## APPLYING DATA RECONCILIATION AND THE DIAGNOSTIC MODEL PROCESSOR TO A PAPER MACHINE

David I. Wilson Engineering Research Institute Auckland University of Technology, New Zealand email: diwilson@aut.ac.nz

#### ABSTRACT

This paper compares the techniques of Data Reconciliation and the Diagnostic Model Processor applied to an industrial full scale two-ply paper machine using actual operating data. Data reconciliation optimally adjusts the raw data to satisfy known constraints whilst simultaneously identifying gross errors. The DMP searches for faults that create the observed discrepancies in these constraint equations. The DR strategy worked well, reconciling the raw measurements and correctly identifying gross errors while the DMP was over enthusiastic in its attempts to identify assumption violations.

## **KEY WORDS**

Data reconciliation, paper machine, gross errors, diagnostic model processor

## **1** Introduction

The trend in a competitive industry is to continually improve and optimise production and the efficient management of information is an important factor to achieve this. Given the ability for a Distributed Control Systems (DCS)to generate vast amounts of data, it is crucial that this data be compacted, so only essential validated information is used to schedule, control and monitor a complex industrial operation.

While techniques for compacting, reconciling and extracting information from raw data are not new, and notwithstanding some applications such as that reported in [1] and [2], it appears that in many processing plants, raw measured data is simply stored in the historian until a periodic purge occurs. Given that key production data used for planning is rarely reconciled with the less important data, any errors due to transducer or mechanical failure are unlikely to be identified. Avoiding this potentially dangerous situation is the aim behind the optimal treatment of measured data. Even when the raw data does not seem to have an immediate value, combined with a model, it is possible to generate an improved data set, thus strengthening the value and confidence of the data.

This paper considers two different techniques for processing the raw data, Data Reconciliation (DR) and the Diagnostic Model Processor (DMP), [3].

### 2 A two-ply paper machine

Fig. 1 shows a simplified schematic of a two-ply paper board machine at Gruvön Mill in central Sweden.

The flow and concentration of long and short pulp fibers are measured by magnetic flow and pulp consistency meters respectively. The almost dry finished product is analysed for moisture and basis weight using a Measurex radioactive scanner. The flowrate of water removed first by vacuum through the wire, and then later by heated rollers is unmeasured, and essentially unmeasurable. This excludes the possibility of a full mass balance across the machine. Nominal operating values for the key process measurements to be reconciled are given in Table 1.

Table 1. The 10 logged variables to be reconciled

	Description	value
$f_s$	short fibre flow	153 ton/hr
$f_l$	long fibre flow	76 ton/hr
$c_b$	conc. after chest (bot.)	2.82 %
$f_b$	mass flow after chest (bot.)	3954.0 1/min
$c_t$	conc. after chest (top)	3.45 %
$f_t$	mass flow after chest (top.)	1868.0 1/min
w	basis weight	$152.9 \text{ g/m}^2$
v	machine speed	310.3 m/min
$\bar{m}$	average moisture profile	7.0 %
p	production rate	12.25 ton/hr

Paper machines are a challenge to any data management system due to the frequent paper breaks, and grade changes indicated by the operating data trended in Fig. 2. Consequently this is an appropriate data set to trial data reconciliation and diagnostics on a full scale, important industrial process.

#### 2.1 Data reconciliation

In the linear data reconciliation problem, [4], we wish to establish a set of reconciled measurements,  $\mathbf{x}$ , given possibly erroneous raw measured data,  $\hat{\mathbf{x}}$ , to minimise the weighted squared residual

$$\mathcal{J} = (\mathbf{x} - \hat{\mathbf{x}})^T \mathbf{Q}^{-1} (\mathbf{x} - \hat{\mathbf{x}})$$
(1)



Figure 1. A simplified process flow schematic of the twolayer paper machine.



Figure 2. Top and bottom sheet flows to the paper machine showing the frequency of the machine stoppages.

where  $\mathbf{Q}$  is a weighting matrix often chosen as the covariance matrix of the anticipated uncertainties in  $\mathbf{x}$ , subject to the constraints

$$\mathbf{A}\mathbf{x} = \mathbf{b} \tag{2}$$

which could include structurally exact mass and energy balance relations, or perhaps regressions from equipment manufacturer's data where we have less confidence. The solution to this constrained optimisation problem is well known, [5],

$$\mathbf{x} = \hat{\mathbf{x}} \mathbf{Q} \mathbf{A}^{T} \left( \mathbf{A} \mathbf{Q} \mathbf{A}^{T} \right)^{-1} \left( \mathbf{b} - \mathbf{A} \hat{\mathbf{x}} \right)$$
(3)

but in realistic cases, the constraint equations will be nonlinear, particularly in the case of energy balances, compromising this closed form solution.

# 2.2 Estimating variances from industrial operating data

The key tuning parameter in the DR strategy is the weighting matrix  $\mathbf{Q}$  in Eqn. 1 and is typically set to the noise covariance of the measurements, but should additionally reflect the relative importance of the measurements.

In industrial applications, it is rare to use anything other than a diagonal covariance matrix, but in many cases the measurement noise is strongly correlated, particularly in the situations where many of the so-called raw measurements are themselves derived from more basic, but inaccessible measurements such as in the proprietary basis weight and moisture scanning systems commonly used on paper machines.

Furthermore it is difficult to estimate  $\mathbf{Q}$  under industrial operation. Fig. 3(a) shows the average moisture sampled every 12 minutes over 160 hours. This trended variable shows periods of *no* noise, periods of noise that would seem reasonable of this industrial measurement, and periods where either the reading is faulty, or the machine is stopped. Using all the data in Fig. 3(a) blindly gives a poor estimate of the instrument noise.

Even if only periods containing 'reasonable' noise are considered, it is evident from the portion shown in Fig. 3(b) that this typical operating data is still not normally distributed, in part due to the presence of excessive outliers. The Jarque-Bera statistical test applied with a confidence level of 95% confirms that this data was not drawn from a normal distribution. This non-normality in key measurements presents a further challenge to the standard DR strategy.



(b) A zoomed portion used for a covariance calculation showing the  $\pm 1\sigma$  limits and the substantial deviation from normality in the distribution of the data.

Figure 3. The variability of the average moisture content during steady operation.

Note that in order for the inverse of the covariance matrix  $\mathbf{Q}$  to exist, all measurements must have some (non-zero) error. While this is not an unreasonable assumption, those measurements that are known to an extremely high

degree of precision can be removed from the optimisation problem.

#### 2.3 Gross error detection

The problem with data reconciliation is that the strategy is based on the assumption of a Gaussian error distribution for the residuals and will likely fail catastrophically in the presence of measurement bias, process leaks or abnormal variances which are collectively known as gross errors. To a lesser degree, we know it will perform sub-optimally in the presence of benign, but still non-normally distributed noise, as exhibited by the moisture reading from the scanner shown in section 2.2.

But identifying gross errors in non-trivial applications requires a prior data reconciliation step which leads naturally to an iterative procedure such as the Modified Iterative Measurement Technique (MIMT) proposed by [6].

- 1. Reconcile the data and compute the measurement residual vector,  $\boldsymbol{\epsilon} = |\mathbf{x} \hat{\mathbf{x}}|$ .
- 2. For each residual compare the statistic  $z_i = \epsilon_i / \sigma_i$ where  $\sigma \stackrel{\text{def}}{=} \sqrt{\text{diag}(\mathbf{Q})}$

There are many variants to this basic scheme as surveyed by [7, 8], which in turn borrows from strategies employed in robust statistics such as iteratively re-weighted least squares, [9].

#### **3** The diagnostic model processor

The diagnostic model processor (DMP) developed in [3], is a generalised fault diagnostic tool which, like the data reconciliation scheme, uses the constraint residuals as a pointer to possible violated assumptions. Each constraint residual,  $e_i$  is transformed to a normalised residual of the form

$$\epsilon_i = \frac{(e_i/\tau_i)^{\zeta}}{1 + (e_i/\tau_i)^{\zeta}} \tag{4}$$

where  $\tau$  is a tuning factor, and shape factor  $\zeta = 4$  is recommended. The DMP scheme generates a vector of failure likelihoods, **f**, for each of the *n* assumptions. For the *i*th assumption,

$$f_i = \frac{\sum_{j=1}^n s_{ij}\epsilon_j}{\sum_{j=1}^n |s_{ij}|} = \frac{(\mathbf{S}\boldsymbol{\epsilon})_i}{\operatorname{rowsum}(\mathbf{S})_i}$$
(5)

where S is the sensitivity matrix quantifying how dependent a particular equation is on a particular assumption. Ideally the only nonzero elements in f are the true faults, although in practice a cut-off limit of 0.5 is recommended to avoid false positives.

## 4 Data reconciliation around the paper machine

A full mass balance around the machine is

$$f_s + f_l + W_1 + W_2 = W_{\text{steam}} + W_{\text{ret}} + wv \left(1 - \frac{m}{100}\right) 4.3$$
(6)

where the width of the paper is 4.3m. Unfortunately Eqn. 6 is impractical since the recycle water,  $W_{\text{ret}}$ , and evaporation,  $W_{\text{steam}}$ , flows are on this machine unmeasured. Note that the dilution flows,  $W_1$  and  $W_2$ , to the machine chests are measured. However a dry material balance gives

$$f_b c_b + f_t c_t = wv \left(1 - \frac{\bar{m}}{100}\right) 4.3 = p \left(1 - \frac{\bar{m}}{100}\right)$$
(7)

Ideally the three terms in Eqn. 7 should all exactly equate, but Fig. 4 shows in reality there is a small fibre gain (!) if the raw data is to be believed uncritically. This highlights the need for data reconciliation.



Figure 4. A fibre balance showing the three supposedly equal terms from Eqn. 7.

To maintain the correct board composition, the flow rates are ratio controlled and the fibre flows are kept at a high concentration to avoid large pumps and buffer storage vessels up until just before the paper machine where is is diluted. Combining these constraints to the fiber balance constraints gives us six constraint equations (in units as given in Table 1) as

$$f_b \frac{c_b}{100} + f_t \frac{c_t}{100} - v \frac{wb}{10^3} \left(1 - \frac{\bar{m}}{100}\right) 4.3 = 0 \tag{8}$$

$$4.3v\frac{wb}{10^3} - 16.67p = 0 \tag{9}$$

$$f_b - 2f_t = 0$$
 (10)

$$f_s - 2f_l = 0$$
 (11)

$$f_b - 25f_s = 0 \qquad (12)$$

$$f_t - 25f_l = 0 \tag{13}$$

Using data reconciliation, we wish to establish the optimal n = 10 measurements which satisfy the m = 6 nonlinear constraint equations. This can be solved using any unconstrained nonlinear optimiser by incorporating Lagrange multipliers.

Fig. 5 trends both the raw measured data with the reconciled values over a period of 35 hours. In this case

the reconciliation is not performed using a dynamic model because the update time of 12 minutes is far outside the dominant dynamics of the plant as established in [10] so a pseudo steady-state is assumed. The tradeoff when rigourously satisfying the constraints is that the reconciled data is far less smooth than the raw data. In this application the mill operating staff were surprised, and a little disappointed, in this consequence.



Figure 5. The raw (heavy) and reconciled (light) short and long fibre flowrates for part of the data series.

#### 4.1 Gross error detection

Fig. 6 shows a gross error at t = -123.8 hours when the concentration transducer after the machine chest to the bottom layer failed. Instead of concentration reading of around 3%, the  $c_b$  reading jumped to over 40%. At that instance, the raw measured data, the reconciled data and the failure vector for this point are given in Table 2.

The DR strategy outlined in section 2.3 correctly identified the gross error,  $\blacksquare$ , in this instance. However one must take care in solving the nonlinear optimisation problem to linearise about the immediate previous point where it is assumed that there are no unidentified gross errors, and that the operating point has not substantially changed. If the linearisation is performed about the current point, the as yet unidentified gross errors tend to destroy the linearisation which subsequently causes the DR strategy to deliver unreliable results.

Conversely, while the DMP analysis correctly shows the failure of  $c_b$ , the method also asserts other measure-



Figure 6. Flow and concentration measurements showing the gross error in the fibre concentration measurement after the machine chest to the bottom layer,  $c_b$ , at t = -124 hours.

Table 2. The raw data,  $\hat{\mathbf{x}}$ , the associated failure vector,  $\mathbf{f}$ , reconciled data,  $\mathbf{x}$ , and gross error detection results for a single gross error suspected in  $c_b$ .

	$\hat{\mathbf{x}}$	f	х	OK ?
$f_s$	157.98	0.05	160.68	
$f_l$	82.06	0.93	80.34	
$c_b$	43.78	1.00	3.01	
$f_b$	3806	-0.10	4017	
$c_t$	3.45	1.00	3.45	
$f_t$	2006	0.2	2008.5	
w	153.1	-0.48	153.05	
v	310.3	-0.48	310.25	
$\bar{m}$	7.08	1.00	7.08	
p	12.25	0.00	12.25	

ments such as  $f_l$ ,  $c_t$  and  $\bar{m}$  to have failed (indicated as a  $\Box$  in Table 2). In this particular case, this is unlikely. This demonstrates that the nonlinear DMP works well if all the values are nearly correct, but tends to over exaggerate the problems if even one gross error exists.

Fig. 7 trends all the elements of the failure vector over 50 hours. Clearly it is rare that the DMP algorithm is ever completely satisfied that there are no violated assumptions.

## 5 A comparison between DR & DMP

Data reconciliation and the DMP have different aims. Data reconciliation produces a reconciled data set, possibly with an extra set of severely suspect measurements. The DMP identifies violated assumptions but it does not attempt to calculate a reconciled data set.

In the paper machine application presented in this paper, the less computationally demanding DMP strategy had



Figure 7. The failure vector trended over time. Anything within  $\pm 0.5$  is deemed acceptable.

trouble isolating the one known fault as compared to the data reconciliation strategy with could identify the correct fault, and then subsequently go on to reconcile the values. However published reports indicate *both* strategies tend to identify too many gross errors which is probably due to the fact that most automated ways of determining the variance of real data gives values that are too small and the nonlinearities are ignored.

Both schemes require tuning to provide informative results. In the case of DR, the tuning involves the selection of the co-variance matrix while for the DMP scheme, one must select the  $\tau$  vector. Of the two schemes, selecting a suitable Q proved more difficult than just setting a diagonal matrix proportional to the estimates of the individual transducer variances. One explanation of the excessive variance exhibited in the reconciled variables in Fig. 5 compared to the raw measurements is that the some of the elements in the weighting matrix are too small, perhaps due to an overly optimistic view of the quality of the transducer. This would have the effect that some of the variables are prevented from deviating too far from their raw values, which in turn would force other measurements to excessive deviations in order to satisfy the required constraints.

Both schemes as presented assume pseudo steadystate conditions, although modifications have been developed for dynamic systems. In this case, the update time of 12 minutes is well outside the dominant time constants of the paper machine. In paper production, the periods of unsteady-state operation are always carefully monitored, and the products are recycled.

## 6 Conclusions

The data reconciliation strategy applied to a two-ply paper machine using actual operating data was reasonably successful. The strategy could reconcile the data, (although admittedly by introducing what some operators might consider excessive variation in the measured variables), and the strategy could correctly identify actual gross errors.

The diagnostic model processor was less successful. The strategy over estimated the incidents of faults and the tuning is less transparent than that for DR. However the computation is considerably easier and more robust.

#### References

- C. Karlsson, E. Dahlquist, and E. Dotzauer. Data reconciliation and gross error detection for the flue gas train in a heat & power plant. In 8th International Conference on Probabilistic Methods Applied to Power Systems, pages 193–198, Iowa, USA, September 2004.
- [2] Shankar Narasimhan and Cornelius Jordache. Data Reconciliation and Gross Error Detection: An Intelligent Use of Process Data. Gulf Publishing, Houston, Tx., 2000.
- [3] Thomas A. Petti, Jim Klein, and Prasad S. Dhurjati. Diagnostic model processor: Using deep knowledge for process fault diagnosis. *American Inst. Chemical Engineers J*, 36(4):565–575, April 1990. See also letters to the Editor, AIChE J. (1990) pp1121–1124.
- [4] Richard S. Mah, Gregory M. Stanley, and Dennis M. Downing. Reconiliation and retification of process flow and inventory data. *Ind. Eng. Chem. Process De*sign Dev., 15(1):175–183, 1976.
- [5] D.R. Kuehn and Harold Davidson. Computer control II Mathematics of control. *Chemical Engineering Progress*, 57(6):44–47, June 1961.
- [6] R.S.H. Mah and A.C. Tamhane. Detection of gross erros in process data. *American Inst. Chemical Engineers J*, 28(5):828–830, September 1982.
- [7] Dider Maquin and José Ragot. Comparison of gross error detection methods in process data. In 30th Conference on Decision and Control, pages 2254–2261, Brighton, UK, December 1991.
- [8] Jian Chen, Zheng Chen, Hongye Su, and Jian Chu. Two-step method for gross error detection in process data. In *Proceedings of the American Control Conference*, pages 2121–2126, Arlington, VA, USA, 25–27 June 2001.
- [9] Antoine Guitton and William W. Symes. Robust and stable velocity analysis using the Huber function. Technical Report SEP-100, Stanford Exploration Project, Stanford University, USA, 1999.
- [10] Jonas Balderud and David Wilson. Parameters affecting disturbance propagation through the wet end of a paper machine. *Appita Journal*, 58(1):40–46, January 2005. ISSN: 1038-6807.