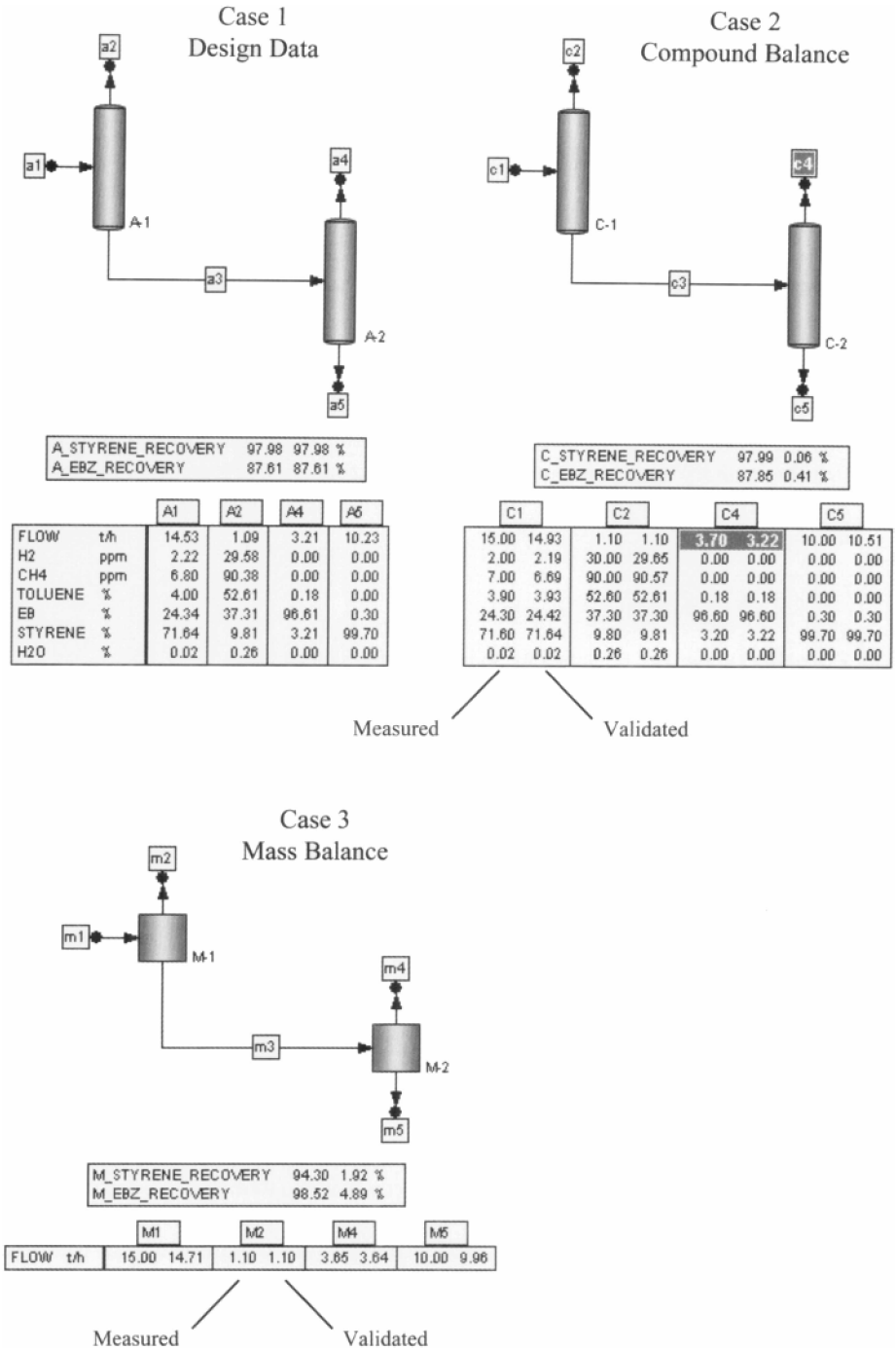


Figure 2.5 Compound balances influence on accuracy

interior point), which has been implemented to solve complex nonlinear data reconciliation problems.

2.3.3.2

Example: Reconciliation of Two Distillation Columns

Two consecutive distillation columns are used to separate styrene (the final product) from unreacted ethylbenzene (EB), which is recycled to the reaction section (see Fig. 2.5).

Case 1 presents the design mass (and compound) balance of the plant. Case 2 presents typical measured values with a significant bias on the flow rate of recycled EB (stream c4) as reconciled in a compound-based data reconciliation system. The bias is clearly identified (3.70) and corrected (3.32) so that the styrene and EB recovery are accurately determined (87.86 ± 0.41 %). Case 3 then presents the same flow rates reconciled using a simple mass balance system, which is unable to detect the measurement error and therefore calculates a wrong recovery of EB and styrene. One can see that the accuracy of the computed recoveries is considerably better when performing a compound balance than with a simple mass balance (in this case, more than ten times better).

2.3.4

Exploiting LV and LLV Equilibria as Source of Information

Variables describing the state of a process must be reconciled to verify consistency constraints representing basic laws of physics: dew point and boiling point constraints in condensers, evaporators, or distillation columns are a source of information exploited by thermodynamic-based validation software.

The process of industrial ammonia production may be subdivided into three distinct parts: synthesis gas production, compression section and ammonia synthesis loop. Process natural gas (PNG) and steam enter the primary reformer reactor, after sulfur removal of PNG. High temperature and low temperature shift sections follow the secondary reforming, where compressed air is also introduced. After the methanator section, synthesis gas is partially recycled upstream in the process and partially introduced in the hyper compressor section. Finally, gas enters the ammonia synthesis loop. Figure 2.6 represents an ammonia synthesis loop process flow diagram (PFD), which can be considered as having an 8-digit structure with a heat exchanger in the middle.

The synthesis gas enters the hyper compressor as well as the recycle gas, then the outlet (process gas) is cooled and partially condensed (106F) to recover ammonia. Afterwards, gas is heated through a counter-current heat exchanger, goes to the reactor section, then again to the same heat exchanger (at lower pressure than the cold process gas) before closing the synthesis loop.

Condenser temperature (see Table 2.5) reflects a compromise between ammonia content and flow rate of the gas entering the reactor section. Considering condenser pressure as constant (158 bar) to simplify the following illustration, and condenser

Table 2.5 Condenser 106F measurements

		Raw measurements	Validated measurements
Condenser 106F	T	−14 °C	*−16.50°C*
	Vapor flow rate ($Nm^3\ h^{-1}$)	456,890	**455,040**
Reactor	P_{in}	165 barg	165 barg
	T_{in}	185 °C	181 C
	%mol NH_3	2.4	***2.829***
	%mol inerts	11.2	11.07

inlet composition and vapor flow rate specified, three different "what if" cases were studied (see Table 2.6). First, temperature was assumed equal to measured temperature −14 °C. In the second column, temperature was considered the same as validated value −16.5 °C. Finally, temperature is computed for ammonia content in vapor phase identical to raw measurement 2.40 %.

Thus, a large amount of information can be extracted from the results:

- At 158 bar, hydrogen solubility rises slightly with temperature.
- If temperature is considered equal to the raw measurement (−14 °C), ammonia vapor composition estimated is considerably different from measurement (3.1 % instead of 2.4 %). This proves inconsistence in the measurement set. On the contrary, vapor flow rate computed seems closer to that of the measurement value.
- In the second "what if" case, we reproduce validated data.
- To reach specified reactor inlet ammonia content (2.4 %), temperature should be −20.8 °C, instead of the −14 °C measured. Therefore, vapor flow rate decreases.

This illustration shows the limitations of any partial "manual" validation.

Why is validated data so important in this particular case? The "what if" computations show the size of uncertainty of different data. The more NH_3 you condense in the condenser the better, but this has a direct cost, the energy spent in the cooling loop. How do you optimize any compromise if only nonvalidated data are available? Does it make sense?

Table 2.6 LV equilibrium calculation results

T (°C)	−14	−16.5	−20.8
Vapor fraction	0.9586	0.9558	0.9517
%mol NH_3 in vapor phase	3.10	2.83	*2.40*
Vapor flow rate ($Nm^3\ h^{-1}$)	456,330	455,039	453,049
%mol H_2 in liquid phase	0.38	0.36	0.33
Liquid flow rate	14.95	15.93	17.44

2.3.5
Exploiting Reactions and Chemical Equilibria as Source of Information

This point can be illustrated with the same ammonia process described previously (see Fig. 2.6), in particular its reactor section. Ammonia is produced in a two adiabatic catalytic stages reactor. Reactants are nitrogen and hydrogen, entering the reactor in a stoichiometric mixture. Ammonia formation reaction is exothermic and reversible; therefore, gas leaving the first adiabatic stage is cooled before entering the second stage. Furthermore, the model considers a performance equation, consisting in the introduction for both adiabatic stages of a ΔT_{eq} parameter, which takes into account deviation from chemical equilibrium. Because reaction is exothermic, ΔT_{eq} will be positive.

Thus, important information that can be extracted from data validation, considering reactions and chemical equilibrium, are performance parameters ΔT_{eq} (see Table 2.7). Results pinpoint a closer approach to equilibrium in the first catalyst bed. In addition, it is possible to visualize validated ammonia concentration profile together with equilibrium curve and plant measurements (see Fig. 2.7). The two vertical lines represent measured inlet and outlet temperatures of the heat exchanger between the two catalyst beds.

One cannot accept a measurement point above the equilibrium curve. This erroneous measurement set could not have been noticed any other way than exploiting reactions and chemical equilibria as an information source.

Table 2.7 Performance parameters

	ΔT_{eq} (°C)
First catalytic bed	6
Second catalytic bed	14

2.3.6
Exploiting Process Information

As explained before, data validation is based on measurement redundancy. The plant structure yields additional information, which is exploited to correct measurements. Consequently, considering a process at a global scale brings more accuracy to validated data than only taking into account a local section of the process. It is the same for the accuracy evolution of key performance indicators.

Considering the same ammonia process as before, the H_2/N_2 ratio in the synthesis loop was estimated in several ways. First, only a local section of the process was considered (the synthesis loop). Then, additional information of the plant was successively added until the whole process was taken into account. Results pinpoint a substantial reduction of the KPI inaccuracy when more and more process information is considered (see Fig. 2.8).

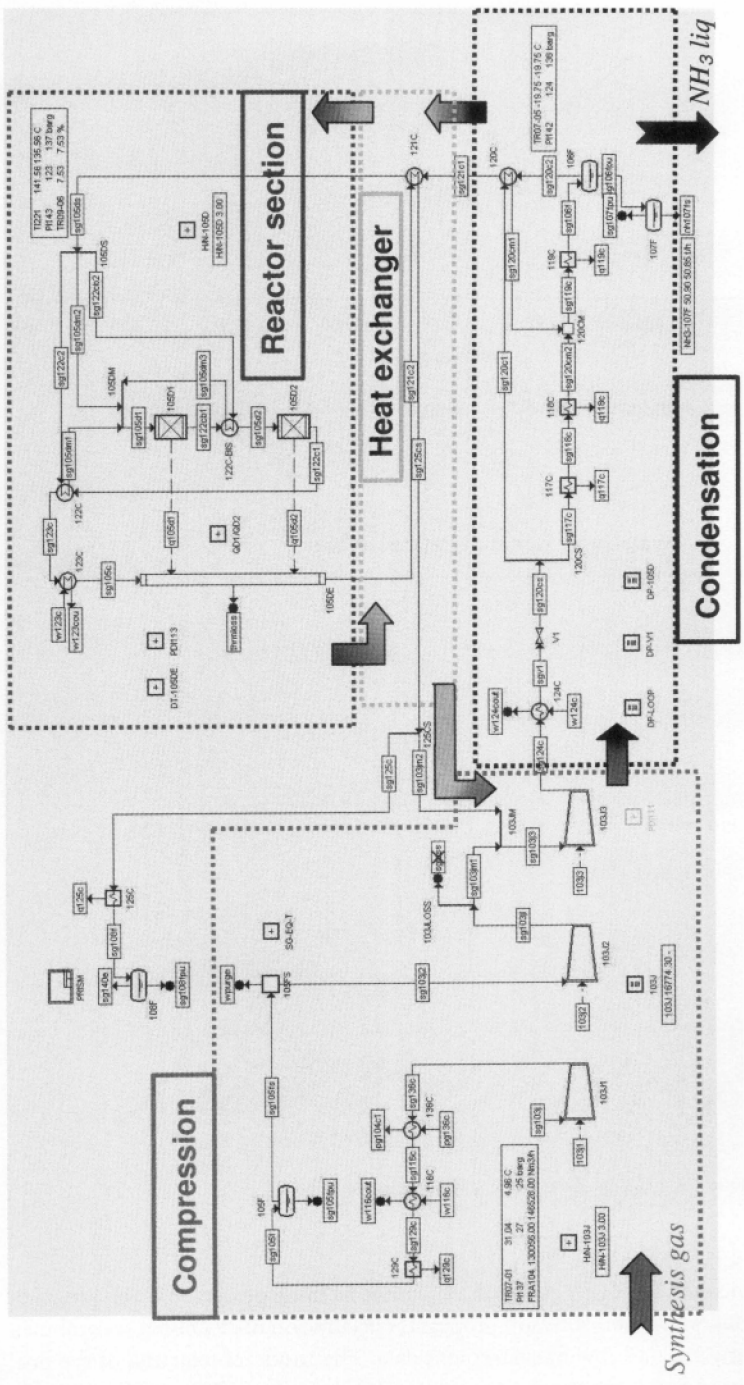


Figure 2.6 Ammonia synthesis loop PFD

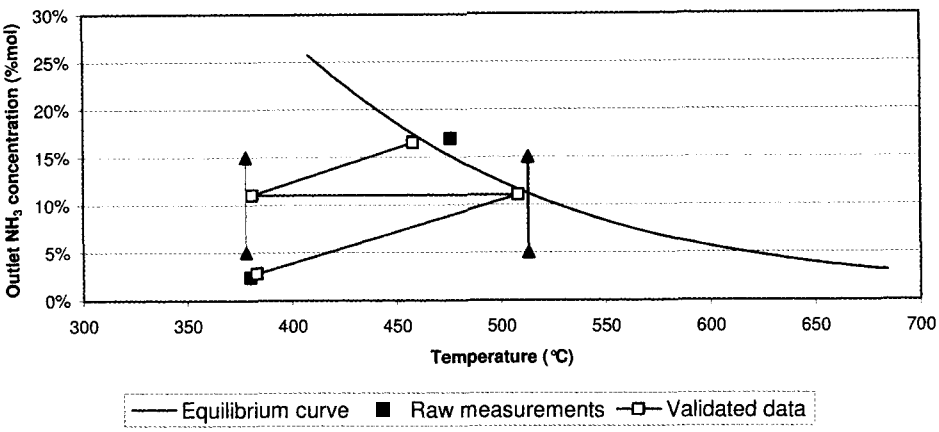


Figure 2.7 Synthesis reactor equilibrium curve

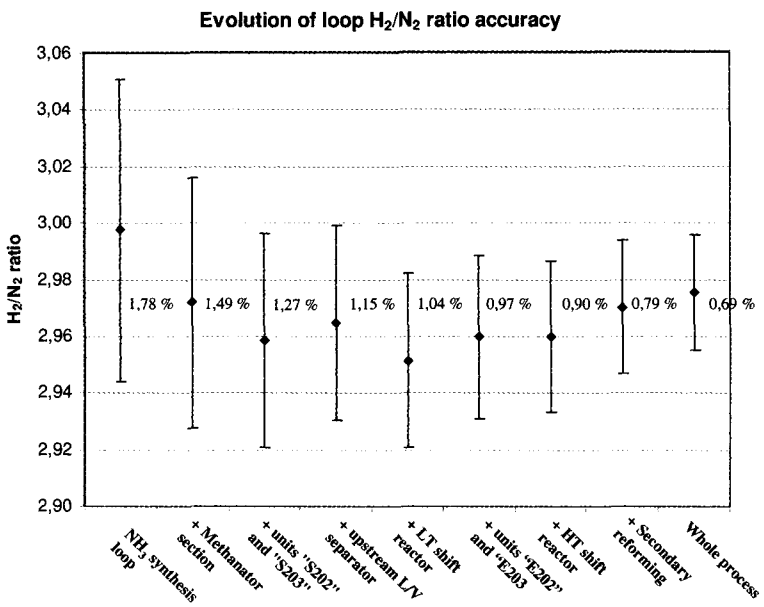


Figure 2.8 Evolution of a KPI imprecision according to process information taken into account

It was previously demonstrated that validation technology avoids error propagation. In fact, data validation software propagates accuracy. This technology combines process information and raw measurement data. The more information of the process taken into account, the more nonmeasured data (and so KPIs) will be accurate and reliable.

2.3.7
Detection of Leaks

Validation technology points out process performance degradation sources and helps to operate the plant closer to its ultimate performance. In particular, validation allows the detection of leaks. This can be illustrated by a practical case study related to a previous ammonia plant, where a leak in a NH_3 synthesis loop was discovered. It would have been hardly detected by tools other than validation technology.

A Carbochim plant operated in Belgium at 90 % of nominal capacity; a retrofit was studied to restore the expected capacity. Validation conveyed a leak in the heat exchanger in the middle of the 8-structure synthesis loop (see Fig. 2.6). Thus, part of the process gas was cycling around from the compressor and condenser section, to the heat exchanger and again to the compressor. That leak had not been suspected. It probably developed and increased smoothly, but the question is, how could it have been discovered in the absence of the appropriate tool? Plant was shut down for isolating leaking tubes in the exchanger and was reopened to easily achieve the expected production rate without any costly additional investment.

2.4
Advanced Features of Validation Technology

2.4.1
Trivial Redundancy

Trivial redundancy cases are met when the validated value of a measured variable does not depend at all upon its measured value but is inferred directly from the model.

This can occur in particular in L/V equilibrium drums, where complementary thermodynamic constraints must be respected. Indeed if, e.g., temperature, pressure, flow rate and composition of a condenser inlet stream were known together with the unit pressure drop, any complementary measurement (e.g., outlet temperature) would be considered as a trivial redundancy. Proper validation software detects trivial measurements, which then are no longer considered as measured. As a consequence, their measurement accuracy will not affect the accuracy of the respective validated variable.

2.4.2
Gross Error Detection/Elimination

Gross errors are detected by means of a chi-square (χ^2) statistical test, which has been previously explained at Section 2.3.1.

2.4.2.1
Detecting Gross Errors

The χ^2 statistical test enables the detection of gross errors in the sets of measurements. The χ^2 value depends on the total number of redundancies of the system, active bounds being considered as adding new levels of redundancy, and on the sta-

tistical threshold of the test, typically 95 %. If the weighted sum of penalties is higher than the χ^2 threshold value, then there is a significant suspicion that gross errors exist. In such a case, all results obtained with that model are to be used with caution: validated values, identified performance factors and their reconciled accuracy.

2.4.2.2

Eliminating Gross Errors: The Highest Impact Method

Identifying the actual source of the gross errors is not always trivial and requires a careful analysis of the results. The conventional technique (highest penalty method) is to ignore the measurements for which the highest corrections are made. This method is known to be inadequate in detecting some gross errors, for example, when the corresponding measurement is specified with a high level of accuracy as compared to the other measurements.

On the contrary, the highest impact method detects the impact on the total sum of penalties by removing each of the measurements. This approach is in principle highly time-consuming and is therefore not used by most data validation packages. However, by means of specific algorithm, one can apply this technique in a calculation time of the same order of magnitude as a single validation run.

2.4.3

How to Validate with Petroleum Fractions

The modeling of a refinery process or a part of it is always confronted by the complexity of the petroleum and its products. Indeed, crudes and petroleum cuts are mixtures of a large number of chemical compounds, thus making it very difficult to model their properties without accurately knowing their composition. Therefore, it is common practice to model such streams by the well known pseudo-component concept.

2.4.3.1

Concept

A pseudo-component is a hypothetical molecule characterized by its density and its boiling temperature. Those parameters are then used to estimate the other thermodynamic properties (like critical properties or specific heat capacity) using empirical correlations as proposed for example by American Petroleum Institute (API). According to the crude type and origin, different pseudo-components must be used to get an accurate representation. The usual way of characterizing petroleum fractions is to generate a defined mixture of pseudo-components, with given boiling point, having the same properties as the Petroleum fraction. Namely, their composition and their density are identified in order to match all stream distillation curves and densities. Most common standards for distillation curves are true boiling point (TBP) and ASTM; each of them can be expressed on a weight basis or on a volume basis (see Fig. 2.9).

Several petroleum cuts involved in a distillation process can be modelled as a data validation system involving:

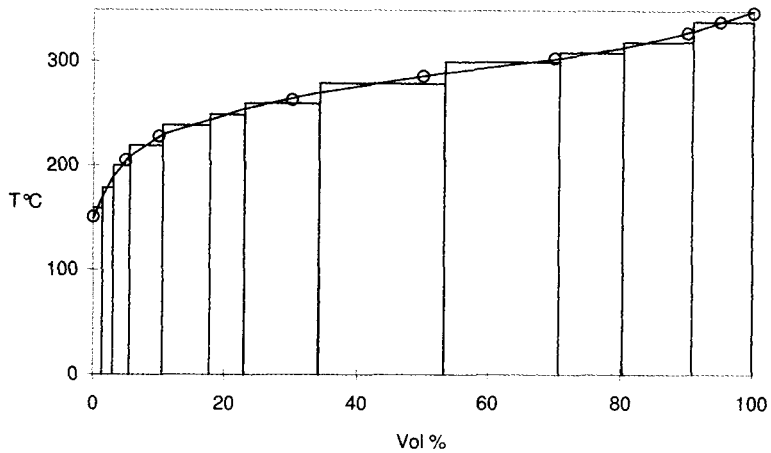


Figure 2.9 Decomposition in pseudo-components based on a TBP curve of a gas oil

- as variables, the pseudo-compounds densities and their measured volume fractions in each stream;
- as equations, the mass balances of the distillation column for each pseudo-component;
- as measurements, the densities and TBP or ASTM curves of all connected streams.

On this basis, data validation will generate calculated distillation curves from measured TBP or ASTM data, as it identifies the density of each pseudo-compound; this involves minimization of weighted deviation between measured and calculated distillation points, under density constraints, mass and thermal balance constraints. The other thermodynamic properties of the pseudo-compounds are also estimated.

2.4.3.2

Crude Oil Atmospheric Distillation Example

Following example concerns the modeling of a crude oil distillation unit (CDU), preceded by the preheating train (see Fig. 2.10) [4]. The crude oil is separated into six Petroleum cuts: naphtha, jet, kerosene, gasoline, diesel and residue.

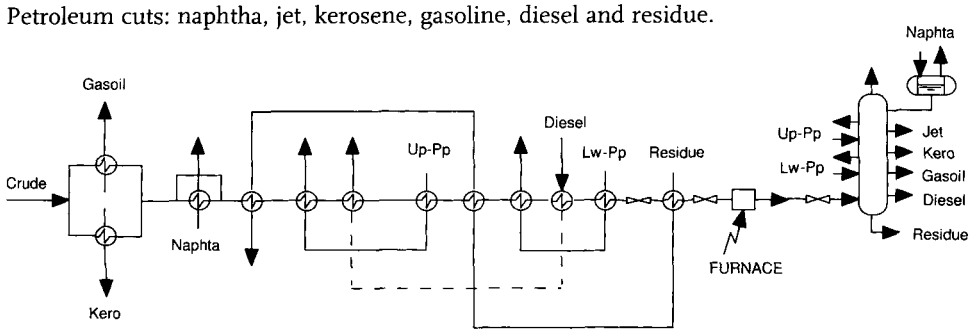


Figure 2.10 Preheat train and CDU

Measurements available to perform the modeling are:

- density and distillation curves (ASTM-D86) of the petroleum cuts,
- temperature, pressure and flow rates of the streams,
- design data of the exchangers.

These measurements are validated and the other thermodynamic properties of pseudo-compounds are subsequently computed. Furthermore, with several sets of measurements taken in one year it was also possible to confirm fouling problems for the exchangers at the end of the preheating train. Indeed, their heat transfer coefficient decreased by a factor of two after one year of operation. Thus, data validation uses a rigorous method integrating robustly complex distillation systems. This forms a sound basis for the analysis of refinery performance, and for instance, of a retrofit potential.

2.4.4
Advanced Process Control Benefits from Working with Data Validation

Nowadays plants face a market where margins are under pressure due to global competition, more stringent environmental regulations, a higher demand for flexible operation and more severe safety requirements. Control techniques are required to increase those margins. Advanced process control (APC) systems can help optimize control to deal with those challenges [5]. Data validation technique enhances the quality of information allowing APC systems to work more efficiently.

2.4.4.1
Data Validation-APC: How They Work Together

Figure 2.11 describes how data validation software works hand to hand with an APC system in order to improve its efficiency. A plant process is permanently submitted to disturbances, causing changes in operating conditions. APC systems use a reduced dynamic model to predict the plant behavior when submitted to disturbances, and thereafter take the adequate actions to counter them (multiple input multiple output, MIMO control). Without data validation, input and output stream measurements are introduced directly into a dynamic model with no ways of chec-

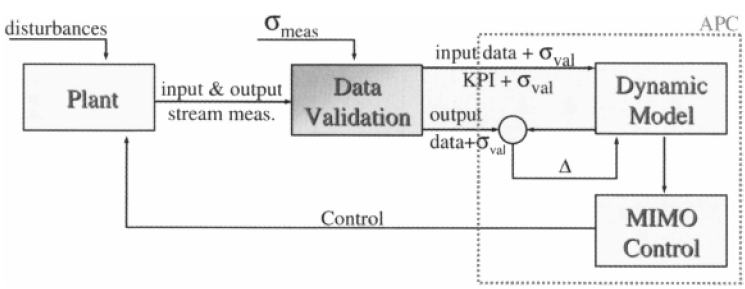


Figure 2.11 Data Validation working together with APC system

king raw information reliability and coherency. Some measurements could be erroneous and balances not be closed.

Data validation software uses input and output streams of raw measurements in order to provide one coherent and accurate data set. With data validation, APC systems are allowed to take actions on the process based on coherent and reliable measured and nonmeasured data. Validated data contains measurements, equipment parameters, KPIs, and many other nonmeasured but validated data. The *a posteriori* accuracy of measurements and KPIs is provided.

When a dynamic model is tuned according to validated data, benefits are generated as early as at the model design stage.

2.4.4.2

Benefits at Model Design Stage

Reduced dynamic model must be certified: dynamic model parameters are chosen and adjusted in order to produce results identical to measurements ($\Delta = 0$ on Fig. 2.11). Benefits using validation techniques are double:

- Measurements, to which dynamic model results are compared, are checked and corrected by data validation techniques. Measurements are much more reliable (they represent the actual process operation) and thus the model will be more reliable as well.
- Data validation technology reduces the number of principal directions needed to represent process variability, allowing the reduced dynamic model to represent the same level of variability using a model with a lower number of principal directions (see Fig. 2.12) [6].

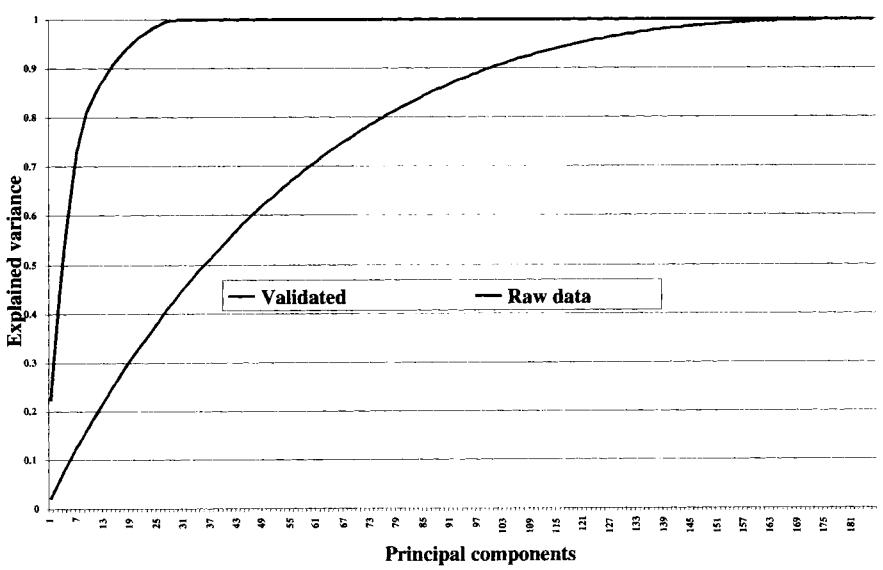


Figure 2.12 Level of variability according to number of principal directions

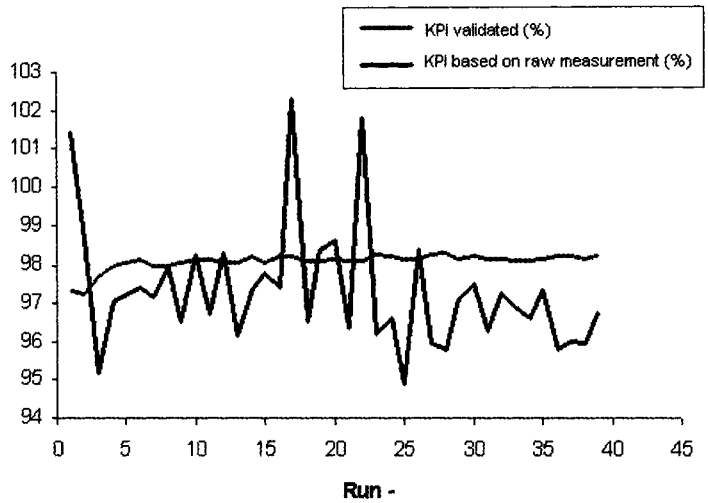


Figure 2.13 KPI follow-up and control

Figure 2.12 illustrates the number of principal directions (or components) necessary to represent the variability of a given system when the latter is based on validated data or on raw measurement data. Taking into account more principal components allows for the explanation of a higher fraction of the total process variability:

- When using raw measurements, a large number of components are needed to explain most of the process variability (upper limit is the number of original variables, 186).
- When using the validated data sets, the number of significant principal components tends to a much lower number than the number of variables (upper limit is the number of degrees of freedom of the data validation model).

This reduction in the problem size allows the dynamic model to be more reduced, when based on validated data (accuracy increased and noise reduced). Control of the process is made easier and the computing demand is decreased.

Furthermore, since data validation technique enforces the strict verification of all mass and energy balance constraints, use of this technology ensures that the principal components represent the proper process behavior.

2.4.4.3

Benefits at Operation Stage

Process control behavior can be very different whether APC is working together with data validation or not. Figure 2.13 presents the evolution of a process yield (KPI) versus time (run) whether data validation software is used or not:

- Without data validation, APC detects a KPI variation and tries to stabilize the process operation. Based on raw data with embedded errors, APC takes actions risking being unuseful, resources-expensive, and even process-disturbing.

- With data validation, APC considers actual process operation (validated, measured and nonmeasured, data are used as inputs to APC). APC can now use all of its resources on optimization of the process rather than on more stable operations.

2.5 Applications

2.5.1 On-line Process Performance Monitoring

The goal is to deliver on a periodic basis (typically each 10 to 60 minutes) a coherent heat and mass balance of a production unit. In addition to the compound balances, the laws of energy conservation are introduced in the form of heat balances. This more detailed modeling of the production unit allows validation software to work as an advanced process soft sensor and to determine reliable and accurate KPIs.

Typical benefits are:

- access to unmeasured data, which is quantified and related accuracy determined;
- early detection of problems: sensor's deviation and degradation of equipment performance are pinpointed;
- quality at process level: anticipate off-spec products by carefully monitoring the process;
- work closer to specifications: as the accuracy of measurement data improves, the process can be safely operated closer to the limits. This feature is reported as being financially the most productive.
- decreased number of routine analyses (up to 40 % in chemical applications);
- reduced frequency of sensor calibration (only faulty sensors need to be calibrated).

Improvement of Product Selectivity in a BASF Plant

This example shows how the operation of a production unit at a BASF operating division of performance chemicals can be improved using data reconciliation [7]. The product C is produced by conversion of component A with component B using 2 reactors. Several undesired by-products are generated, thus selectivity has to be maximized. Process model generated took into account only component mass and atomic balances. Several data sets at different process conditions were validated and from those the selectivity of product C was calculated. The diagram Fig. 2.14a (courtesy of BASF) shows this selectivity as a function of residence time in the first reactor, calculated from measured values; Fig. 2.14b shows the results from validated data.

The selectivity, calculated from crude data, is spread widely and in some cases selectivity values of more than 100 % were obtained, which is meaningless. The corresponding unfeasible area is marked on the charts. One could estimate in this case that a residence time of about 45 minutes is enough to maximize selectivity. However the selectivity based on reconciled data shows a clearer trend and does not

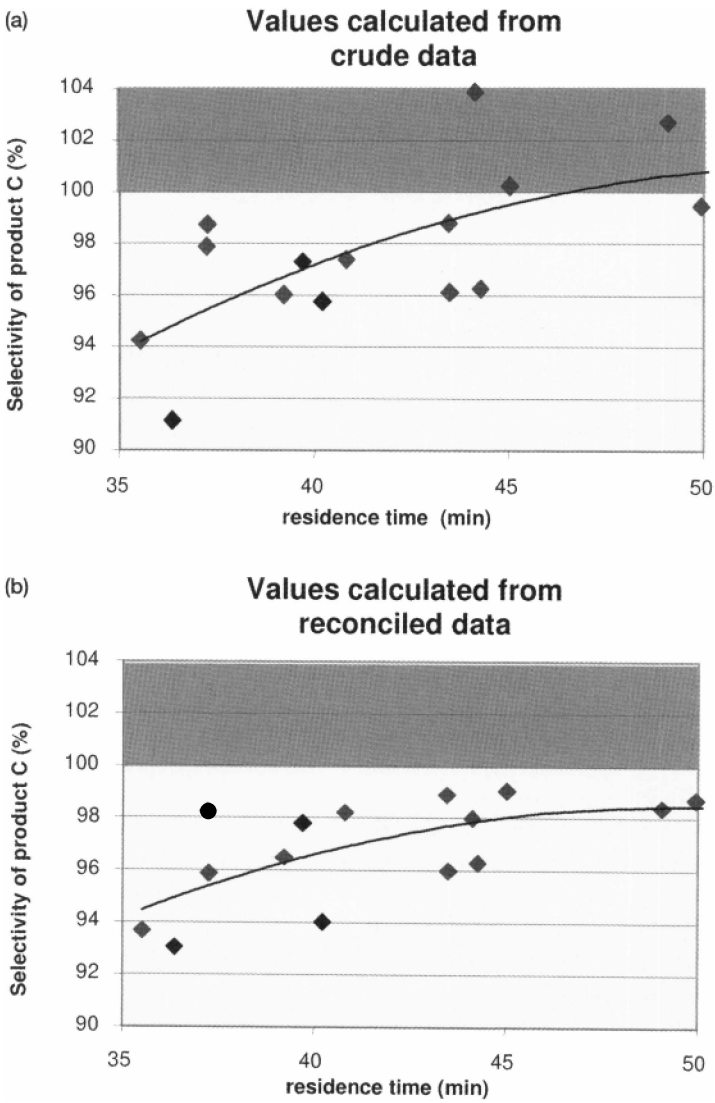


Figure 2.14 Nitrile selectivity as a function of residence time in the reactor: raw data versus reconciled data (courtesy of BASF [7])

exceed the 100 % boundary. One realizes that residence time should be larger than the one estimated without data validation, in order to achieve the product optimal selectivity (residence time of about 48 minutes). This example (considering only a restricted part of a process) shows that the evaluation of selectivity is meaningful only on the basis of validated operational data. These lead to a safe interpretation of measurements. By doing so, a selectivity close to 99 % can be obtained systematically, which is 2 % higher than the average figure obtained without data validation.

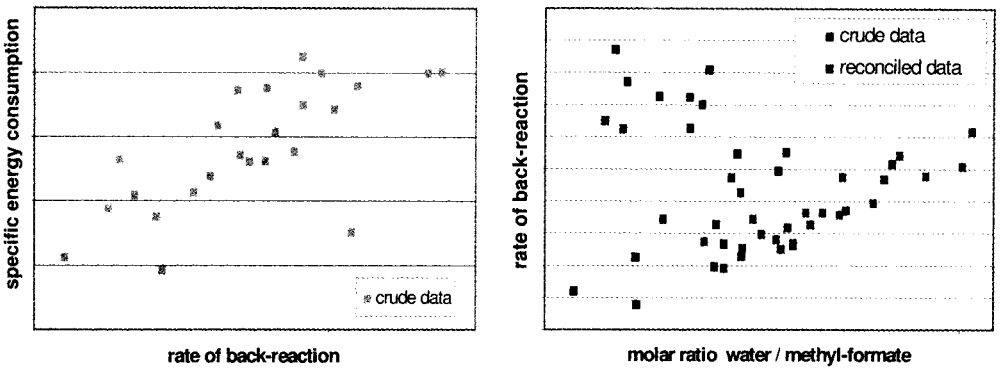


Figure 2.15 Evolution of the specific energy consumption as a function of the rate of back-reaction and parameters that influence this rate: crude data versus reconciled data (courtesy of BASF [7])

Reducing Energy Consumption in a Formic Acid Plant of BASF

A main problem at the formic acid production is the undesired back-reaction of formic acid during distillation, which increases the specific energy consumption [7]. This is shown in the left diagram of Fig. 2.15, on the basis of measured values within a time interval of 6 days. BASF looked for process parameters, which may influence the back-reaction, in order to decrease operation costs (specific energy consumption). One of them is the molar ratio of water to methyl formate, both educts of the formic acid synthesis.

The diagram on the right shows the influence of the mentioned molar ratio to the rate of the unwanted back-reaction:

- Without data validation (raw data , black symbols), no influence is visible but only a cloud of data.
- Using validated values (grey symbols) a clear trend is visible, which means that reducing the molar ratio decreases the rate of back-reaction.

Both parameters could be correlated only by data validation. Due to these results the specific energy consumption can be reduced by 5 %. Data validation allows the most effective command variables for the control of a process to be determined. This study led to the discovery of which control variable had a dominant effect on the said rate of back-reaction, and consequently on the specific energy consumption.

Performance Monitoring at KKL Nuclear Power Plant

On-line implementation of validation software in the nuclear power plant (NPP) of Leibstadt – Switzerland (KKL) generated substantial benefits (two million USD per year) over the past 10 years. The priority of NPP operators is to run their plant as close as possible to the licensed reactor power in order to maximize the generator power. To meet this objective, plant operators must have the most reliable evaluation of the reactor power. The definition of this power is based on a heat balance using several

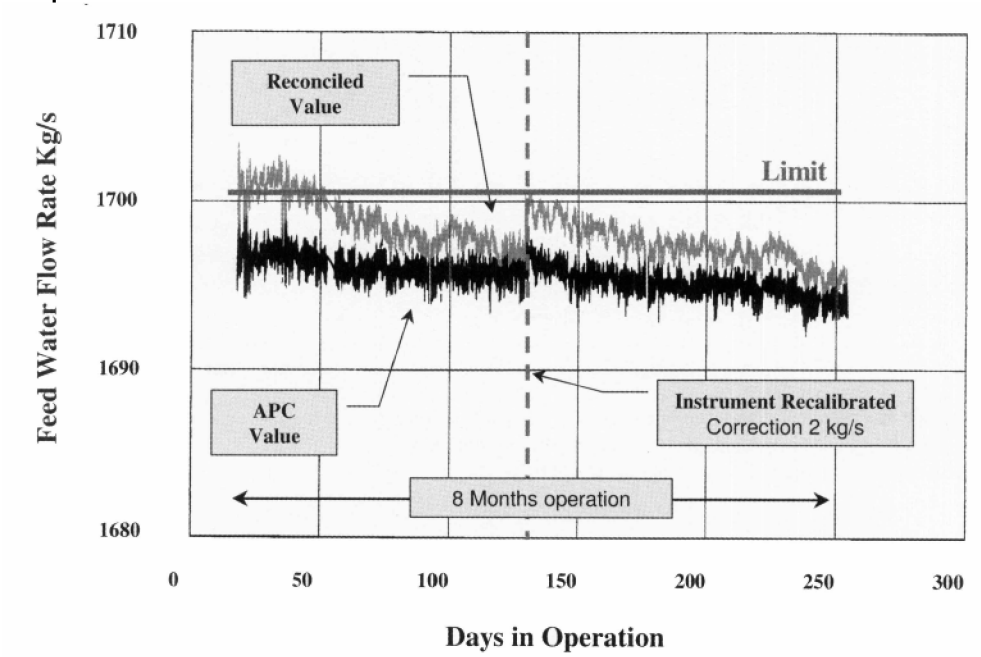


Figure 2.16 Operating closer to the limits: site feed water flow of the NPP Leibstadt (courtesy of Kernkraftwerk Leibstadt [8])

measured process parameters, among which the total feed water flow rate is the most critical value.

On-line implementation of validation software in the NPP of Leibstadt – Switzerland (KKL) has quantified the deviation between the actual and the measured feed water flow rate (see Fig. 2.16). In Fig. 2.16, only one recalibration is illustrated. This was used to convince the legal authorities about the reliability of the implemented validation technique. Validation results were also compared to test runs.

In agreement with the authorities in charge of safety of NPP, KKL nowadays recalibrates the measured flow rate based on the validated value, as soon as a deviation becomes significant. This enables the power plant to work close to its maximum capacity throughout the whole year (1145 MW). Prevention of losses due to heat balance errors increased the plant output by 5 MW. In addition, the use of this technology also made the annual heat cycle testing obsolete and significantly reduced the cost for mechanical and instrumentation maintenance [8].

Performance Monitoring of Refinery Units at LOR (Lindsey Oil Refinery), U.K.

On-line validation software is used at LOR for the performance monitoring of refinery units for several years. One set of applications is about the follow-up of fouling of the heat exchangers of several preheat trains. The main goal of the application is to determine the appropriate amount of anti-fouling product in order to maintain an adequate operation of the preheat trains and thus energy efficiency of the plant.

Another set of applications concerns the follow-up of furnaces and power plant boilers. The goal here is to determine with sufficient reliability and accuracy their energy efficiency. Any inappropriate operation can easily be detected and corrected when necessary.

Performance Monitoring of PE Plant at Gonfreville, France

The application enables any deviation within the instrumentation to be detected and provides guidance to the operators for the recalibration of the on-line analyzers. In addition, it ensures that the on-line soft sensors remain valid by counter-checking the quality of the instrumentation on which they rely.

2.5.2 On-line Production Accounting

2.5.2.1 Description and Benefits

This solution aims at providing a clear view of the production accounting, on a daily basis, of a whole industrial site: rigorous and automatic procedure for production accounting based on closed material balances. These material balances can be performed either:

- On a global mass balance basis: mass flow rates, in terms of tons entering and tons going out of each production unit, are reconciled to generate a coherent mass balance of the whole site. This approach is typically applied in refineries and covers the whole site including the tank farm.
- On a chemical compound basis: additional information is then required on the composition of the various streams and the reactions schemes. This approach is typically applied in chemical and petrochemical production plants.

Typical benefits are:

- Actual plant balances: closed balances are key elements as much for effective production accounting as for efficient performance monitoring.
- Decrease of unidentified losses and surpluses: abnormal conditions leading to losses and/or apparent surpluses are identified and can be corrected before they impact the economics of the plant.

Several real cases can be referred to, namely an adiponitrile plant and two refineries.

Production Accounting at ERE and Holborn Refineries

On-line validation software establishes the daily mass balance of the whole ERE refinery (BP refinery located at Lingen, Germany), covering about 150 tanks and about 50 production or blending units. Only a global mass balance (in tons) is made around each unit. The person in charge of the use (and maintenance) of the system spends about 30 minutes per day to generate all the validated reports and inputs for the production accounting. More recently the Holborn refinery in Germany has

installed a similar system, which also automatically detects abrupt changes in measured data identifying possible changes in operation or instrumentation failure.

Production Accounting at Butachimie, France

In this application the modeling includes compound balances of each main piece of equipment of an adiponitrile production facility. Reconciled compound balances are provided on a daily basis. All main chemical compounds as well as the catalysts used in the system are rigorously tracked all over the process unit.

2.6 Conclusion

Data reconciliation and validation is nowadays a mature technology. However it is often confused with flow sheeting and process simulation. Still, much has to be done to inform engineers and managers who have not learned about this technology during their studies. We have tried to convey the importance of this technology, and the very high diversity of applications and benefits that it can provide for the process industry.

References

1 *Belsim* VALI 4 User's Guide, Belsim, Belgium, 2005

2 BP-ERE, AN-ERE-01.pdf, available at: www.belsim.com, 2004

3 WACKER AN-Wacker-01.pdf, available at: www.belsim.com, 2002

4 *Delava P. Maréchal E. Vrielynck B. Kalitventzeff B.* Modeling of a Crude Oil Distillation Unit in Terms of Data Reconciliation with ASTM or TBP Curves as Direct Input. Application: Crude Oil Preheating Train, Proceedings of ESCAPE-9 conference, Budapest, May 31-June 2 1999, *Computers and Chemical Engineering* Suppl., (1999) pp. 17–20

5 APC Systems www.ipcos.be, 2005

6 *Amand T. Heyen G. Kalitventzeff B.* Plant Monitoring and Fault Detection: Synergy between Data Reconciliation and Principal Component Analysis. Computers and Chemical Engineering 25 (2001) p. 501–507

7 BASF, AN-BASF-02.pdf, available at: www.belsim.com, 2002

8 Kernkraftwerk Leibstadt, AN-KKL-01.pdf, available at: www.belsim.com, 1995