CHEMICAL ENGINEERING TRANSACTIONS

A publication of AIDIC
The Italian Association of Chemical Engineering
www.aidic.it/cet

VOL. 39, 2014

Guest Editors: Petar Sabev Varbanov, Jiří Jaromír Klemeš, Peng Yen Liew, Jun Yow Yong

ISBN 978-88-95608-30-3; **ISSN** 2283-9216

DOI: 10.3303/CET1439280

Application of Interval Analysis to the Reconciliation of Process Data when Models Subject to Uncertainties Are Used

Marco Vocciante[a], Valter Mantelli[b], Nicoletta Aloi[b], Andrea P. Reverberi[c], Vincenzo G. Dovì*[c]

[a]Dipartimento di Ingegneria Civile, Chimica e Ambientale, Università di Genova, Via Montallegro 1, I-16145 Genova
[b]IPLOM SpA, Via Carlo Navone 3B, I-16012 Busalla (Genova)
[c]Dipartimento di Chimica e Chimica Industriale, Università di Genova, Via Dodecaneso 31, I-16146 Genova
dovi@istic.unige.it

Temperatures and flow rates are the most frequently measured variables in industrial processes. They generally form the basis of data reconciliation based on mass and energy balances.
While flow rate data rectification is routinely carried out using rigorous linear models of mass balances, energy balances, necessary for the reconciliation of temperature values, introduce two main difficulties. Indeed the presence of the product of two variables that need reconciling (temperature and flow rate) changes the original linear equations into a system of bilinear equations.
Additionally energy balances are subject to modelling errors due to the presence of parameters (specific heats, latent heats, heats of reaction). The uncertainties on these parameters can affect the reliability of the data reconciliation considerably.
It is shown in this article that interval analysis can provide a useful tool for reducing the sensitivity of the reconstructed values of the process variables. An important simple case is examined for illustration purposes.

1. Introduction

The reconciliation of process data is today a widespread practice in most industrial plants for the achievement of optimal production targets (Sarabia et al., 2012), quality standards (Fabiano et al., 2012) and safety conditions (Fabiano et al., 2013).
The reconciliation of flow rate measurements is routinely carried out in most plants using well established algorithms based on linear algebra techniques (Mah and Tamhane, 1982). A robust linear approximation can be applied also if the mass balances of single components are considered (Crowe et al., 1982). Typically in these algorithms a maximum likelihood function is optimised subject to the linear constraints representing the mass balances of the model.
The introduction of additional constraints related to further equations of the model could in principle increase the information available for a more accurate rectification of the variables measured, but the uncertainties contained in them make their use problematic, because experimental errors and model discrepancies could cumulate to produce largely inaccurate reconstructions of the reconciled variables. However, a fundamental difference should be made between energy balances and the remaining closure relations that completely define the model. Indeed in the former case a rigorous conservation law is used, even if it can contain approximately known parameters (e.g. the temperature coefficients of specific heats), whereas closure (constitutive) relations may contain structurally unreliable assumptions (e.g. the assumption of thermodynamic equilibrium). This is the reason why only energy balances are occasionally considered in the overall reconciliation procedure.
However, even energy balances may introduce computational difficulties and statistical biases.

Please cite this article as: Vocciante M., Mantelli V., Aloi N., Reverberi A.P., Dovì V.G., 2014, Application of interval analysis to the reconciliation of process data when models subject to uncertainties are used, Chemical Engineering Transactions, 39, 1675-1680 DOI:10.3303/CET1439280

The computational difficulties are due to the presence of non linearity in the resulting system of equations that represent the overall model. For instance, if the energy flux through a section is given by $\dot{m}c_pT$ where $\dot{m}$ is the flow rate, c_p is the specific heat and T is the temperature, the energy balance containing this term is bilinear with respect to $\dot{m}$ and T if c_p is constant, or generally non linear if c_p is allowed to vary with the temperature.
This difficulty is frequently dealt with using two different strategies. If c_p can be assumed constant (for instance using the value of c_p corresponding to the measured value of the temperature), the overall system of mass and energy balances can be split into two separate subsystems of equations, the first of which contains mass balances only and is used for reconstructing reconciled values of the flow rates. The second subsystem contains the energy balances, which can also be considered linear if the specific heats are constant and the reconciled values of the flow rates are used for $\dot{m}$. This approach is not satisfactory from a statistical point of view because the original maximum likelihood is arbitrarily split into two independent terms, but can turn out to be satisfactory for practical purposes, provided the redundancy of measurements of the flow rates is sufficient to reconstruct their values within reasonable confidence intervals using mass balances only. Generally splitting mass and energy balance equations proves more reliable when the exchange of mass and energy takes place in different units, such as the mixing of streams under isothermal conditions or the transfer of energy in heat exchangers without merging of the streams involved.
Alternatively a rigorous optimisation with non linear constraints is applied, which is the only option if the specific heats cannot be assumed constant and must be allowed to vary with the temperature. This procedure can be computationally burdensome (and consequently not suitable for real time evaluations) and may occasionally lead to suboptimal solutions.
A more serious inconvenience brought about by the use of energy balances in the general reconciliation procedure is the presence of parameters which are only approximately known, such as specific heats, latent heats, heats of reaction (Solisio et al., 2012), as well as thermal emissivity of flames (Palazzi and Fabiano, 2012). The uncertainties on these parameters can considerably affect the reliability of the adjusted data. Indeed, if both c_p and T are error prone (either because of poor modelling or due to limited experimental accuracy) the resulting product $\dot{m}c_pT$ can present large deviations even if $\dot{m}$ has been estimated accurately.
A general framework for dealing with uncertain models has been introduced by Maquin et al. (2000) using penalty functions and introducing suitable weights for each constraint as a function of the uncertainties contained in the model. Being based on nonlinear optimization techniques their method can prove algorithmically burdensome and may locate local (non globally optimal) solutions. Furthermore the method for estimating the weights to be attached to each energy balance due to the uncertainties contained in the thermal parameters may not be straightforward.
Ragot and Maquin (2004) have also developed a general strategy that combines both experimental errors and model approximations, but by merging the two sources of inaccuracies they disregard the information contained in the statistical properties of the measurements.
Interval analysis may provide a useful frame of reference and powerful computational tools for dealing with uncertainties, but it can require a prohibitively high computing effort, unless suitable techniques for dealing with particular classes of problems are adopted.
Benothman et al. (2007) have described a general strategy for the use of interval analysis in data analysis based on linear systems, without considering the special characteristics of reconciliation problems.
Similarly, interval analysis techniques have been used by Dubois et al. (2013) for the adjustments of mass flow rates whose balances give rise to fuzzy linear equations. However, the statistical information provided by the measurements is not fully taken advantage of and the simultaneous presence of balances based on rigorous and approximate models is not considered in their approach.
The main advance of this article beyond the present state-of-art provided by the articles mentioned before is precisely the development of a reconciliation strategy that employs interval analysis techniques adapted to the particular class of process measurements reconciliation and uses the statistical information contained in the data. The method developed reduces the overall computational effort and can be employed for real time analysis of complex process configurations.

2. Application of Interval Analysis to the Reconciliation of Process Measurements

Typically a reconciliation of process data implies the optimization of a suitable objective function (generally the maximization of the likelihood function based on the available measurements) subject to the constraints provided by the balance equations.
If temperatures are supposed to be measured with greater accuracy than flow rates (especially component flow rates), the balance equations can be written in the form

$$\sum_i a_{ij} \Delta L_i = b_j \qquad j=1,..., n_1 \tag{1}$$

$$\sum_i a_{ik}^* \Delta L_i = b_k^* \qquad k=1,..., n_2 \tag{2}$$

where the asterisk means that the corresponding coefficient is not perfectly determined. Typically equations $k=1,..., n_2$ include enthalpy balances containing inaccurate thermal parameters. Additionally the uncertainty may be extended to include measurement errors of temperatures if they cannot be assumed to be negligible. Equations $j=1,..., n_1$ include all the mass balances (in this case $a_{ij} = \pm 1$), as well as enthalpy balances with accurately determined thermal parameters.
The solution of systems of linear equation with coefficients varying inside given intervals implies determining suitable enclosures including values of the unknown variables ΔL_i that satisfy each equation.
Determining the best enclosure to the solution set is an NP-hard problem, which makes its application to complex process configurations too computationally expensive for all practical purposes.
Several methods have been developed in the literature to detect feasible methods for the determination of enclosures less general than those theoretically possible, but capable of including the solutions of interest for the particular problem under consideration.
To illustrate the application of interval analysis algorithms to the class of problems under consideration, let us define first
$[\boldsymbol{x}]_r = \left\{ \boldsymbol{y} \in \Re^n : \|\boldsymbol{y} - \boldsymbol{x}\|_\infty \le r \right\}$, i.e. the interval set centred at $\boldsymbol{x}$ with radius r. The infinite norm $\|\bullet\|_\infty$ implies that no component of $(\boldsymbol{y} - \boldsymbol{x})$ can exceed r. Typically inequalities based on infinite norms are verified using minimax algorithms based on linear programming.
This definition makes it possible to verify whether the corrections that could be introduced considering enthalpy balances do improve the solution obtained neglecting them.
To this purpose the following linear programming problem can be considered:

$$\sum_i \eta_i = \min \tag{3}$$

$$\sum_i \vec{a}_{ij} \left(\hat{x}_i + \eta_i \right) = \vec{b}_j \tag{4}$$

$$\sum_i \vec{a}_{ij} \left(\hat{x}_i - \eta_i \right) = \overleftarrow{b}_j \tag{5}$$

where the arrow pointing left indicates the lower bound and the arrow pointing right the upper bound. The hatted variables indicate the values obtained through reconciliation of the mass balances only.
If the value of η_i is much larger than the standard deviation of the corresponding measurements, the enthalpy balances are not further considered. The remaining enthalpy balances, together with the mass balances are used for the final reconciliation step, which consists in the solution of a system of linear equations, some of which contain fuzzy coefficients.
To this purpose the iterative procedure originally proposed by Krawczyk (1984) is a useful basis. Given any differentiable function *F(**x**)*, the Krawczyk set associated with $[\boldsymbol{x}]_r$ is defined as

$$\left[F[\boldsymbol{x}]_r \right] = \boldsymbol{x} - DF(\boldsymbol{x})^{-1} F(\boldsymbol{x}) + \left[I - DF(\boldsymbol{x})^{-1} DF\left([\boldsymbol{x}]_r\right) \right]\left([\boldsymbol{x}]_r - \boldsymbol{x}\right) \tag{6}$$

Starting with an assigned interval for each $[\boldsymbol{x}]_r$ (provided the statistical properties of the estimates), the iterative procedure is repeated until successive intervals coincide according to a specified convergence criterion. Alternatively the method recently developed by Dymova et al. (2010) can be employed.
Thus the various parts of the algorithm have been analysed. In the next section a general outline of the algorithm will be provided.

3. Outline of the overall algorithm

The overall algorithm can now be summarised as follows:

a) In the first step the data are reconciled neglecting the equations of the model that are subject to uncertainty. Typically the reconciliation procedure includes only mass balances and is carried out using a traditional maximum likelihood approach.
b) In the second step a measure of the uncertainty in the rectified data introduced by the models included in the remaining equations is evaluated using linear programming.
c) If the value of the uncertainty is much larger than experimental errors, the equations containing the streams affected by them are not further considered
d) The variables including uncertainties are replaced by intervals and the reconciliation procedure is applied to the whole set of equations (except those dropped in step 3) using an iterative method capable of estimating the resulting interval of the variables to be reconciled.

Although the solution of the resulting linear interval systems of equations remains a NP-hard problem, it is computationally feasible (as opposed to computationally intractable) in the sense of Lakeyev (Kreinovich et al., 1996), depending on the number of fuzzy coefficients.

4. Application of the algorithm to an industrial case

The algorithm developed in the previous sections has been applied to the example shown in Figure 1. Furnace oil (stream 1) and fuel gas (stream 2) are used to heat the crude oil charge in a topping furnace. Both streams are measured with known (but different) accuracies. The flow rate of the air (stream 3) fed to the fired heater after exiting a preheater in counter current exchange with the flue gases of the combustion (stream 4) is also measured. The temperatures of the streams entering/exiting the preheater are also measured.

The flow rate of the flue gas can be evaluated using the stoichiometric relations based on the composition of the furnace oil and of the fuel gas, as well as on the amount of excess air. The uncertainty on its value can be obtained by appropriately

This information makes it possible to write both mass and enthalpy balances around the preheater.

While the use of enthalpy balances has been successfully used in several applications (Jiang et al., 2013), it is shown that in this case it can give rise to large deviations, unless interval estimation techniques rather than traditional point estimation methods are used.

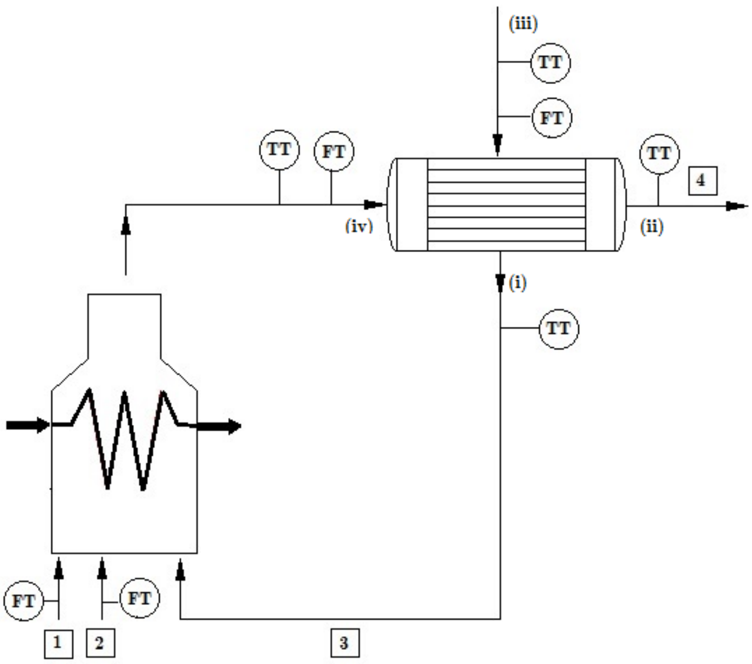

Figure 1: Mass and enthalpy balances around the preheater

For illustration purposes the inaccuracy in the measurement of the air flow rate is considered negligible with respect to the measurement errors of both fuel streams and the (unknown) discrepancies between theoretical and measured values of the two fuel streams are indicated by ΔL and ΔG respectively.
It can be shown that the mass balance can be written as: $a_{11}\Delta L + a_{12}\Delta G = b_1$, where the coefficients a_{11}, a_{12} and b_1 can be evaluated using stoichiometric relations based on the information on fuel gas components and on the fraction of C,H,N,S in the furnace oil.
A similar relation for the enthalpy balance around the preheater can be obtained: $a_{21}\Delta L + a_{22}\Delta G = b_2$. However, in this case the coefficients a_{21}, a_{22} and b_2 contain thermal parameters that may be not perfectly known. In other terms the coefficients are to be regarded as fuzzy variables.
Using the notation introduced in the previous section the overall system can be written as:

$$a_{11}\Delta L + a_{12}\Delta G = b_1 \tag{7}$$

$$[a_{21}]\Delta L + [a_{22}]\Delta G = [b_2] \tag{8}$$

The simplifications brought about by the assumption on higher accuracies in the measurement of the air stream reduce the reconciliation problem to the solution of the linear system:

$$\Delta L = \frac{b_1[a_{22}] - [b_2]a_{12}}{a_{11}[a_{22}] - a_{12}[a_{21}]} \tag{9}$$

$$\Delta G = \frac{[b_2]a_{11} - b_1[a_{21}]}{a_{11}[a_{22}] - a_{12}[a_{21}]} \tag{10}$$

Allowing b_2 to vary in a ±10 % relative interval causes the adjustments to vary in a 200 % interval as shown in Figure 2.
The presence of large deviations in the reconstructed flow rates is an indication that the use of enthalpy balances reduces the reliability of the reconciled values. Dropping the corresponding equation and minimising the resulting likelihood provides the solution based on the mass balance only:

$$\Delta L = b_1 \frac{a_{11}\sigma_L^2}{a_{11}^2\sigma_L^2 + a_{12}^2\sigma_G^2} \tag{11}$$

$$\Delta G = b_1 \frac{a_{12}\sigma_G^2}{a_{11}^2\sigma_L^2 + a_{12}^2\sigma_G^2} \tag{12}$$

where σ_L and σ_G are the variances of the measurement errors of the two fuel streams. The resulting values of $|\Delta L|$ and $|\Delta G|$ are larger than the corresponding values if the enthalpy balance with a perfectly

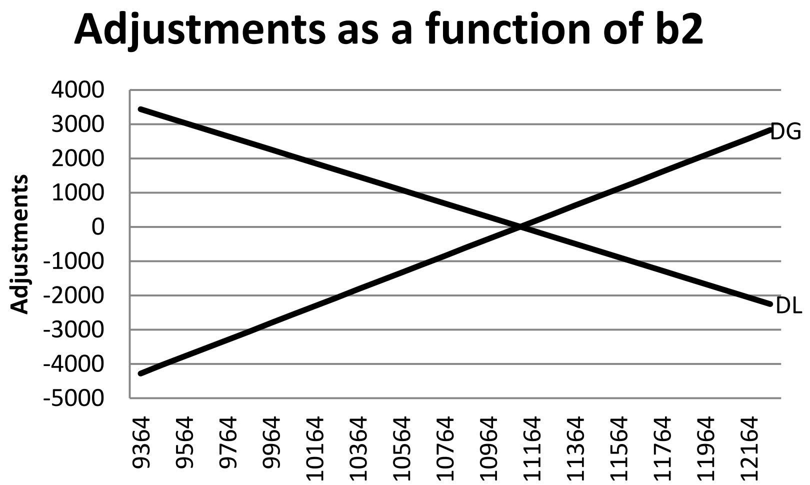

Figure 2: Adjustment brought about by simultaneous mass and enthalpy balances

determined parameter b_2 is used, but they are smaller if the uncertainties on the value of b_2 exceed 2 %. Depending on the accuracy of the parameter either strategy will be selected, as outlined before.

5. Conclusions

The use of enthalpy balances can introduce a severe bias in the reconciliation of flow rates whenever the thermal parameters of the models contained in them are approximate. In this case mass balances only are to be used. In this article a general strategy to assess the convenience of using enthalpy balances in the reconciliation of flow rates has been described. The resulting algorithm, based on interval analysis, provides a general framework for the selection of the equations to be included in the reconciliation procedure as well as a numerical method that takes advantage of the particular structure of the overall model to reduce the computational burden generally associated with interval analysis procedures.
Its real applicability in terms of computational feasibility depends on the general process configuration and on the number of equations subject to uncertainties and has to be decided on case-by-case basis.

Acknowledgements

The financial support from the EC FP7 project EFENIS (ENER/FP7/296003) is gratefully acknowledged.

References

Benothman K., Maquin D., Ragot J., Benrejeb M., 2007, Data validation and diagnosis using interval analysis, Int. J. of Sciences and Techniques of Automatic Control & Computer Engineering, 1, 2, 136−154.

Crowe C. M., Garcia Campos Y. A., Hrymak A., 1983, Reconciliation of process flow rates by matrix projection. Part I: Linear case, AIChE J, 29, 6, 881-888.

Dubois D., Fargier H., Guyonnet D., 2013, Data Reconciliation under Fuzzy Constraints in Material Flow Analysis, 8th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT 2013) DOI:10:2991/eusflat.2013.4

Dymova L., Pilarek M., Wyrzykowski R., 2010, Solving the systems of interval linear equations with use of modified interval dividing procedure, Lecture Notes in Computer Science 2010, 6068.

Fabiano B., Reverberi A.P., Del Borghi A., Dovi V.G., 2012, Biodiesel production via transesterification: Process safety insights from kinetic modelling, Theoretical Theor. Found. Chemical Engineering, 46, 673-680.

Fabiano B., Currò F., Reverberi A.P., Palazzi E., 2013, Coal dust emissions: From environmental control to risk minimization by underground transport. An applicative case-study, Proc. Safety and Envir. Protection, DOI:10.1016/j.psep.2013.01.002.

Jiang X., Liu P., Li Z., 2013, A Data Reconciliation Based Approach to Accuracy Enhancement of Operational Data in Power Plants, Chemical Engineering Transactions., 35, 1213-1218, DOI: 10.3303/CET1335202

Krawczyk R., 1984, Interval Iteration for Including a Set of Solutions, Computing, 32, 13-31.

Kreinovich V., Lakeyev A.V., Rohn J., 1996, Computational complexity of interval algebraic problems: some are feasible and some are computationally intractable: a survey, Eds. Alefeld G,. Frommer A., Scientific Computing and Validated Numerics, Akademie-Verlag, Berlin, 293–306.

Mah R.S.H., Tamhane A.C., 1982, Detection of gross errors in process data, AIChE J., 28, 5, 828–830.

Maquin D., Adrot O., Ragot J., 2000, Data reconciliation with uncertain models, ISA Transactions ®, 39, 35-45.

Palazzi E., Fabiano B., 2012, Analytical modelling of hydrocarbon pool fires: Conservative evaluation of flame temperature and thermal power, Process Safety and Environmental Protection 90, 2, 121-128.

Ragot J., Maquin D., 2004, Reformulation of Data Reconciliation Problem with Unknown-but-Bounded Errors, Ind. Eng. Chem. Res., 43, 1530-1536.

Sarabia D., De Prada C., Gómez E., Gutierrez G., Cristea S., Sola J.M., Gonzalez, R., 2012, Data reconciliation and optimal management of hydrogen networks in a petrol refinery, Control Engineering Practice, 20, 4, 343–354.

Solisio C., Reverberi A.P., Del Borghi A., Dovi' V.G., 2012, Inverse estimation of temperature profiles in landfills using heat recovery fluids measurements Int. J. of Applied Mathematics, Vol. 2012, Article n. 747410, DOI: 10.1155/2012/747410

Sarabia D., De Prada C., Gómez E., Gutierrez G., Cristea S., Sola J.M., Gonzalez, R., 2012, Data reconciliation and optimal management of hydrogen networks in a petrol refinery, Control Engineering Practice, 20(4), 343–354.