WS - PEMS: Using data reconciliation for predictive emission monitoring systems

Joachim Weel Rosbo* and Peter Jul-Rasmussen*

**Weel & Sandvig, Kgs. Lyngby, 2800, Denmark*
(jwr@weel-sandvig.dk, pjr@weel-sandvig.dk)

Keywords: Predictive Emission Monitoring Systems (PEMS), data reconciliation, gross error detection, observability, redundancy, soft constraints, SLSQP.

ABSTRACT

Background

Predictive Emission Monitoring Systems (PEMS) estimate pollutant emissions from a plant by coupling the available sensor measurements to a mechanistic model of the emission source, in which mass and energy balances connect a network of plant units (Knopf, 2011). The reliability of the predicted emissions is therefore dictated by the quality of the input measurements. Two classes of measurement error must be distinguished: random errors, modelled as zero-mean Gaussian noise, and gross errors, which are large, non-random deviations caused by sensor malfunction, miscalibration, or operator mistakes (Narasimhan and Jordache, 2000). Random errors are routinely handled by data reconciliation, which adjusts the raw measurements so that the physical balances are satisfied. Gross errors, however, violate the underlying Gaussian assumption and, if left untreated, are propagated by the reconciliation optimisation, distorting every reconciled variable.

Consider a steady-state process described by a vector of process variables $x \in \mathbb{R}^n$, which must satisfy the (in general nonlinear) constraints

$$g(x) = 0, \quad g : \mathbb{R}^n \to \mathbb{R}^m. \tag{1}$$

The process variables split into a measured part $x_m \in \mathbb{R}^{n_m}$, sensed by instruments with raw values $y \in \mathbb{R}^{n_m}$ and known covariance Q, and an unmeasured part $x_u \in \mathbb{R}^{n_u}$ that must be inferred from the model. Linearising g around the operating point and partitioning the Jacobian as $J = (J_m, J_u)$, the unmeasured variables can be classified as follows:

- an unmeasured variable is *observable* if it can be uniquely determined from the measurements and the constraints, which holds globally when

$$rank(J_u) = n_u. \tag{2}$$

- a measured variable is *redundant* if it can be recovered from the remaining variables and the constraints when its sensor is removed. The number of redundant pieces of information, the *degree of redundancy*, is

$$\nu = m - rank(J_u). \tag{3}$$

Redundancy makes reconciliation meaningful: a redundant measurement carries surplus information that can be used to adjust the raw value toward a reconciled estimate. Specifically, data reconciliation replaces y by a reconciled vector $\hat{x}$ that minimises the weighted squared residuals subject to the constraints,

$$\min \sum_{i \in n_m} \left(\frac{y_i - \hat{x}_i}{\sigma_i}\right)^2 \quad s.t. \;\; g(\hat{x}) = 0. \tag{4}$$

where σ_i is the standard deviation of the i-th measurement. For a physical system the constraints g usually comprise the mass and energy balances connecting the plant units.

The implementation already incorporates a global test (GT) that compares the objective value at the optimum to the upper quantile of a chi-squared distribution with ν degrees of freedom (Narasimhan and Jordache, 2000):

$$GT = \min \sum_{i \in n_m} \left(\frac{y_i - \hat{x}_i}{\sigma_i}\right)^2 < \chi^2{}_{1-\alpha}(\nu). \tag{5}$$

If the inequality is violated the dataset is declared inconsistent and an error is raised, but the existing code does not identify which measurement is responsible. The remainder of this paper addresses that gap.

The Case: WS-PEMS gas turbines

At Weel & Sandvig, the WS-PEMS software predicts emissions of methane slip, CO_2, NO_x, carbon monoxide and more from a range of industrial combustion processes. The present work focuses on gas turbines used both for power generation and for driving pumps and compressors. Figure 1 shows a Solar Titan 130 industrial gas turbine, and Figure 2 the unit-component model used to represent it in WS-PEMS.

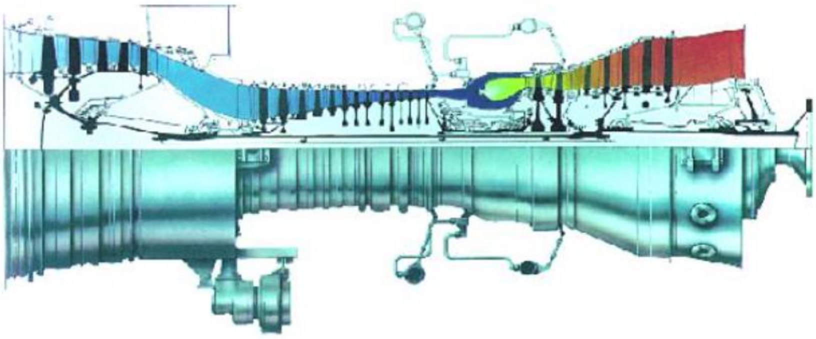

Figure 1. Solar Titan 130 industrial gas turbine.

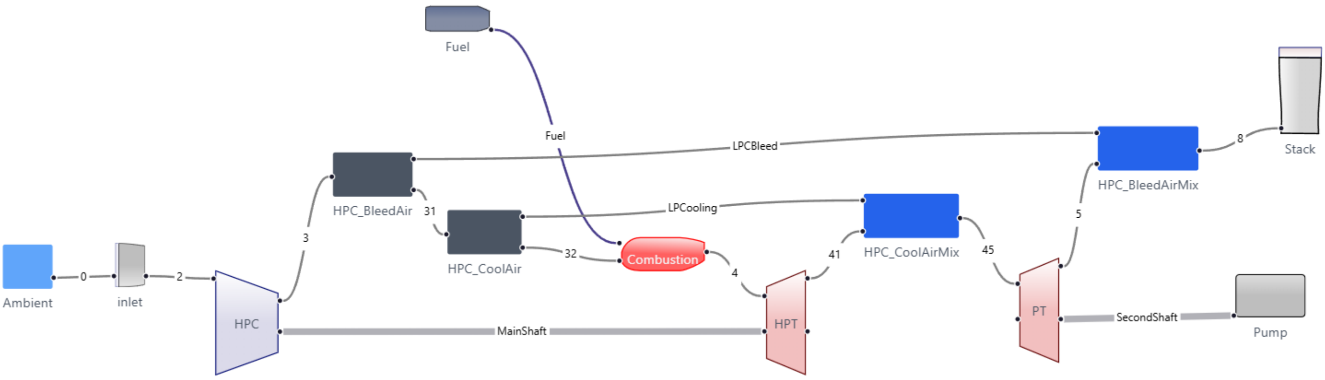

Figure 2. Schematic of the unit-component model used to represent the gas turbine in WS-PEMS. The high-pressure compressor (HPC), high-pressure turbine (HPT) and power turbine (PT) are connected by the main and second shafts; fuel enters the combustor and exhaust leaves through the stack.

Reliable emission estimates require a reliable estimate of the underlying system state, which is what data reconciliation provides. However, with only mass and energy balances, the gas turbine system is typically under-determined or, at best, marginally observable: standard plant instrumentation does not measure enough state variables for every unknown to be uniquely determined. To restore observability, we adopt the framework of generalised data reconciliation (Szega, 2009; Heyen and Kalitventzeff, 2006; VDI, 2017), in which equations describing the characteristic behaviour of individual units are added to the model. For each turbine stage we use Stodola's ellipse, which links inlet pressure, outlet pressure and mass flow, and for the compressors we use Greitzer-style maps relating corrected mass flow, pressure ratio and corrected speed (Greitzer, 1976; Guo et al., 2016). These characteristic equations are not enforced as hard constraints. Instead, each characteristic residual $h_j(\hat{x}),\ h: \mathbb{R}^n \rightarrow \mathbb{R}^{n_e}$, is associated with a variance ${\sigma_j}^2$ and appended to the objective as a penalty term, so that the reconciled state minimises,

$$\min\left(\sum_{i \in n_m}\left(\frac{y_i - \hat{x}_i}{\sigma_i}\right)^2 + \sum_{j \in n_e}\left(\frac{h_j(\hat{x})}{\sigma_j}\right)^2\right), \quad s.t.\ \ g(\hat{x}) = 0. \tag{6}$$

In this way $h_j(\hat{x})$ is a soft-constraint, or virtual-measurement: every characteristic equation acts as an additional pseudo-measurement of zero residual, with its own declared uncertainty.

A second challenge for real-time monitoring is that faulty measurements, if left undetected, are propagated by the reconciliation optimisation and bias every reconciled variable. The optimisation in eq. (6) is solved by Sequential Least-Squares Programming (SLSQP). If the global test is outside the 95% chi-squared confidence interval the reconciliation is flagged.

We implemented and compared four strategies for identifying the measurement causing the gross error leading to the failed chi-squared test:

Method 1 — Measurement test

Following Knopf (2011), the constraint system is linearised by a forward finite-difference Taylor expansion around the reconciled point $\hat{x}$, which gives the Jacobian J. The covariance matrix of the residual $a_i = y_i - \hat{x}_i$ is then $V = Q\,A^{\mathrm{T}}\,(A\,Q\,A^{\mathrm{T}})^{-1}AQ$, and the per-measurement test statistic is

$$MT_i = \frac{|a_i|}{\sqrt{V_{ii}}} \tag{7}$$

The index maximising MT_{i} is flagged as suspicious. Termination is guaranteed: each iteration removes at least one measurement, so the global test is eventually driven to zero.

Method 2 — Exhaustive index masking

Method 2 trades algebraic effort for combinatorial effort. Each measured tag is masked in turn, the optimisation is re-solved without it, and the resulting global test is recorded; the tag whose removal yields the smallest GT is declared suspicious. The approach is intuitive and accurate but requires one full reconciliation per candidate index.

Methods 3 and 4 — Early termination and MT pre-ranking

Methods 3 and 4 retain the masking principle of Method 2 but introduce early-termination criteria, so that the candidate loop exits as soon as the global test has been reduced enough. Method 4 additionally sorts the candidate indices in descending order of MT_i before masking, so that the indices most likely to be faulty are tested first. This combination preserves the accuracy of masking while sharply reducing the average number of optimisations per iteration.

The four methods were evaluated on a Solar Titan 130 emission source over 7,200 perturbations (600 iterations of 12 measurements each), with each measurement shifted within $\pm 400\sigma$ around its nominal value. Detection outcomes were classified as true or false positives (TP, FP) and true or false negatives (TN, FN), and the standard summary metrics were computed: error rate (FP + FN) / (TP + FP + FN + TN), false-positive ratio FP / (FP + TN), and precision TP / (TP + FP).

Table 1. Detection accuracy and computation time of the four methods on the Solar Titan 130 dataset.

Metric	M1	M2	M3	M4
Error rate	0.026	0.007	0.023	0.013
FP ratio	0.018	0.004	0.015	0.008
Precision	0.684	0.919	0.726	0.845
Computation time for one error [s]	0.8	6.8	1.4	2.0

Method 2 attains the highest precision (0.92) but its computation time is roughly an order of magnitude larger than that of Method 1. Method 4 closes most of the accuracy gap (precision 0.85) while staying within an acceptable execution budget; it is therefore selected as the default high-precision option, with Method 1 retained where speed dominates.

REFERENCES

Crowe, C.M. (1996). Data reconciliation – progress and challenges. *Journal of Process Control*, 6(2–3), 89–98.

Crowe, C.M., Garcia Campos, Y.A., and Hrymak, A. (1983). Reconciliation of process flow rates by matrix projection. Part I: Linear case. *AIChE Journal*, 29(6), 881–888.

Greitzer, E.M. (1976). Surge and rotating stall in axial flow compressors – Part I: Theoretical compression system model. *Journal of Engineering for Power*, 98(2), 190–198.

Guo, S., Liu, P., and Li, Z. (2016). Inequality constrained nonlinear data reconciliation of a steam turbine power plant for enhanced parameter estimation. *Energy*, 103, 215–230.

Heyen, G. and Kalitventzeff, B. (2006). Process monitoring and data reconciliation. In Puigjaner, L. and Heyen, G. (eds.), *Computer Aided Process and Product Engineering*, pp. 517–540. Wiley-VCH, Weinheim.

Knopf, F.C. (2011). *Modeling, Analysis and Optimization of Process and Energy Systems*. Wiley, Hoboken, NJ.

Narasimhan, S. and Jordache, C. (2000). *Data Reconciliation and Gross Error Detection: An Intelligent Use of Process Data*. Gulf Publishing, Houston, TX.

Romagnoli, J.A. and Sánchez, M.C. (2000). *Data Processing and Reconciliation for Chemical Process Operations*. Academic Press, San Diego, CA.

Stanley, G.M. and Mah, R.S.H. (1981). Observability and redundancy in process data estimation. *Chemical Engineering Science*, 36(2), 259–272.

Szega, M. (2009). Application of the generalised method of data reconciliation in thermal technology. *Archives of Thermodynamics*, 30(4), 219–232.

VDI (2017). *VDI 2048 Part 1 — Uncertainties of Measurement during Acceptance Tests on Energy-Conversion and Power Plants: Fundamentals*. Verein Deutscher Ingenieure, Düsseldorf.