On-line processing and steady-state reconciliation of pulp and paper mill process data

BY S. BELLEC, T. JIANG, B. KERR, M. DIAMOND AND P. STUART

Abstract: In order to increase the value of real-time data accessible through implemented data management systems, on-line data treatment techniques have been evaluated. Processing and reconciliation techniques can be used to provide better quality data by correcting various sources of errors present in real-time measurements, including random noise, abnormalities, and inconsistencies. The on-line application of these techniques could potentially assist mills in quickly identifying instrument maintenance requirements, in troubleshooting operating problems, and in improving decision-making related to process activities.

Measurements are inherently corrupted with various sources of error, including instrument miscalibration or malfunction, power supply fluctuation, wiring, process noise, and other reasons. Unless they are carefully treated, using process data to make process decisions in real-time can lead to misinterpretations caused by this poor data quality.

Figure 1 shows a set of flowrate measurements including the 3 different types of errors present in data measurements:

- Random errors are mostly concentrated at high frequencies and can be associated with power supply fluctuations and signal conversion noise.
- Abnormalities, considered to be supernatural random errors characterized by a sudden change in the measurement signal. This high amplitude measurement can be caused by electric voltage surges or wiring problems.
- Gross errors associated with bias, drifting in measurements, complete failure, and precision degradation of instruments [2]. They are related to instrument wear, corrosion, miscalibration or malfunction, and to inconsistencies in a set of measurements around a system, e.g., process leaks or accumulation.

Data processing and reconciliation techniques have already been applied in the pulp and paper industry for various off-line applications. For example, Jacob and Paris [3,4] used data reconciliation techniques to provide more precise data for the calibration of integrated newsprint mill simulations, and suggested redundancy and sensitivity analysis for the mill sampling protocol design. The actual paper suggests going a step further by applying a systematic method suitable for on-line applications to improve the steady-state data quality of real-time measurements.

ON-LINE DATA PROCESSING METHODOLOGY

In order to correct the different sources of error present in real-time process measurements, an on-line methodology based on data processing techniques and steady-state data reconciliation is proposed. Figure 2 illustrates the systematic methodology developed in this study. This on-line procedure is inspired by an off-line strategy used...
to reduce various types of errors included in a set of measurements [5]. First, abnormal measurements are detected and corrected, and then a data filtration technique is used to correct random errors. In order to correct gross errors, near steady-state data sets are carefully identified and then reconciled using a weighted least squared method.

Detection of abnormalities
In the first step of the methodology, abnormal measurements are detected using a method based on wavelet transforms. Sudden changes (abnormalities) in the measurement signal are detected by identifying a coupling of wavelet modulus maxima with opposite signs and large amplitude [6].

Data filtration
Wavelet transforms are useful to represent a series of measurements as a limit of successive approximations. This manner of describing temporal measurements has the advantage of highlighting the high frequency components of the signal [6]. Therefore, wavelet transforms are used in the proposed methodology to identify and discard random noise present in on-line measurements.

Steady-state identification
In order to reconcile on-line steady-state measurements, one should first identify adequately the near steady-state operating conditions for the process under study. This is done using a novel technique based on wavelet transforms and statistical theories [7].

Data reconciliation
Data reconciliation techniques are useful to correct gross errors present in redundant measurements from a system. This is done by evaluating the reliability of the measurements based on mass and energy balance relationships. Due to the importance of redundancy in data reconciliation, the first task to accomplish is classifying the variables under study [8]. If the system is over-specified, then gross error detection and reconciliation can be performed.

Errors between measurements and real process values are minimized under physical constraints using a weighted least square equation. Figure 3 illustrates an example of reconciliation between measurements (100, 61, 38) acquired around a splitter. Results obtained by the minimization of errors under steady-state mass balance constraints show that reconciled measurements (100, 60.3, 39.7) are near the real process values (100, 60, 40).

On-line application
The 4-step methodology (detection of abnormalities, data filtration, steady-state identification, and data reconciliation) proposed by Jiang et al. [5], and modified here for on-line application, requires the calibration of various parameters. This calibration is greatly influenced by the measurement signal variance and the process conditions of operation. In other words, whether optimal parameter values can be obtained depends on the status of the system under study. Therefore, for on-line application, the methodology includes a system status identification. This step has two main purposes:

• To identify the conditions of operation for the process under study,
• To select data processing parameters associated with the identified process status.

The selection of these parameter values is based on an off-line investigation of properties performed for various steady-state conditions of operation. In each case, the parameters are set to achieve optimal results in the data treatment procedure.

Case study
The applicability and performance of the proposed methodology were tested using real-time data from a paper machine stock approach system at an integrated newsprint paper mill, Fig. 4. In order to assess and evaluate the methodology, the system was investigated in 3 steps:

1) Historical process data were analyzed and hypothetical process signals were developed,
2) A systematic analysis of the data treatment methodology was performed on the hypothetical data sets,
3) The proposed methodology was applied to real-time process measurements.

Historical Process Data Analysis and Creation of Hypothetical Process Data Sets
To be able to characterize the accuracy and precision of the methodology, treated data must be compared with the true process values. Since on-line measurements are always corrupted with errors, their true value is never exactly known. Therefore, in order to investigate the accuracy of the proposed methodology, hypothetical process data were created.

In order to create hypothetical data that are representative of real process operations, a study of the system was performed in 3 steps:

1) Based on historical data, near steady-state operating regimes were identified for the system under study,
2) Characteristics associated with the different operating regimes were compiled,
3) Based on the characteristics of the different operating regimes, hypothetical data sets with known errors were created.

Identification of near steady-state regimes of operation
To investigate the different operating regimes of the system, historical process data were analyzed. For each process variable measurement, steady-state periods were selected and combined to identify overall system steady-state conditions of operation, Fig. 5.

For the entire system, 27 different operating regimes were grouped into 4 general system classes, namely: production of paper grade A, production of paper grade B, maintenance stops, and occurrence of a paper break, Fig. 5. An example demonstrating the differences between two regimes of operation within the same class is also presented.

One should be aware that, as the system under study becomes larger, the number of operating regimes quickly increases, and information needed for on-line data reconciliation will grow considerably. Therefore, process operating regimes must be carefully and logically combined in order to minimize their overall number.

Characteristics of the different operating regimes
Once near steady-state operating regimes were identified, the probability of occur-
rence, the duration of each regime and the delays occurring between them were compiled based on historical data. Statistics related to the mean and variance of process measurements within each regime were also compiled and stocked in a database. In the on-line application of the proposed methodology, these characteristics are useful to properly identify the conditions of operation of the process.

Creation of hypothetical data with known errors

Based on the system analysis and characteristics associated with each process measurement within a regime of operation, hypothetical data sets were created. According to statistical properties and the probability of abnormal measurements in the historical data, abnormalities and random errors were added to the created signals. This method of generating the measurement signals ensured that the “true process values” and the errors associated with the hypothetical data would be known. Therefore, the accuracy of the proposed methodology could be verified using the generated measurements.

To demonstrate the representativeness of the created data sets, the system operating regimes were analyzed and compared to hypothetical data over a one-week period. Identified regimes were reproduced and errors were added to create a hypothetical corrupted set of data (generated data). Figure 6 shows the results of the investigation for a particular flowrate measurement over time. Although the general characteristics of the two signals are the same, the hypothetical data can contain more or less noise in particular regions and part of the signal can be corrected to fit overall mass balances around the system.

Investigation of Data Processing and Reconciliation Methodology Using Hypothetical Data Sets

The step-by-step data processing and reconciliation methodology was applied to a variety of hypothetical corrupted data sets. Figure 7 shows an example of the results obtained, for a particular flowrate measurement. At each step of the methodology, the raw “measurements” were corrected and compared to the targeted values (hypothetical data sets without errors). In the first two steps, abnormal measurements and random errors were corrected using wavelet transform properties. Later, steady-state identification was performed and steady-state data sets were selected. Finally, the selected data were reconciled through mass balances on the overall system. The results indicate an improvement in the accuracy of the corrected steady-state measurements compared to the raw data (hypothetical data).

Figures 8 and 9 summarize the improvement of steady-state data by comparing the average accuracy and precision of the data before and after the application of the proposed methodology.

Accuracy is defined as the average relative error (%RE) compared to the targeted value, equation 1, while precision is defined as the standard deviation (SD%RE) of such errors, equation 2. Therefore, lower values on the graphs indicate improvements in data accuracy and precision. An improvement in accuracy indicates an overall amelioration of steady-state data due to the application of the methodology, while an improvement in precision demonstrates the ability of the method to correct gross errors.

\[
\text{%RE} = 100\% \times \frac{|\text{Targeted value} - \text{measurement}|}{\text{Targeted value}} \quad (1)
\]

\[
\text{SD}_{\text{RE}} = \sqrt{\frac{\sum_{i=1}^{n} 100\% \times (\text{average(RE)} - \text{RE}_i)^2}{n - 1}} \quad (2)
\]

Through the application of the proposed methodology to generated data sets, it can be concluded that the overall accuracy and precision were improved. It is apparent that the methodology has a greater effect on flowrate measurements than on consistency data. This can be explained by the fact that flowrate measurements are associated with two levels of mass balance constraints (overall and fiber mass balances), while consistency...
measurements are only affected by fiber mass conservation equations.

Evaluation of Methodology Robustness on Real-time Process Measurements

A methodology used to improve process measurement quality must improve the overall accuracy and precision of measurements, but should also be robust. Robustness is defined here as the ability of the methodology to be successfully applied to real-time process measurements under different operating conditions.

To investigate the robustness of the methodology, it was applied to real-time data (directly measured from the process). The investigation was performed on 10 flowrate and 8 consistency signals over a period of 3 days (sampling time: 10 seconds). In order to ensure that process time lags were taken into account, only steady-state periods longer than a threshold value (system delay + safety factor) were considered. The method under investigation was found to be robust for treating real-time process data.

Figure 10 shows 3 different examples of data quality improvement obtained by the application of the proposed methodology. Abnormalities and random noise have been corrected, and data under near steady-state conditions were efficiently corrected using data reconciliation. Fig. 10a. Instrument offset, Fig. 10b, and total failures, Fig. 10c, were also identified and corrected.

IMPLICATION OF RESULTS

A novel methodology used to improve online measurement quality has been investigated with real-time data from an integrated newsprint mill. It has been shown that by applying this data treatment methodology, data accuracy and reliability are improved. Therefore, greater confidence in data quality for well-defined process operating regimes can be achieved - the “direct benefits” in Fig. 11. This benefit alone can likely justify the effort and expense of implementing such a method. In this section, the implications and further benefits related to the on-line application of data quality improvement techniques are outlined.

By compiling on-line data reconciliation results for near steady-state operating regimes, mills can develop a relational database of accurate process data. These results can help to express the probability of occurrence curve for different operating parameters. Figure 12 shows an example of the probability of occurrence curve for a flowrate measured at a process point, during a specified regime of operation.

This could potentially lead to benefits related to process operation analysis and troubleshooting, classified into two major categories, Fig. 11:

• Benefits obtained as a consequence of changes in the data reconciliation outcome compared to many previous on-line analyses,
• Benefits obtained as a consequence of analyzing many sets of steady-state process simulation data.

In the first case, the outcome of data reconciliation measurements can be compared to the previous reconciled data available in the database. The following potential benefits can be realized:

• By comparing the measurement correction applied to the reconciled data, uncalibrated instruments and/or process leaks can be identified,
• Real-time yield accounting (performance analysis) for near steady-state operating conditions can be performed.

In the second case, information issued
from the compiled data reconciliation results, illustrated in Fig. 12, can be used to localize, assess, and analyze process operations. The better resolved data from different process operating regimes improves the understanding of the process during near steady-state operating conditions. Therefore, these results can also enable more precise conditions for process optimization studies used to enhance process planning activities, i.e., scheduling and plant-wide optimization [9,10]. The possibilities in this latter example are considerable; coupling process and cost data sets for a specific product would allow enhanced decision-making and mill profits [11].

**CONCLUSIONS**

A robust methodology to improve steady-state data accuracy and precision was developed in this study. First, measurement signals were improved by abnormality and noise removal, then the process trend was evaluated by a novel steady-state identification technique. If near steady-state conditions could be identified, then data inconsistencies were removed using data reconciliation. Improvements in data precision and accuracy were demonstrated by applying the proposed methodology to hypothetical and real process signals. The application of this methodology on-line could yield significant benefits, including rapid instrument failure identification, process leak detection, improved process yield calculations, and more reliable relational data for overall process optimization.

**ACKNOWLEDGEMENTS**

This work was completed with support from the Natural Sciences and Engineering Research Council of Canada (NSERC) Environmental Design Engineering Chair at École Polytechnique.

**LITERATURE**


