DATA RECONCILIATION

Georges Heyen
Department of Chemical Engineering, University of Liège, Belgium

Keywords: Measurement, redundancy, uncertainty, validation, error detection, error correction, data reconciliation, process operation, monitoring, state estimation, parameter identification, mathematical model.

Contents

Summary

Reliable process data, such as flow rates, compositions, temperatures, pressures and phase fractions, are the key to efficient operation of chemical plants. With the increasing use of computers in industry numerous data are acquired and used for on-line optimization and control. However, raw measurements are not accurate enough; they are affected by random or systematic errors, due to sensor drift, calibration errors, instrument malfunctions, leaks and so forth. Hence the measurements cannot satisfy exactly material and energy balances or other model constraints. The goal of data validation is to reconcile the contradictions between the measurements and their constraints, to estimate the true values of measured variables, to detect gross errors and solve for some unmeasured variables. Thus one can obtain the required process data with high accuracy and reliability, and generate consistent balances for accounting.

Algorithms used to correct random errors and allow closing process balances are

discussed, both for steady state and dynamic systems.

Practical applications are described, and the benefits of data validation are illustrated. Up to now steady-state data reconciliation and gross error detection began to be applied widely in industrial plants in 1980s. This technology is now a mature field with certain challenges remaining. The reconciliation for dynamic systems is an active development field; however, its on-line application to large industrial systems is still in its infancy.

1. Scope, Aims and Benefits of Data Reconciliation

Nowadays, industrial processes are more and more complex and difficult to master. They process large quantities of valuable goods, and thus should be run efficiently to avoid wasting raw materials and ensure a high product quality. The operation of many chemical plants involves also potentially dangerous operations: strict process monitoring is necessary to avoid unsafe operating conditions that could lead to fire, explosion or release of toxic components in the environment. The size of the equipment, the value of products they transform, the requirements for safety thus dictate that processes should be monitored and controlled efficiently.

1.1. Importance of Measurements for Process Monitoring

Efficient and safe plant operation can only be achieved if the operators are able to monitor all key process parameters. Instrumentation is used to measure many process variables, like temperatures, pressures, flow rates, compositions or other product properties. Measuring these variables should allow the operators to verify that the equipment is operating according to the design. Without good measurements, the operators would be blind: similarly, to drive a car, one needs to see the road, locate the car position with respect to obstacles, and know its speed. When visibility is poor, the safe decision is to reduce speed, or even to stop the car. In the same way, when measurements do not allow assessing a plant operating condition, it cannot be run safely at maximal efficiency.

In practice, direct measurements do not provide always all the required information. What is needed is an estimation of some performance indicators. These are the variables that either contribute to the process economy (e.g. the yield of an operation), or are linked to the equipment quality (e.g. fouling in a heat exchanger or activity of a catalyst), to safety limits (e.g. departure from detonation limit) or to environmental considerations (e.g. amount of pollutant emissions). Most performance parameters are not directly measured, and are evaluated by a calculation based on one or several measured values.

For instance, a car driver is interested in knowing how much fuel is left in the tank. What is measured is the level in the tank. Thus some knowledge about the physical plant (here the shape and size of the tank) must be known to calculate the useful value (amount of fuel) from the raw measurement. In many cases, several independent measurements must be combined to assess the value of some process variable (e.g. the mileage for a vehicle is computed from the fuel consumption (based on the variation of the fuel tank level, or from a direct flow rate measurement) and from the distance traveled, obtained from an odometer).

However some difficulties arise when one considers experimental errors.

1.2. Sources of Experimental Errors

Experimental data is always affected by experimental errors. No sensor can be built that is absolutely exact and accurate. Besides uncertainty linked to the measuring device, errors can also arise from sampling or positioning the sensors (e.g. measuring local properties in a material that is not homogeneous), from inappropriate calibration, from transcription or reporting errors (during signal conversion for instance).

One should make a distinction between permanent bias or systematic errors, and random deviations. The overall error results from summing both contributions. Systematic errors are related to deficient instrumentation or inexact calibration: an example would be using erroneous weights or a chronometer that runs late. No matter how careful the measurement is carried out, the error will remain undetected, even if the measurement is repeated. The only way out is to compare the measurement with an independent assessment using a different sensor, and such a procedure allows then to calibrate properly the defective sensor. In other respects random errors are due to a multiplicity of causes, and may result from fluctuations in sampling or external perturbations (e.g. variation of atmospheric pressure, voltage fluctuations for electric instruments). They can be detected by repeating the measurement, and noticing that the outcome is different.

Measurement error is the sum of both contributions: systematic and random errors.

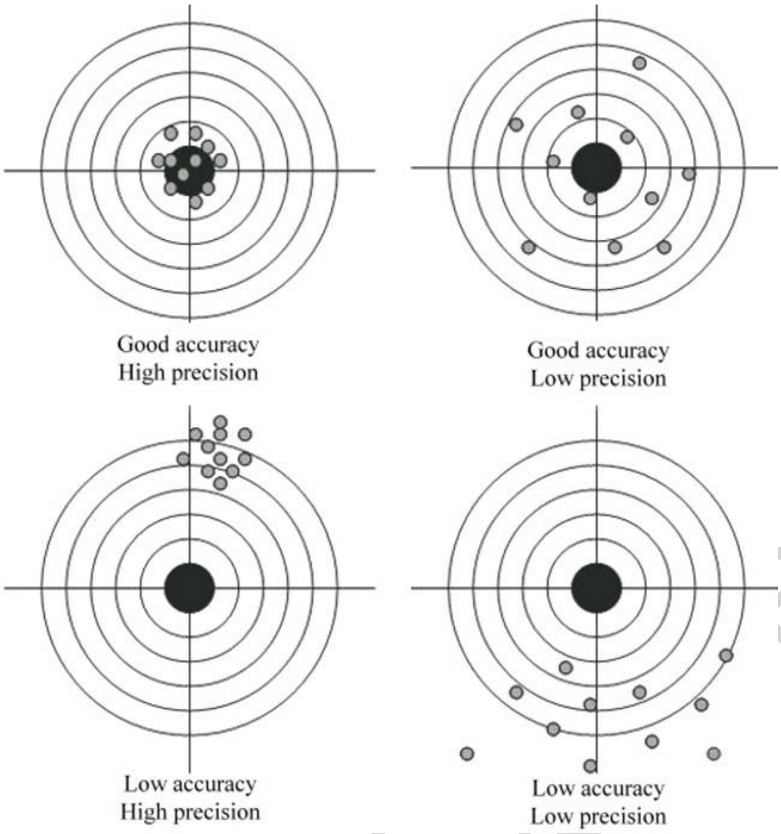

Figure 1: Comparison between precision and accuracy

Figure 1 allows illustrating the difference between accuracy and precision, by comparing the measurement process with shooting at a target. Accuracy represents the systematic departure of the measurement with respect to the true value (usually unknown). For the shooting analogy, this could be corrected by adjusting the sight. For a measurement, inaccuracy results from instrument bias and improper calibration. Precision, for the shooting analogy, is related to the spread of the bullets on the target. Low precision results from imperfect instrumentation and variation in operating procedures. Precision is linked to the repeatability of the measurement: a clock can be very precise (exactly 3600 ticks every hour) and give systematically the wrong time.
Repeating measurements allows estimating their precision, by assessing the spread of their distribution around the average value (assuming that the measured variable remains constant during the measurement process). Thus we can expect that measurement redundancy is a way to improve the quality and reliability of the measurement results.

Random errors that always affect any measurement also propagate in the estimation of performance parameters. When redundant measurements are available, they allow the estimation of the performance parameters based on several independent data sets; this provides different estimates, which may lead to confusion if not properly interpreted. Data validation is the method applied to properly exploit measurement redundancy in order to improve the assessment of process variables.

1.3. How to Achieve Measurement Redundancy

Measurement redundancy can be obtained in several ways.

A first approach is to repeat several times the same measurement using the same sensor. This is called temporal redundancy. By taking the average of the measurements, one can expect to decrease the uncertainty arising from random errors. In fact, according to the theory of probability, the variance of the mean value $\sigma_{\overline{X}}^2$ is proportional to the inverse of the number of measurements N :

$$\sigma_{\overline{X}}^2 = \frac{\sigma^2}{N} \qquad (1)$$

In a process whose variables are likely to fluctuate with time, measurement redundancy can also be achieved by installing multiple sensors in order to obtain several simultaneous measurements of the same variable(s). This procedure allows not only to reduce uncertainty by averaging the measured values, but also to detect gross errors resulting from sensor failures. Such an approach is used for some safety critical measurements, coupled with comparison software that implements a voting scheme in case contradictory measurements are obtained. However this type of sensor redundancy is costly and not applied systematically to all process variables.

Redundancy can also be achieved by using several measurements combinations and a process model to estimate the required process variables. To explain this method, we need first to evaluate the uncertainty of an estimate when several measurements are needed to assess a variable value.

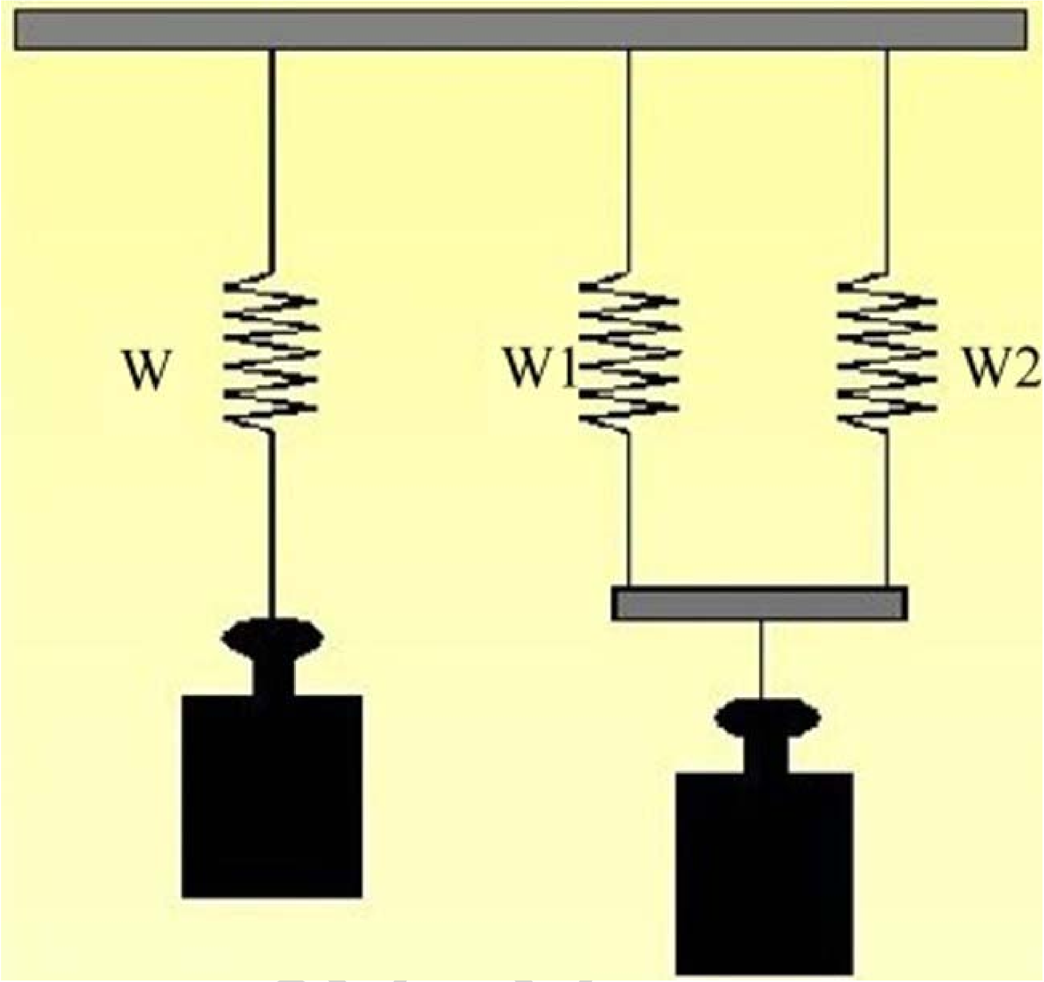

Figure 2: Using several measurements to assess a value

An example is shown in Figure 2, where a weight is evaluated by summing two measurements. The variance of the estimate is obtained by summing the variance of both measurements.

$$\begin{aligned} W &= W_1 + W_2 \\ \sigma_W^2 &= \sigma_{W_1}^2 + \sigma_{W_2}^2 \end{aligned} \qquad (2)$$

For the more complex set up shown in Figure 3, we need to use a model of the set up, and use the equilibrium condition to obtain the value of the weight W from the measurements W_1 and W_3:

$$\begin{aligned} W_1 &= 2(W + W_3) \\ W &= 0.5W_1 - W_3 \\ \sigma_W^2 &= 0.25\sigma_{W_1}^2 + \sigma_{W_3}^2 \end{aligned} \qquad (3)$$

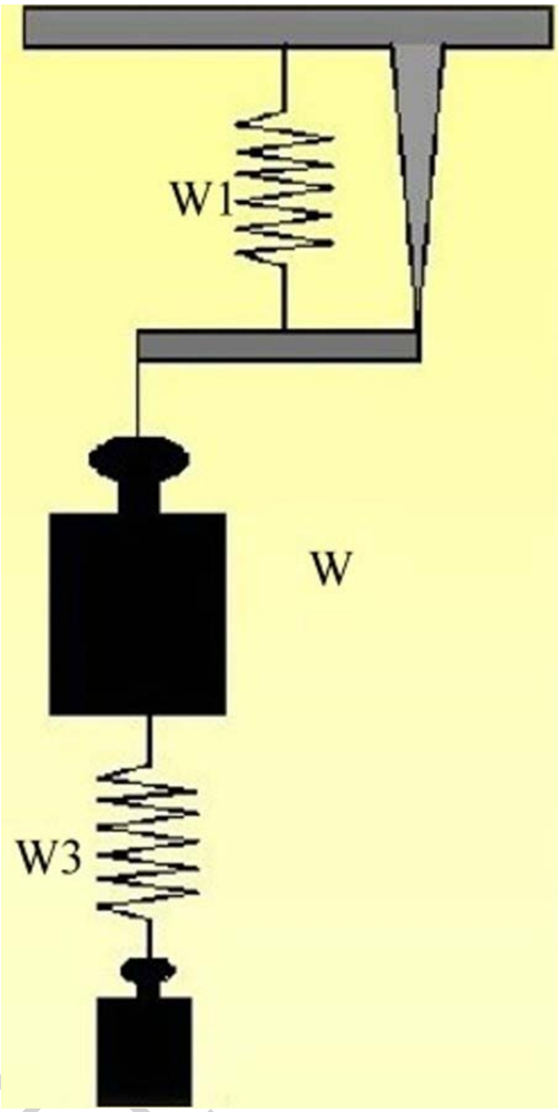

Figure 3: Indirect evaluation by combining several measurements

In general, if a variable W can be calculated using a model f and the value of several independent measured variables x_i, the variance of its estimate will be related to the measurement variances, using the following relationship obtained by linearizing the model f:

$$W = f\left(x_i\right) \qquad i = 1 \ldots n$$
$$\sigma_W^2 = \sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i} \right)^2 \sigma_{x_i}^2 \tag{4}$$

The variables appearing in the model need not to be of the same type. In fact, some process variables can be estimated in several independent ways. As an example, let us consider the case shown in Figure 4.

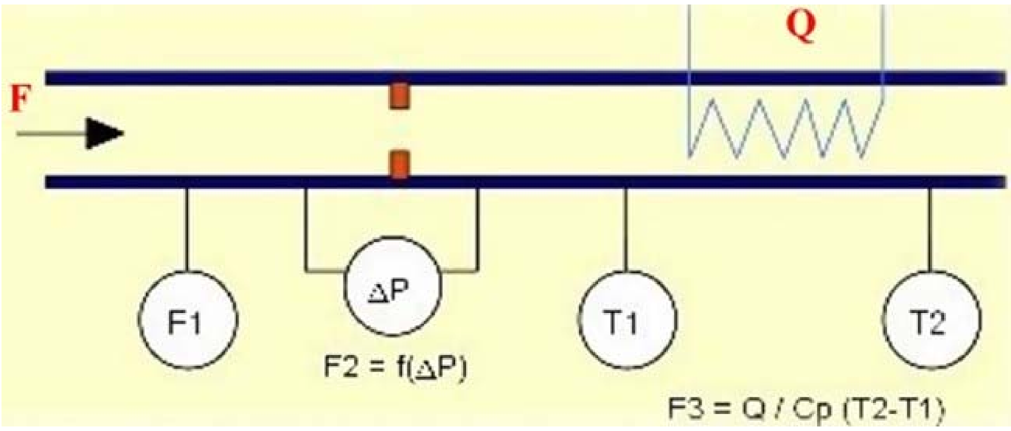

Figure 4: Multiple evaluations by combining several techniques

A flowrate in a pipe can be directly measured as $F1$ using a flow meter (e.g. using Doppler-effect). The flowrate can also be estimated by measuring the pressure drop through an orifice, which will provide estimate $F2$. It can also be obtained from an energy balance, for instance by heating the fluid using electrical power and measuring the temperature increase. If the fluid specific heat C_p is known, the flowrate estimate $F3$ will be related to the power dissipated Q and the temperature increase by:

$$F3 = \frac{Q}{C_p\left(T_2 - T_1\right)} \tag{5}$$

A data validation algorithm will provide a way to merge those independent estimates and pool their variances in order to provide a consistent value of the flowrate.

-
-
-

Bibliography

Albuquerque J.S., L.T. Biegler, (1996) *Data reconciliation and Gross-Error Detection for Dynamic Systems*, AIChE Journal, (42) 2841-2856 [good introduction paper on error detection for dynamic systems]

Bagajewicz M.J., (2001) Design and Retrofit of Sensor Networks in Process Plants, AIChE J., 43(9), 2300-2306 [strategy to design or upgrade measurement systems]

Crowe C.M., (1989) *Observability and redundancy of process data for steady state reconciliation*, Chem. Eng. Sci. 44, 2909-2917 [variable classification and redundancy checks]

Heyen G. , Kalitventzeff B. (2006) *Process monitoring and data reconciliation*, in Computer Aided Process and Product Engineering , Puigjaner, L. Heyen, G. (eds.), Wiley VCH [review article, including considerations on the rational design of sensor networks]

Kalitventzeff B., Heyen G., Mateus M., (2006) *Data Validatio: a Technology for Intelligent Manufacturing*, in Computer Aided Process and Product Engineering , Puigjaner, L. Heyen, G. (eds.), Wiley VCH [focus on industrial applications and benefits for process monitoring]

Kalman R.E., (1960) A new Approach to linear Filtering and Prediction Problems, Trans. ASME J. Basic Eng. 82D, 35-45 [seminal paper on model based observation for linear dynamic systems]

Liebman M.J., T.F. Edgar, L.S. Lasdon, (1992) *Efficient Data Reconciliation and Estimation for Dynamic Processes using nonlinear Programming Techniques*, Computers and Chemical Engineering, 16, 963-986 [seminal paper on dynamic data reconciliation]

Musch H., List T., Dempf D., Heyen G., (2004) *Online estimation of Reactor Key Performance Indicators : An Industrial Case Study* ,in, *Computer-Aided Chemical Engineering volume 18*, Elsevier Science [description of an industrial application for online estimation of reactor performance]

Narasimhan S., C. Jordache, (2000) *Data Reconciliation and gross Error Detection, an intelligent use of Process Data*, Gulf Publishing Company [good introduction, clearly written with examples]

Romagnoli J.A., M.C. Sanchez, (2000) *Data Processing and Reconciliation for chemical Process Operations*, Academic Press [thorough theoretical and mathematical developments]

Vaclavek V., (1969) *Studies on System Engineering : Optimal choice of the balance measurements in complicated chemical engineering systems*, Chem. Eng. Sci. 24, 947-955 [a pioneering paper on the subject]

Veverka, V.V., Madron, F., (1996) *Material and energy balancing in the process industries. From microscopic balances to large plants, Computer-Aided Chemical Engineering volume 7*, Elsevier Science [clear introduction to the topic]

Additional information to be found on the web sites of some commercial software publishers:

Belsim VALI 4 : http://www.belsim.com/Products_main.htm

DATACON : http://www.simsci-esscor.com/us/eng/simsciProducts/productlist/datacon/DATACON.htm

SIGMAFINE : http://www.osisoft.com/Sigmafine_Brochure_screen.pdf

DATREC : http://www.technip.com/english/pdf/DATREC.pdf

RECONCILER : http://www.ris-resolution.com/reconciliation.shtml

ADVISOR : http://www.aspentech.com/brochures/Aspen_Advisor.pdf

Biographical Sketch

Georges Heyen is a professor in Chemical Engineering at the University of Liège, Belgium, with teaching, research and consultancy interests in process systems engineering. He teaches courses ranging from applied chemical thermodynamics to modeling and plant design.

He took part in several international projects financed by the European Union to carry research and development on process optimization tools and methods, in collaboration with major European universities and research centers.

He is one of the founders of the company Belsim, which is active in software development for process modeling and optimization, especially using online data validation techniques.

Georges Heyen had his education as under:
Ph.D: Université de Liège, Chemical Engineering, 1983
M.S.: University of Texas, Austin, Chemical Engineering, 1975
B.S.: Université de Liège, Chemical Engineering, 1973