PolyPublie
Polytechnique Montréal

POLYTECHNIQUE MONTRÉAL
UNIVERSITÉ D'INGÉNIERIE

Titre: Title:	Réconciliation de données en ligne
Auteur: Author:	Steve Bellec
Date:	2004
Type:	Mémoire ou thèse / Dissertation or Thesis
Référence: Citation:	Bellec, S. (2004). Réconciliation de données en ligne [Master's thesis, École Polytechnique de Montréal]. PolyPublie. https://publications.polymtl.ca/7337/

Document en libre accès dans PolyPublie
Open Access document in PolyPublie

URL de PolyPublie: PolyPublie URL:	https://publications.polymtl.ca/7337/
Directeurs de recherche: Advisors:	Paul R. Stuart, & Taiwen Jiang
Programme: Program:	Unspecified

Ce fichier a été téléchargé à partir de PolyPublie, le dépôt institutionnel de Polytechnique Montréal
This file has been downloaded from PolyPublie, the institutional repository of Polytechnique Montréal

https://publications.polymtl.ca

NOTE TO USERS

This reproduction is the best copy available.

UMI®

UNIVERSITÉ DE MONTRÉAL

RÉCONCILIATION DE DONNÉES EN LIGNE

STEVE BELLEC
DÉPARTEMENT DE GÉNIE CHIMIQUE
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION
DU DIPLÔME DE MAÎTRISE ÈS SCIENCES APPLIQUÉES
(GÉNIE CHIMIQUE)
DÉCEMBRE 2004

© Steve Bellec, 2004.

Library and Archives Canada

Published Heritage Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et Archives Canada

Direction du Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 0-494-01280-3
Our file Notre référence
ISBN: 0-494-01280-3

NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé :

RÉCONCILIATION DE DONNÉES EN LIGNE

présenté par : BELLEC Steve
en vue de l'obtention du diplôme de : Maîtrise ès sciences appliquées
a été dûment acceptée par le jury d'examen constitué de :

M. CHAOUKI Jamal, Ph.D., président
M. STUART Paul, Ph.D., membre et directeur de recherche
M. JIANG Taiwen, Ph.D., membre et codirecteur de recherche
M. Roche Alain, M.Sc.A., membre

Dédicace

À tous ceux qui ont cru moi et en ce projet du premier au dernier moment.

Remerciements

Je tiens à remercier :

La chaire CRSNG en conception environnementale pour son appui financier.

Le professeur Paul Stuart pour sa participation active à nos discussions de recherche et surtout pour m'avoir partager sa vision de l'ingénierie et de la conception avec.

Taiwen Jiang pour sa disponibilité, ses conseils, sa patience et son enseignement tout au long de ce projet.

Agnes Devarieux-Martin pour sa disponibilité et son soutient.

Les gens qui ont contribués de près ou de loin à mes travaux chez Kruger Inc. (Trois-Rivières et Montréal).

Les membres de la chaire en conception environnementale pour avoir contribuer à agrémenter mon séjour parmi vous.

Résumé

En industrie, les mesures prises sur un procédé sont généralement contaminées par plusieurs sources d'erreurs : bruit associé aux différents capteurs utilisés, mauvaise calibration des instruments de mesure, variation dans les conditions ambiantes et erreurs humaines. La présence de ces erreurs est amplifiée par l'inconsistance entre les valeurs mesurées et les bilans, de masse et d'énergie, reliés au procédé. Cette constatation rend l'utilisation directe des données brutes inefficace. Afin de pallier à ces problèmes, on utilise depuis plusieurs années, des techniques de filtration et de réconciliation des données. En plus de rendre les données consistantes aux lois de conservation de la masse et de l'énergie, une élimination adéquate des erreurs, présentes dans les mesures, permet d'augmenter l'efficacité des tâches de hauts niveaux reliés à l'opération des procédés chimiques : configuration des systèmes de contrôle, détection et diagnostic des fautes directement reliées à l'opération du procédé (fuites ou dépôts), planification et ordonnancement des opérations du procédé. La complexification et la spécificité grandissante des procédés chimiques obligent une automatisation de plus en plus importante de ces tâches directement dépendantes d'une interprétation adéquate des tendances associées au procédé étudié. L'extraction qualitative et quantitative automatisée des caractéristiques du procédé devient donc une étape essentielle pour l'amélioration éventuelle de ces tâches. Par conséquent, le traitement de données en ligne se veut une étape cruciale, permettant l'utilisation de données propres dans l'automatisation éventuelle de ces différentes tâches.

Plus spécifiquement, depuis quelques années, les usines de pâtes et papier utilisent de plus en plus un système de gestion des données couplé à un système de gestion de l'information. Selon le sondage mené par Janssen et al. (2004), le personnel des usines reconnaît que l'implantation de ces systèmes d'information améliore grandement l'accessibilité des données en temps réel ainsi que l'opportunité de les utiliser afin de régler les problèmes reliés à leur procédé. Cependant, jusqu'à maintenant, avant d'être

utilisées pour dépanner temporairement le procédé, ces données sont analysées de manière impromptue. Ce type d'analyse ne permet malheureusement pas de tirer totalement profit des systèmes d'information mis en place. Par conséquent, l'implantation d'une méthodologie permettant la correction des erreurs de mesure, en ligne, améliorerait l'efficacité de ces systèmes.

En se basant sur la nécessité d'amélioration des systèmes, une méthodologie permettant un traitement de données en ligne et l'utilisation de données fiables dans l'automatisation éventuelle des tâches de hauts niveaux est élaborée dans cette étude. Cette méthodologie se devra d'être robuste et de permettre d'améliorer, le plus efficacement possible, l'exactitude et la précision des mesures effectuées en temps réel. Cet objectif général est approché à l'aide de sous objectifs spécifiques tels que :

- La comparaison de l'efficacité des techniques de détection des anormalités en ligne.
- La comparaison des performances d'une nouvelle méthode de détection, en ligne, du régime permanent en 3 étapes avec les méthodes existantes.
- La démonstration des capacités d'implantation de la méthodologie sur des données industrielles.

Afin d'atteindre les objectifs présentés précédemment, une approche méthodologique est mise en place. Puisque l'objectif principal est relié à la méthodologie générale du traitement de données en ligne, elle englobe les autres objectifs et ne peut être vérifiée qu'après l'analyse de ces dernières. Cette approche méthodologique permet donc, en premier lieu, d'atteindre les sous objectifs pour ensuite s'attaquer à l'objectif global.

La méthodologie, en ligne, proposée dans ces travaux est basée sur la méthodologie appliquée par Jiang et al. (2003) à des problèmes hors-ligne. Les étapes de détections de anormalités, de filtration, d'identification du régime permanent et de réconciliation des données y sont présentes. En plus, la méthodologie proposée inclue une étape

d'identification du régime d'opération du système. À cette étape, l'état d'opération du système est évalué grâce à un suivi de l'évolution de variables représentants les opérations du système. Une fois que l'évaluation et l'identification du régime d'opération sont effectuées, cette étape joue deux rôles importants. Premièrement, les paramètres, utilisés pour corriger les erreurs dans les différentes étapes de la méthodologie, sont contenus dans une librairie basée sur des données historiques et sont sélectionnés ou ajustés en ligne. Par la suite, ces paramètres sont acheminées vers les différentes étapes du traitement de données. En procédant de cette façon, les paramètres nécessaires aux différentes étapes sont ajustés en fonction de l'amplitude et de la fréquence propre à chaque régime d'opération.

Afin de permettre la validation de cette méthodologie, un système d'étude a été sélectionné : système d'approche d'une machine à papier. Puisque les mesures prises en temps réel sont corrompues par la présence de plusieurs types d'erreurs, il est impossible d'utiliser ces données de manière directe afin de juger de l'amélioration de l'exactitude et de la précision des résultats. Une étape de génération de données expérimentales est donc nécessaire. Cette étape permet la création de données dont la valeur parfaite est connue et par conséquent elle permet une caractérisation de l'exactitude et de la précision des mesures.

Afin de comparer les techniques de détection des anormalités applicables en ligne, les données générées ont été utilisées. Les deux tests comparés sont: le test de Hampel et la méthode basée sur les caractéristiques de la premier transformées d'ondelettes. L'application des deux techniques permet de constater que l'utilisation de la transformée d'ondelettes est plus efficace pour détecter les mesures anormales présentes sur une longue période de temps. Cette constatation est attribuable au point de brisure de 50% associé au test de Hampel.

Une méthode permettant l'identification, en ligne, des régimes permanents en 3 étapes a également été élaborée lors de cette étude. Premièrement, le point de départ du régime permanent est détecté à l'aide des caractéristiques propres aux transformées d'ondelettes et de sa première dérivée, ensuite les données du régime sont filtrées et enfin, le point final du régime est identifié à l'aide de la première transformée d'ondelettes. L'application de cette méthode est comparée à l'utilisation de deux autres techniques utilisées pour détecter, en ligne, la présence de régimes permanents. L'application de la méthode en 3 étapes permet d'améliorer l'efficacité d'identification des régimes permanents autour d'un système en ayant un taux d'erreurs de type I et de type II reliées à l'identification des régimes qui est plus bas que les deux autres techniques étudiées.

Par la suite, une étude comparative des différentes combinaisons de techniques, applicables à chacune des étapes de la méthodologie proposée, a été effectuée. La comparaison de l'amélioration de l'exactitude et de la précision des mesures en appliquant ces méthodes permet d'identifier la combinaison : utilisation de la transformée d'ondelettes de premier ordre pour détecter les anormalités, filtration par ondelette et détection du régime permanent à l'aide de la méthode en 3 étapes comme étant la combinaison de techniques la plus efficace. Par la suite les données en régime permanent sont réconciliées à l'aide du logiciel Sigmafine de OsiSoft Inc.

En ce qui concerne l'aspect d'amélioration de la précision, la comparaison de l'utilisation des combinaisons de techniques ne permet pas d'identifier la supériorité évidente de l'une ou l'autre d'entre elles. Cette constatation s'explique par le fait que le calcul de précision des mesures est grandement influencé par la présence d'erreurs brutes. Puisque ces combinaisons possèdent toutes la même technique de détection des erreurs brutes, la précision se voit améliorée relativement de la même façon pour chaque application.

Finalement, la combinaison de technique permettant d'optimiser les résultats est intégrée à la méthodologie générale et est implantée, en ligne, afin de traiter les données en temps réel d'une machine à papier. L'application de la méthodologie a permis d'améliorer la qualité des données mesurées ainsi que d'identifier et de corriger des erreurs de décalage et de défaillances totales des instruments. Ces résultats permettent de démontrer la robustesse de la méthode et sa possibilité d'implantation en industrie.

Les résultats obtenus permettent de conclure que la méthode proposée est robuste et permet d'améliorer l'exactitude et la précision des mesures. De plus, il est démontré que la méthode de détection des anormalités basée sur l'utilisation des transformées d'ondelettes est plus efficace, pour les applications en ligne, que l'application du test de Hampel. De plus, l'utilisation d'une nouvelle méthode d'identification du régime permanent en 3 étapes et basée sur l'utilisation de la transformée d'ondelettes de premier ordre, de théories statistiques et de technique de filtration est plus efficace que les techniques disponibles présentement.

En terminant, mentionnons qu'en implantant la méthodologie proposée de traitement des données en ligne, les usines pourront développer une liste de reproductions fidèles des régimes d'opérations quasi permanents de leurs usines. Cette implantation peut mener à plusieurs bénéfices tel :

- Une détection efficace des instruments hors de calibrations
- Une amélioration du contrôle de procédé
- Une connaissance plus détaillée des différents régimes d'opérations du procédé
- L'identification des tendances locales et globales du procédé étudié
- L'identification de fuite et d'accumulation dans le procédé.
- Une amélioration des opérations de planification et d'optimisation reliées au procédé étudié.

Abstract

In industries, process measurements are usually corrupted with various sources of errors : noise associated with the sensors usage, miscalibration or malfunction of instruments, variation in ambient conditions and human errors. Errors issued from those different sources are amplified by the inconsistencies between measured values and mass or energy balances around a process. This statement makes a direct usage of measured data inefficient. Since a coupled of years, filtration and reconciliation techniques are used to overcome that problem. Those techniques are correcting the measured data set and making them consistent with mass and energy conservation laws, which allows the improvement of higher-level tasks in chemical industries : process control configuration, detection and diagnosis of process faults (process leaks or deposits), planning and scheduling of process operations. Increasing complexity of chemical plants necessities the automation of high-level tasks, which are largely influenced by the characterisation of process trends. Therefore the automatic extraction of qualitative and quantitative features related to process trend is a crucial steps in the eventual amelioration of high level-tasks. Consequently, on-line data treatment is, by extension, an important task to accomplish in order to provide high quality data for the assessment of those important tasks.

More specifically, pulp and paper mills are starting to use, more frequently, information management systems that are coupled with data management systems. According to the survey of Jansen and al. (2004), mill personnel recognize that the implementation of information systems improves considerably the access to real-time data and the opportunity to use them in order to correct problems related to process operation. However, until now, those data have been used through impromptu analysis, which does not allowed the full exploitation of the advantages related to the implementation of information systems. Consequently, the implementation of an on-line methodology allowing the correction of measurements would ameliorate system efficiencies.

A methodology allowing the utilization, by applying an adequate data treatment strategy, of reliable on-line data in the automation of high-level tasks is elaborate in this study. This methodology requires to be robust and to allow the improvement of real-time data accuracy and precision. This general objective is partially achieve by the investigation of those sub-objectives :

- To compare the efficiency of on-line abnormalities detection techniques.
- To compare the performance of a new on-line steady state identification method, performed in 3 steps, with existing methods.
- To demonstrate the capacity of implementation of the methodology on industrial data.

In order to achieve the objectives presented above, a methodological approach as been used. Since the main objective is related to the general methodology of on-line data treatment, it includes the 3 sub-objectives. Therefore, the main objective can only be verified after those. Hence, the elaborated approach is structured to first achieve the 3 sub-objectives and then is oriented toward the main objective.

The on-line methodology proposed in this study is inspired from Jiang et al. (2003) methodology applied on off-line data sets. Abnormalities detection, filtration and steady state identification steps are present in both methodologies. Also, the proposed on-line methodology includes a step to identify the system regime of operation. At that step, process status is evaluated through a follow up of high impact variables on the process operation. Once the assessment and identification of process regime of operation is done, this particular step achieves two main purposes. First, the parameters, used to correct measurements, are contained in a library based on historical data and are selected or adjusted by on-line measurements. After, those parameters are send to the different steps of the methodology. In other words, that steps allows the selection and adjustment of the parameters, used in the different steps of the methodology, according to the amplitude and frequency of each operation regimes.

In order to allow the validation of the proposed methodology, a system as been selected : a papermachine stock approach system from an integrated newsprint mill. Since on-line measurements are corrupted by various sources of errors, it is impossible to use directly those corrupted data to investigate the amelioration of accuracy and precision in the data issued from the proposed methodology. In order to investigate the performance of the proposed methodology, a generation of experimental data is indispensable. This step allows the creation of data, which true value is known, and by consequence also allows a characterization of the accuracy and precision of measurements.

To compare on-line abnormalities detection techniques, data generated in the previous step as been used. The two tests compared in this study are : the Hampel test and the method based on the characteristics of wavelet first order transform. The application of both techniques conclusion is : utilization of wavelet transform is more suitable to detect long period abnormalities. This statement can be explain by the 50% break point value associated with the Hampel test.

A 3 steps online steady state identification methodology as also been elaborated during this study. First, the starting point of the steady state period is detected with the help of wavelet transform and is first derivative characteristics, then steady state data are filtered and finally, the steady state final point is detected by looking at the first order wavelet transform modulus. The application of the new method is compared to the utilization of two other techniques used to identify, on-line, steady state periods. The application of the 3 steps method allows the improvement of steady state identification efficiency by creating an amount of type I and II errors that is lower than the two other techniques.

Later, a comparative study around different combination of techniques, applicable at the different steps of the general methodology, is performed. The comparison of accuracy and precision improvement in measurements by applying each combination allows the identification of the following arrangement: utilization of wavelet first order transform

to detect abnormalities, filtration achieved by applying multi-scale wavelet processing and steady state identification by the novel 3 steps method, as the most suitable combination of techniques. After passing the data through those 3 steps, steady state data are reconciled by using the Sigmafine software from OsiSoft Inc.

Concerning the aspect of precision improvement, the comparison of the different combination of techniques does not point out any arrangement as the most appropriate one to use. This can be explain by the fact that precision calculation is highly affected by the presence of gross errors. Since all the combination are using the same technique of detection of gross errors, precision is relatively improved the same way by one or the other arrangement.

Finally, the combination of techniques providing optimal results is integrated to the general methodology and is applied, on-line, in order to treat real-time data extracted from a papermachine stock approach system. The application of the proposed methodology allows the improvement of measured data quality and the identification (and correction) of offset and total failure of instruments. Those results demonstrate the robustness of the method and it's possible industrial implementation.

Results, obtained in this study, can be used to concluded that the proposed methodology is robust and improves the accuracy and precision of measurements. Also, it as been proven that the abnormality detection method based on wavelet transform is more adequate, for on-line application, then the utilisation of Hampel test. In addition, this study concludes that a novel 3 steps method based on wavelet first order transforms, statistical theories and filtration techniques is more efficient to identify, on-line, steady state periods around a process than other techniques.

To conclude, the implementation of the proposed methodology to treat data, on-line, will provide to the mills the possibility to develop a list of reliable quasi-steady state regimes

of operation of their processes. This implementation can imply multiple potential benefices :

- Efficient detection of miscalibrated instruments
- Improvement of process control
- Detailed knowledge of the different process regimes of operation
- Identification of local and global process trends
- Identification of leaks and accumulation in the process
- Improvement of planning and optimization related to the process operation.

Table des Matières

DÉDICACE ..... **IV**

REMERCIEMENTS ..... **V**

RÉSUMÉ ..... **VI**

ABSTRACT ..... **XI**

TABLE DES MATIÈRES ..... **XVI**

LISTE DES TABLEAUX ..... **XXIII**

LISTE DES FIGURES ..... **XXIV**

LISTE DES SIGLES ET DES ABRÉVIATIONS ..... **XXVIII**

LISTE DES ANNEXES ..... **XXX**

CHAPITRE 1. INTRODUCTION **1**

CHAPITRE 2. REVUE DE LA LITTÉRATURE: QUALITÉ DES DONNÉES ET TYPES D'ERREURS **3**

2.1 TYPES D'ERREURS 3

2.1.1 Erreur aléatoire 4

2.1.2 Anormalité 5

2.1.3 Erreur brute 6

2.2 QUALITÉ DES DONNÉES 7

2.2.1 Exactitude 8

2.2.2 Précision 9

2.2.3 Robustesse 11

CHAPITRE 3. REVUE DE LITTÉRATURE: TRAITEMENT DES ERREURS 12

3.1 MÉTHODOLOGIE GÉNÉRALE UTILISÉE POUR TRAITER LES DONNÉES 12

3.2 TECHNIQUES UTILES POUR LA CORRECTION DES ERREURS EN LIGNE 13

3.2.1 La théorie des ondelettes 13

3.2.1.1 Découper le signal en niveau 13

3.2.1.2 Transformée d'Ondelettes de premier et second ordre 15

3.2.2 Détection des anormalités 16

3.2.2.1 Détection par test statistique 16

3.2.2.2 Détection à l'aide des ondelettes 18

3.2.3 Filtration de données 21

3.2.4 Détection du régime permanent 24

3.2.4.1 Tests statistiques –Approche de Cao et Rhinehart 25

3.2.4.2 Technique basée sur les ondelettes 28

3.3 RÉCONCILIATION DE DONNÉES 31

3.3.1 Classification des variables 31

3.3.2 Détection des erreurs brutes 32

3.3.3 Réconciliation en régime permanent 35

3.3.3.1 Problème de minimisation 36

3.3.3.2 Sigmafine 37

3.3.3.3 Effet des mesures indirectes sur la supposition de normalisation 37

CHAPITRE 4. SYNTHÈSE 38

4.1 RETOUR SUR LES CONCEPTS IMPORTANTS 38

4.1.1 La qualité des données 38

4.1.2 La méthodologie générale 38

4.1.3 La détection des anormalités 39

4.1.4 Filtration de données 39

4.1.5 Détection du régime permanent 40

4.1.6 La réconciliation des données 42

4.2 PROBLÈMES RELIÉS AU TRAITEMENT DES DONNÉES EN LIGNE 43

4.2.1 Relié à la méthodologie générale 43

4.2.2 La détection des anormalités 43

4.2.3 La détection du régime permanent 44

4.2.4 Relié au domaine des pâtes et papier 44

CHAPITRE 5. PROBLÉMATIQUE ET MÉTHODOLOGIE DE RECHERCHE 45

5.1 PROBLÉMATIQUE 45

5.2 HYPOTHÈSES DE RECHERCHE 46

5.3 SUPPOSITIONS 47

5.3.1 Normalité des erreurs aléatoires 47

5.3.2 Mesures effectuées directement sur les variables mesurées 48

5.4 MÉTHODOLOGIE DE RECHERCHE 48

5.4.1 Méthodologie exacte, précise et robuste 48

5.4.2 Efficacité de la détection des anormalités 50

5.4.3 Détection du régime permanent 52

5.4.4 Application de la méthodologie sur des données industrielles 54

5.4.4.1 Méthodologie utilisée 54

5.4.4.2 Description du procédé 54

CHAPITRE 6. MÉTHODOLOGIE POUR LA RÉCONCILIATION DE DONNÉES EN LIGNE 56

6.1 PRÉSENTATION DE LA MÉTHODOLOGIE 56

6.1.1 Méthodologie de Jiang et problèmes pour les applications en ligne 57

6.1.2 Méthodologie proposée 57

6.2 PRÉSENTATION DU PREMIER ARTICLE 59

6.3 ON-LINE PROCESSING AND STEADY STATE RECONCILIATION OF PULP AND PAPER MILL PROCESS DATA 59

6.3.1 Abstract 59

6.3.2 Introduction 60

6.3.3 Methodology 61

6.3.4 Case study 64

6.3.5 Process data analysis and creation of hypothetical "true" and "measured" data sets 65

6.3.5.1 Signal processing and near steady state operation regime identification 65

6.3.5.2 Steady state detection measurement correction through mass balances 65

6.3.5.3 Creation of hypothetical data with known errors 66

6.3.6 Investigation of Data Processing and Reconciliation Methodology Using the Hypothetical Data Sets 67

6.3.7 Evaluation of Methodology Robustness Using Real Time Process Measurements 70

6.3.8 Implications of results 71

6.3.9 Conclusions 74

CHAPITRE 7. DÉTECTION DES ANORMALITÉS 75

7.1 MÉTHODOLOGIE UTILISÉE POUR DÉMONTRER L'EFFICACITÉ DES MÉTHODES ÉTUDIÉES 76

7.2 RETOUR SUR LES MÉTHODES COMPARÉES 77

7.2.1 Méthode de Hampel 77

7.2.2 Méthode basée sur les caractéristiques propres aux ondelettes 78

7.3 RÉSULTATS ET DISCUSSION 78

7.3.1 Erreur de type I 79

7.3.2 Erreur de type II 81

CHAPITRE 8. IDENTIFICATION DES PÉRIODES DE RÉGIME PERMANENT 82

8.1 PRÉSENTATION DU SECOND ARTICLE 82

8.2 STEADY STATE IDENTIFICATION FOR ONLINE DATA RECONCILIATION BASED ON WAVELET TRANSFORM AND FILTERING 83

8.2.1 Abstract 83

8.2.2 Introduction 83

8.2.3 Elimination of random noise and abnormalities 88

8.2.3.1 Representing the process measurements features by wavelet multi-scale approach 88

8.2.3.2 Identification of abnormal measurements 89

8.2.3.3 Denoising based on wavelet multi-scale features and historical process data. 90

8.2.4 Steady state detection based on WT features, filtering and statistical theories 91

8.2.4.1 3 Steps methodology to perform steady state detection 91

8.2.4.2 Selecting the threshold values and filtering parameters for steady state detection 97

8.2.5. Case study : Application to real time papermachine data 99
8.2.5.1 Performance indices 99
8.2.5.2 Efficiency of the online steady state detection 100
8.2.5.3 Impact on online steady state data reconciliation 102
8.2.6. Conclusions 104

CHAPITRE 9. RÉCONCILIATION DES DONNÉES 106

9.1 CLASSIFICATION DES VARIABLES 106
9.2 DÉTECTION DES ERREURS BRUTES ET RÉCONCILIATION DES DONNÉES 108
9.2.1 Méthode itérative 108
9.2.2 Détection des erreurs brutes 109
9.2.3 Réconciliation à l'aide du logiciel Sigmafine 110
9.2.3.1 Caractéristiques du logiciel 110
9.2.3.2 Méthodologie utilisée 110
9.3 APPLICATION DE COMBINAISONS DE TECHNIQUES À LA MÉTHODOLOGIE GÉNÉRALE 111
9.3.1. Présentation des combinaisons étudiées 111
9.3.2 Résultats et discussion 112
9.3.2.1 Exactitude des différentes combinaisons de techniques candidates . 113
9.3.2.2 Précision des différentes combinaisons de techniques candidates 114

CHAPITRE 10. INTERPRÉTATION, ANALYSE ET DISCUSSION GÉNÉRALE 115

10.1 DISCUSSION RELIÉE À LA DÉTECTION DES ANORMALITÉS 115
10.2 DISCUSSION REILÉE À L'IDENTIFICATION DES RÉGIMES PERMANENTS 116
10.3 DISCUSSION ASSOCIÉE À LA DÉTERMINATION DE LA COMBINAISON OPTIMALE DE TECHNIQUES 118
10.4 DISCUSSION RELIÉE À L'APPLICATION DE LA MÉTHODE OPTIMALE À DES DONNÉES EN TEMPS RÉELS 119

CHAPITRE 11. CONCLUSIONS **121**

RÉFÉRENCES **125**

ANNEXES **129**

Liste des tableaux

Tableau 3. 1 : Résultats des simulations pour les essais standardisés de Davies et Gather 18

Liste des figures

Figure 2. 1 : Schématisation des différentes catégories d'erreurs 3
Figure 2. 2 : Exemple de distribution normale 5
Figure 2. 3 : Illustration d'anormalités mesurées 5
Figure 2. 4 : Représentation des quatre types d'erreur brute 7
Figure 2. 5 : Exactitude d'un essai 8
Figure 2. 6 : Précision d'une série d'essais 9
Figure 2. 7 : a) Série de données imprécises, b) Série de données précises 10
Figure 2. 8 : a) Série d'essais inexacts et précis, b) Série d'essais exacts et imprécis ... 10

Figure 3. 1 : Méthodologie utilisée pour traiter les différents types d'erreurs 13
Figure 3. 2 : Représentation de la décomposition d'un signal à l'aide des ondelettes ... 15
Figure 3. 3 : Représentation de la détection des anormalité à l'aide des ondelettes 20
Figure 3. 4 : Différence de résolution entre a) La fenêtre utilisée dans la méthode par transformée de Fourier b) La fenêtre utilisée par les méthodes basée sur les transformées d'ondelettes 23
Figure 3. 5 : Représentation graphique de la méthode de détection du régime permanent à l'aide des transformées d'ondelettes. 29
Figure 3. 6 : Étapes de résolution de la méthode compensatoire 34
Figure 3. 7 : Modélisation d'un cuvier à l'aide d'un courant entrant/sortant d'un nœud 35

Figure 4. 1 : Exemple d'utilisation de la méthode de Cao et Rhinehart 41

Figure 5. 1 : Méthodologie générale pour juger de la robustesse de la méthode implantée et son impact sur l'exactitude et la précision des résultats..... 50
Figure 5. 2 : Partie de la méthodologie générale reliée à la deuxième hypothèse 51
Figure 5. 3 : Les quatre tests utilisés pour comparer les deux techniques étudiées 52
Figure 5. 4 : Parties de la méthodologie générale reliée à la troisième hypothèse 53
Figure 5. 5 : Partie de la méthodologie générale reliée à la quatrième hypothèse 54
Figure 5. 6 : Système à l'étude 55

Figure 6.1 : Méthodologie proposée 58
Figure 6.2 : Types of errors in online measurements 61
Figure 6.3 : General methodology applied in this study to improve data quality 62
Figure 6.4 : Example of reconciliation problem around a splitter 63
Figure 6.5 : Paper machine stock approach system studied 64
Figure 6.6 : System status and regimes of operation identified 66
Figure 6.7 : Comparison between hypothetical and real time data for the same operation regimes 67
Figure 6.8 : Step by step results from the application of on-line processing and reconciliation methodology 68
Figure 6.9 : Summary of improvement in data accuracy from the application of the data treatment methodology 69
Figure 6.10 : Summary of improvement in data precision from the application of the data treatment methodology 70
Figure 6.11 : Examples of corrections required to real time data: 71
Figure 6. 12 : Types of Benefits Associated with the Application of the Proposed Methodology 72
Figure 6. 13: Example of Cumulated Process Information from Successive Online Data Reconciliation Results 73

Figure 7. 1 : Effet de la présence d'anormalités sur la détection du régime permanent et sur le résultat de la réconciliation 75
Figure 7. 2 : Quatre types d'anormalités 76
Figure 7. 3 : Taux d'erreurs de types I associé aux techniques de détection des anormalités 79
Figure 7. 4 : Effet de la présence de plus de 50% d'anormalités dans l'échantillon étudiée sur le test d'Hampel. 80
Figure 7. 5 : Taux d'erreurs de types II associé aux techniques de détection des anormalités 81

Figure 8. 1 : Inaccurate Estimation of Steady State Periods 84
Figure 8. 2 : Steady State Detection Problems Related to Method of Filtering Variances by a Low Pass Filter 86
Figure 8. 3 : Scaling Choice for On-Line Data Treatment 87
Figure 8. 4 : Multi-Scale Wavelet Representation of Real-Time Measurements 89
Figure 8. 5 : Using Wavelet Transform Features to Represent Process Trend 92
Figure 8. 6 : Detection of Steady State Starting Point Using WT and First Derivative 93
Figure 8. 7 : Problems Related to Detection of Slow Changes Trend over a Long Period 94
Figure 8. 8 : Representation of Deviation from Mean and Difference in Successive Data Points 95
Figure 8. 9 : Evolution of Deviation from Average Value (A) During Transient Periods 96
Figure 8. 10 : Type I error related to online steady state identification 101
Figure 8. 11 : Type II error related to online steady state identification 101
Figure 8. 12 : Accuracy improvement by the application of different online steady state detection techniques 103

Figure 8. 13 : Precision improvement by the application of different online steady state detection techniques 104

Figure 9. 1 : Étapes de la réconciliation des données 106
Figure 9. 2 : Méthode itérative pour détecter les erreurs brutes et les réconciliés 109
Figure 9. 3 : Représentation des sept combinaison de techniques étudiées 112
Figure 9. 4 : Exactitude associée à l'utilisation des différentes combinaison de techniques étudiées 114
Figure 9. 5: Précision associée à l'utilisation des différentes combinaison de techniques étudiées 114

Liste des sigles et des abréviations

α : Intervalle de confiance

β : Index permettant d'identifier le degré de permanence d'un signal ou un système

$\sum_{i \epsilon I_j} C_{j,i} \varphi_{j,i}$: Signal lisse

$\sum_{k \epsilon K_L} d_{j,k} \psi_{j,k}$: Signal détaillé

Def : Taux de défaillance

$\xi[x]$: Fonction de transfert lisse permettant le calcul de β

ER : Erreur relative globale

ER_i : Erreur relative au temps i

F : Variable mesurée

$\tilde{F}$: Variable corigée

$g(N, \alpha_N)$: Fonction de Davies et Gather

λ_1 : Paramètre d'ajustement pour la détection des anormalités à l'aide des ondelettes

λ_2 : Paramètre d'ajustement pour l'identification des régimes permanents à l'aide des ondelettes

MDA : Médiane de la déviation absolue

μ : Moyenne d'un échantillon

n : Nombre de mesures effectuées

n_{Def} : Nombre de défaillance recensée

$\varphi_{j,i}$: Fonction dyadique discrétisée de niveau j à la $i^{ème}$ mesure

$\psi_{j,k}$: Fonction d'ondelettes dyadique discrétisée de niveau j à la $i^{ème}$ mesure

Q : Matrice des covariances

R : Ratio permettant d'identifier les régimes permanents par la méthode de Cao

ρ_1 : Paramètre de pondération permettant de filtrer la moyenne dans la méthode de Cao

ρ_2 : Paramètre de pondération permettant de filtrer la déviation entre la mesure effectuée et la moyenne dans la méthode de Cao

ρ_3 : Paramètre de pondération permettant de filtrer l'estimation de la variance à l'aide de la différence entre 2 données successives au moment de la mesure

σ : Écart-type d'un échantillon

δ^2 : Estimation de la variance basée sur la différence entre deux mesures successives

$\delta_{f_i}^{\ 2}$: Estimation filtrée de la variance basée sur la différence entre deux mesures successives

T_1 : Seuil de tolérance d'acceptation des mesures anormales par la méthode des ondelettes

T_2 : Seuil de tolérance de $Wf(t)$ pour l'identification de la durée d'une anormalité

T_u : Seuil de tolérance associé à l'identification des régimes permanents

t_p : Limite supérieure de durée d'une anormalité utilisé dans la méthode des ondelettes

$\theta(t)$: Facteur représentant l'influence des transformées d'ondelettes

v^2 : Déviation entre la mesure effectuée et la moyenne de l'échantillon

$v_{f_i}^{\ 2}$: Déviation filtrée entre la mesure effectuée et la moyenne de l'échantillon

Var : Variance

VM_i : Valeur mesurée au temps i

VC_i : Valeur ciblée au temps i

WT : Transformée d'ondelettes

$W_j f(t)$: Transformée d'ondelettes de premier ordre de la fonction f(t).

$Wf(p_i)$: Transformée d'ondelettes de premier ordre au point i.

$WW_j f(t)$: Transformée d'ondelettes de deuxième ordre de la fonction f(t).

$\bar{x}$: Moyenne des mesures

$\tilde{x}$: Médiane

Liste des annexes

ANNEXE I : Régimes d'opérations du procédé 129
ANNEXE II : Données expérimentales 132
ANNEXE III : Détection des anormalités 135
ANNEXE IV : Niveau de coupure associée à la filtration des données 140
ANNEXE V : Identification des régimes permanents 142
ANNEXE VI : Résultats de la réconciliation 151
ANNEXE VII : Publications 168

Chapitre 1. Introduction

En industrie, les mesures prises sur un procédé sont généralement contaminées par plusieurs sources d'erreurs : bruit associé aux différents capteurs utilisés, mauvaise calibration des instruments de mesure, variation dans les conditions ambiantes et erreurs humaines. La présence de ces erreurs est amplifiée par l'inconsistance entre les valeurs mesurées et les bilans de masse et d'énergie reliés au procédé (Bazin 1998). Cette constatation rend l'utilisation directe des données brutes inefficaces. Afin de palier à ces problèmes, on utilise de plus en plus, depuis plusieurs années, des techniques de filtration et de réconciliation des données (Bagajewicz 1999).

En plus de rendre les données consistantes aux lois de conservation de la masse et de l'énergie, une élimination adéquate des erreurs, présentes dans les mesures, permet d'augmenter l'efficacité des tâches de haut niveau reliées à l'opération des procédés chimiques : configuration des systèmes de contrôle, détection et diagnostic des fautes directement reliées à l'opération du procédé (fuites ou dépôts), planification et ordonnancement des opérations du procédé. La complexification et la spécificité grandissante des procédés chimiques obligent une automatisation de plus en plus importante de ces tâches directement dépendantes d'une interprétation adéquate des tendances associées au procédé étudié. L'extraction qualitative et quantitative automatisée des caractéristiques du procédé devient donc une étape essentielle pour l'amélioration éventuelle de ces tâches (Bakshi et Stephanopoulos 1995). Par conséquent, le traitement de données en ligne, se veut une étape cruciale, permettant l'utilisation de données propres dans l'automatisation éventuelle de ces différentes tâches.

L'industrie des pâtes et papier est l'une des branches vieillissantes de l'industrie chimique. Ce vieillissement des procédés entraîne des problèmes d'opérations et de production dans les différentes usines. Les différents acteurs de l'industrie soulèvent le

problème et pointent du doigt l'utilisation peu efficace des informations détenues. Par conséquent, depuis quelques années, les usines de pâtes et papier utilisent de plus en plus un système de gestion des données couplé à un système de gestion de l'information. Selon le sondage mené par Janssen et al. (2004a), le personnel des usines reconnaît que l'implantation de ces systèmes d'informations améliore grandement l'accessibilité des données en temps réel ainsi que l'opportunité de les utiliser afin de régler les problèmes reliés au procédé. Cependant, jusqu'à maintenant, avant d'être utilisées pour dépanner les problèmes reliés à l'opération d'un procédé, ces données sont analysées de manière impromptue. Ce type d'analyse ne permet malheureusement pas de tirer totalement profit des systèmes d'informations mis en place. Par conséquent, l'implantation d'une méthodologie, permettant la correction des erreurs reliées aux mesures en ligne, améliorerait l'efficacité de ces systèmes.

Une méthodologie permettant l'automatisation de la correction des erreurs, présentes dans les données mesurées, permettrait l'apparition éventuelle d'une automatisation efficace des tâches de haut niveau dans l'industrie chimique. Plus spécifiquement, cette méthodologie permettrait à l'industrie des pâtes et papiers de tirer totalement profit des données présentes dans leur système de gestion de l'information et par conséquent, ceci devrait améliorer l'efficacité des tâches décisionnelles à l'aide de modifications apportées au procédé.

Chapitre 2. Revue de la littérature : qualité des données et types d'erreurs

2.1 Types d'erreurs

Plusieurs variables influencent la performance des instruments de mesures. Les variations dans les conditions ambiantes, telles que la température, la pression et l'humidité, affectent la précision des mesures et/ou introduisent des biais. Cependant, si on pouvait améliorer les conditions ambiantes de manière optimale, les séries de données mesurées contiendraient encore des erreurs. En effet, en plus des conditions ambiantes, les mesures effectuées sur un procédé sont influencées par l'efficacité du câblage utilisé, la qualité de la calibration et la probabilité de dysfonction des instruments, ainsi que l'encrassement et la corrosion des capteurs. Afin de caractériser ces différents types d'erreurs, on les regroupes sous trois différentes catégories : les erreurs aléatoires, les anormalités (observations aberrantes) et les erreurs brutes (erreurs systématiques). Ces trois types d'erreurs sont illustrés à la figure 2.1.

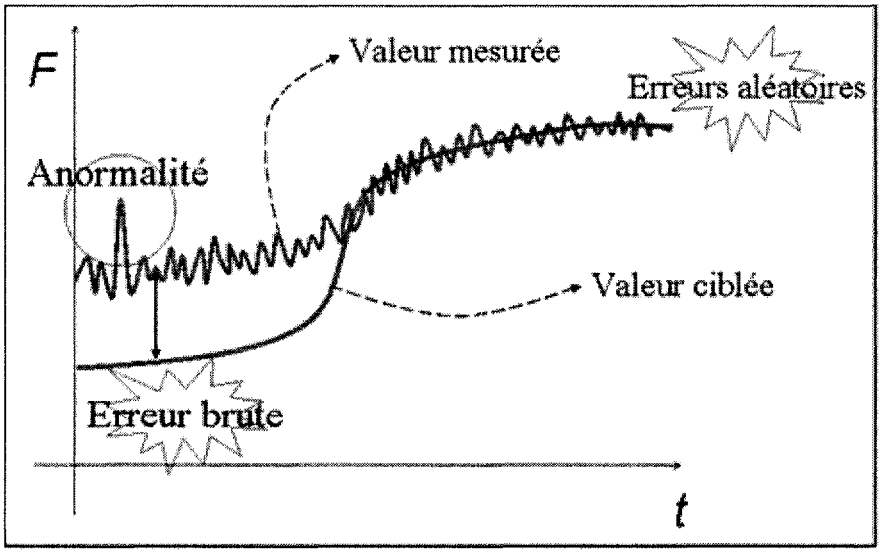


Figure 2. 1 : Schématisation des différentes catégories d'erreurs

2.1.1 Erreur aléatoire

Cette première catégorie d'erreurs regroupe les erreurs reliées aux variations dans les conditions ambiantes, au problème de câblage, aux fluctuations dans le courant électrique, à la vibration des éléments de procédé (ex : pompe, compresseur) et à l'étape de conversion du signal qui a lieu entre le capteur et l'instrument de mesure finale. Ces erreurs sont surtout concentrées dans les hautes fréquences et sont dites non reproductibles. C'est-à-dire qu'il est impossible de prédire leur amplitude et/ou leur signe.

Puisque les erreurs aléatoires sont causées par plusieurs types d'erreurs au même moment et que ces erreurs sont non reproductibles, il est possible de les associés au théorème central limite. Ce théorème statistique stipule que la somme d'un grand nombre de perturbations, possédant chacune ses propriétés propres, a tendance de se rapprocher d'une distribution normale. Par conséquent, une supposition liée à la distribution des normales de ces erreurs est généralement acceptée (Bagajewicz 2001).

Une distribution normale est caractérisée par une distribution en forme de cloche, symétrique par rapport à sa moyenne, dont la fréquence d'occurrence est élevée au centre et faible aux extrémités (Figure 2.2). Ce type de distribution possède une moyenne (μ) qui se situe au centre de sa distribution et une variance (σ^2). Les points auxquels la courbe de la cloche passe d'une forme concave descendante à une forme convexe ascendante se situent à $\mu+\sigma$ et $\mu-\sigma$ (Turner et McElrath 1996). Cette distribution possède également, au niveau des probabilités d'occurrence, des propriétés spécifiques intéressantes:

- 68.26% des mesures sont situées à plus ou moins 1 écart-type de la moyenne
- 95.44% des mesures sont situées à plus ou moins 2 écart-types de la moyenne
- 99.72% des mesures sont situées à plus ou moins 3 écart-types de la moyenne

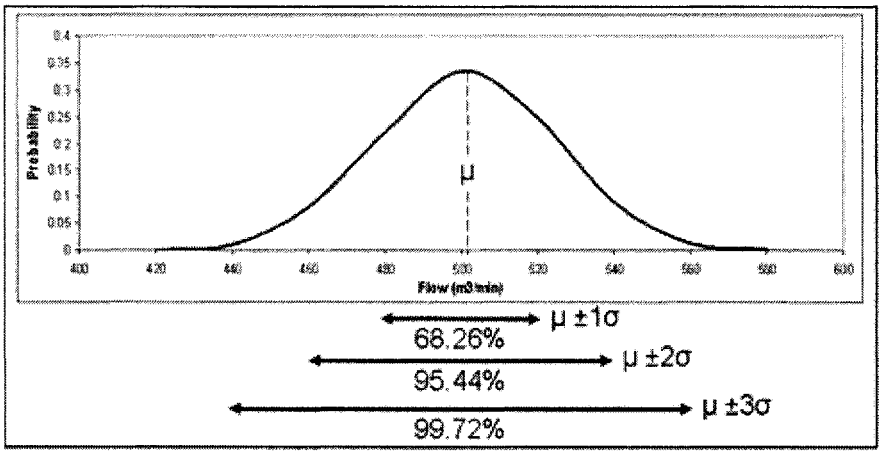


Figure 2. 2 : Exemple de distribution normale

2.1.2 Anormalité

Les erreurs appartenant à cette catégorie sont principalement attribuables à une surcharge électrique, à des problèmes de câblage, à une dégradation des instruments de mesure ou à des erreurs humaines. Une anormalité est définie comme étant une erreur de haute fréquence et de haute amplitude qui ne fait pas partie de la distribution caractérisant la majorité des mesures effectuées. Elle ne peut donc pas être calculée, estimée ou anticipée. On fait donc référence à une mesure qui possède les mêmes caractéristiques de non reproductibilité et de fréquence élevée que les erreurs aléatoires. Par conséquent, cette erreur est caractérisée d'aléatoire et de supernaturelle. Cependant, puisqu'elle est par définition située à l'extérieur de la distribution qui caractérise la majorité des données recueillies, cette mesure n'est pas associée à une supposition de distribution normale (Chen et Romagnoli 1997).

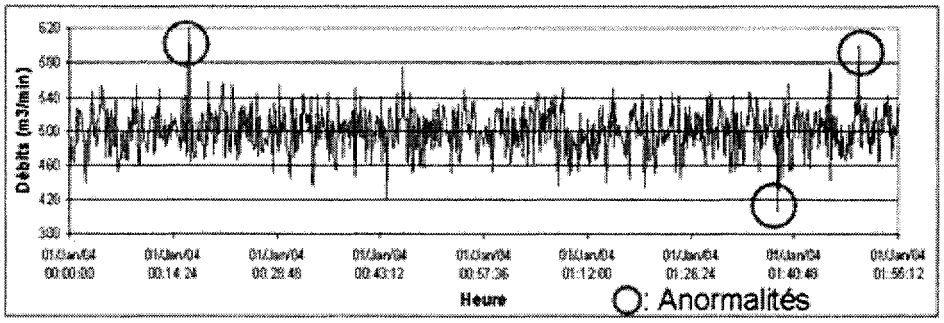


Figure 2. 3 : Illustration d'anormalités mesurées

2.1.3 Erreur brute

Ces erreurs se retrouvent dans une gamme de fréquences beaucoup plus faible que celles appartenant aux 2 premières catégories présentées précédemment. Elles sont surtout associées à une inconsistance physique dans une série de données (loi de la conservation de matière, bilan d'énergie ...). Les erreurs brutes sont attribuables à deux familles différentes de causes : celles reliées aux instruments de mesures et celles inhérentes au procédé. Les erreurs appartenant à la première catégorie sont causées par une mauvaise calibration, le dysfonctionnement, l'encrassement ou la corrosion des instruments utilisés. Celles reliées au procédé sont plutôt associées à des dépôts solides ou à des fuites dans le système.

Selon Narisimhan et Jordache (2000), il existe quatre types d'erreurs brutes (Figure 2.4) :

A. *Le Biais* : erreur constante et de longue durée pouvant être expliquée par différentes causes dont une installation inadéquate des instruments de mesures.
B. *La défaillance totale* : aucune réponse transmise par l'instrument utilisé.
C. *Le phénomène de dérive* : erreur provoquée par une détérioration de la calibration des instruments.
D. *La dégradation de la précision* : erreur de grande amplitude variante présente sur de longue durée qui peu être provoquée par la détérioration de l'équipement.

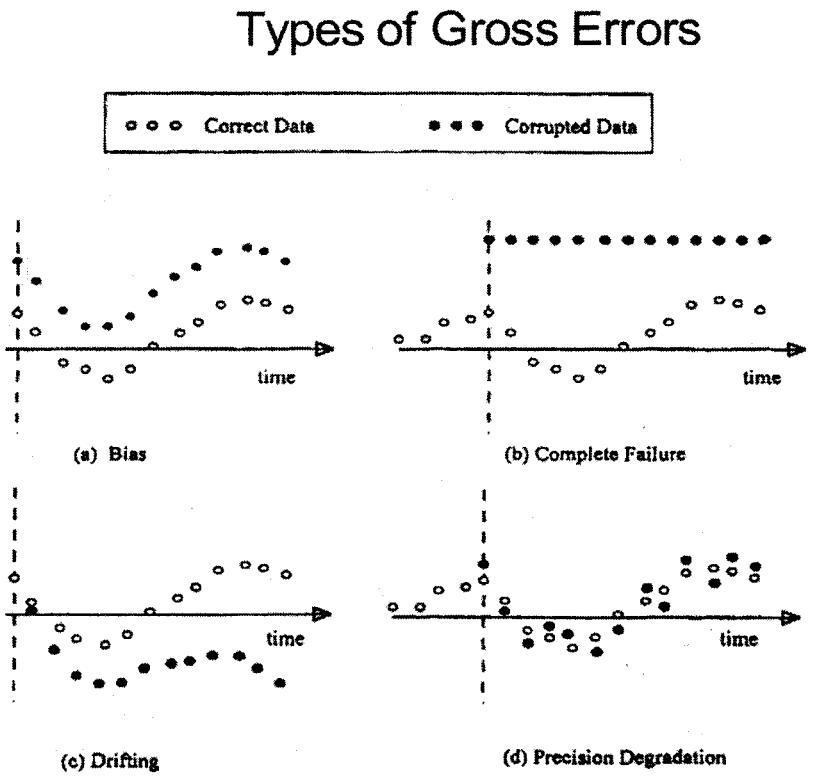


Figure 2. 4 : Représentation des quatre types d'erreur brute
(Figure de Narisimhan et Jordache 2000)

2.2 Qualité des données

La qualité des données est un terme général qui est souvent utilisé pour faire référence à une liste de propriétés que devraient détenir une série de données idéales. Afin de mieux cerner ce concept Bagajewicz (2001) défini la qualité des données. Selon cette définition, les propriétés permettant de juger de la qualité des données sont toutes regroupées dans les trois caractéristiques suivantes : l'exactitude, la précision (reproductibilité) et la robustesse (confiance).

2.2.1 Exactitude

L'exactitude est définie comme étant la qualité de ce qui ne contient pas d'erreur. On fait donc référence à la capacité à atteindre une valeur cible donnée (Figure 2.5). Lorsque l'on fait référence à une mesure effectuée sur un procédé, cette définition ce résume à la capacité d'un instrument de mesurer la vraie valeur de la variable étudiée.

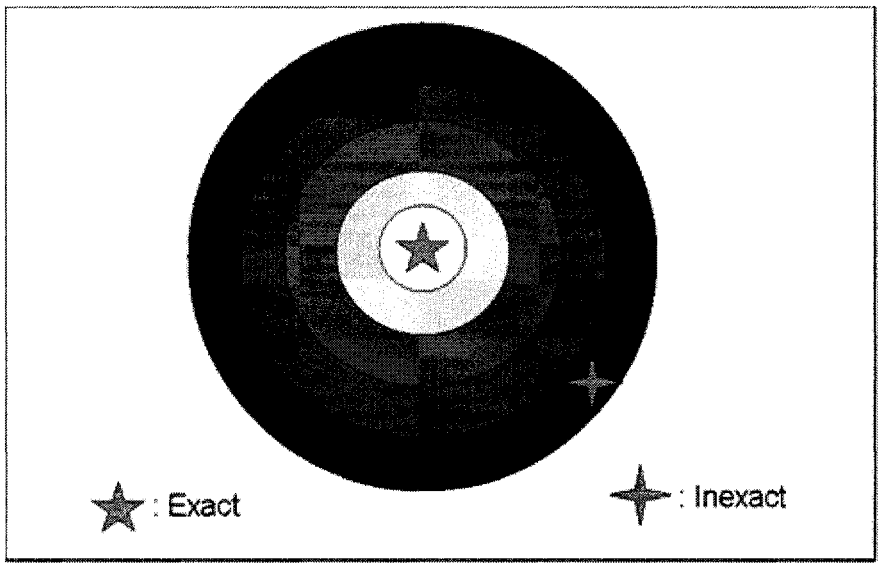


Figure 2. 5 : Exactitude d'un essai

Afin de quantifier l'exactitude d'une mesure à un moment donné, on se réfère à l'erreur relative entre la valeur mesurée et la valeur ciblée (vraie valeur). Plus l'erreur relative est faible, plus la mesure effectuée est exacte. Lorsque l'erreur relative est nulle, la mesure effectuée est dite parfaite.

$$ER_i = 100\% * \frac{|VM_i - VC_i|}{|VC_i|} \quad (1)$$

Dans le cas d'une série de données, on utilise la moyenne des différences relatives afin de quantifier l'exactitude d'un instrument ou d'une méthode utilisée.

$$ER = \frac{\sum_{i=1}^{n} ER_i}{n} \quad (2)$$

2.2.2 Précision

La précision est définie comme étant la capacité à reproduire une valeur à l'intérieur d'un intervalle prédéterminé (Figure 2.6). En appliquant cette définition à un capteur, la précision représente la proximité à laquelle se situe l'ensemble des mesures consécutives d'une série comparativement à la valeur d'une variable qui demeure statique (Figure 2.7).

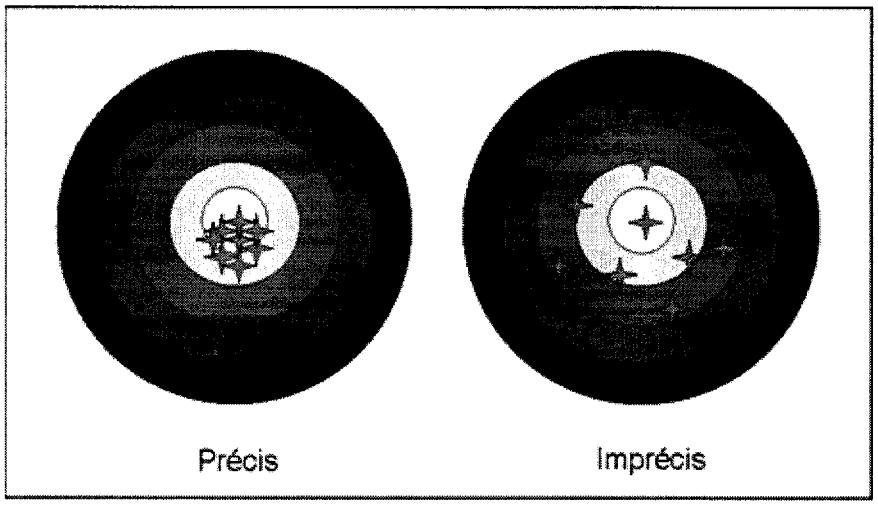


Figure 2. 6 : Précision d'une série d'essais

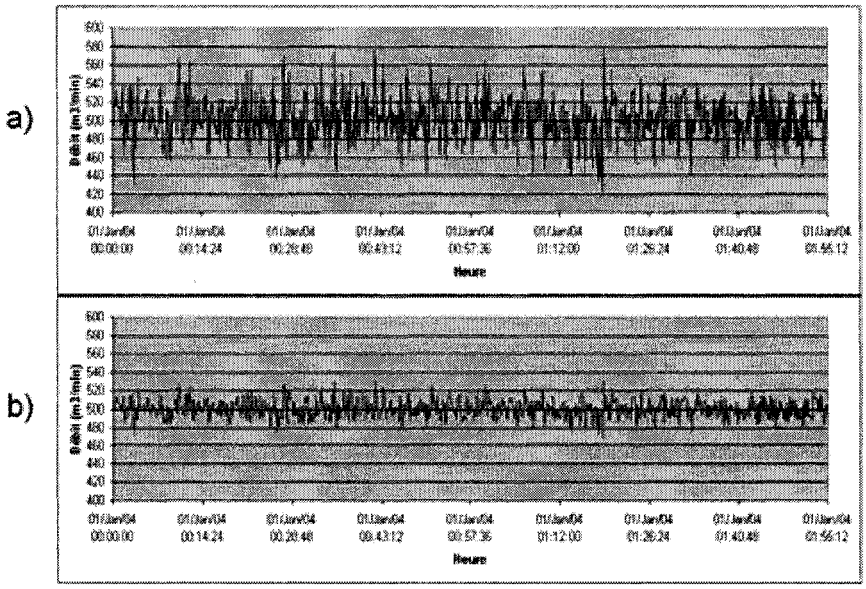


Figure 2. 7 : a) Série de données imprécises, b) Série de données précises

Il est à noter que cette caractéristique ne fait pas référence à la valeur ciblée (vraie valeur) et à la variable mesurée. Il est donc possible qu'un instrument soit précis et au même moment inexact. Cette situation se produit lorsque, pour une situation donnée, les mesures effectuées par un capteur se concentrent dans un petit intervalle, mais sont éloignées de la valeur ciblée (Figure 2.8 a). Inversement, un instrument peut être exact tout en étant imprécis. C'est le cas lorsque la moyenne des mesures effectuées est très près de la valeur ciblée, mais que la variance entre ces mesures est élevée (Figure 2.8 b).

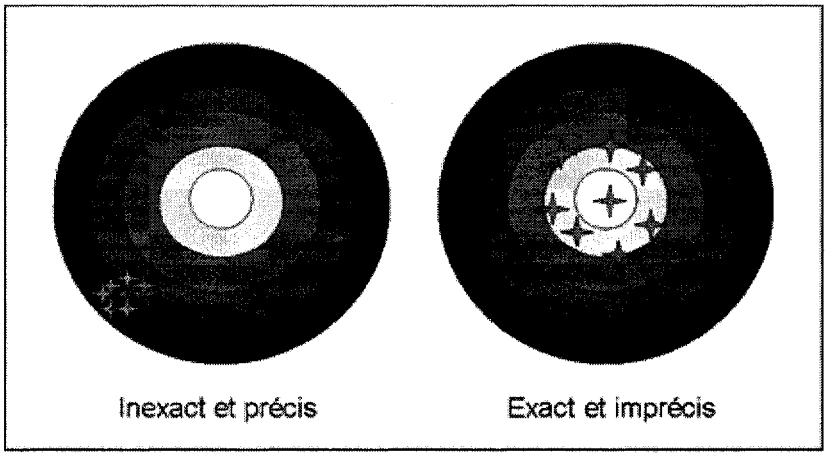


Figure 2. 8 : a) Série d'essais inexacts et précis, b) Série d'essais exacts et imprécis

Afin de quantifier la précision d'un instrument, dans une situation donnée, on utilise la variance des mesures effectuées par le capteur lors de la période étudiée. En supposant que la distribution des erreurs autour de la moyenne est normale, il est possible de calculer la variance à l'aide de l'équation suivante :

$$Var = \frac{\sum_{i=1}^{n} (VM_i - \bar{x})^2}{n-1} \qquad (3)$$

2.2.3 Robustesse

La robustesse représente la possibilité qu'une donnée soit présente lors d'un intervalle de temps prédéterminé. Les termes : robustesse (confiance) et disponibilité sont souvent confondus. La disponibilité représente plutôt la probabilité qu'une donnée ne soit pas absente à un moment donné. La robustesse est donc une caractéristique plus stricte que la disponibilité. Afin de quantifier la robustesse d'un instrument, on calcule son taux de défaillance. Plus la proportion de défaillance est faible, plus l'instrument est robuste.

$$Def = 100\% * \frac{n_{Def}}{n} \qquad (4)$$

Chapitre 3. Revue de littérature : Traitement des erreurs

Afin de tirer le maximum d'informations pertinentes des mesures qui sont effectuées sur un procédé, il est nécessaire de corriger les erreurs qui s'y sont glissées. Pour se faire, différentes techniques de traitement de données sont utilisées. Pour chaque classe d'erreurs présentée dans le deuxième chapitre, un différent type de technique sera favorisé. Les erreurs aléatoires sont filtrées, les anormalités sont détectées, tandis que pour éliminer les erreurs brutes ont fait appel aux techniques de réconciliation de données.

3.1 Méthodologie générale utilisée pour traiter les données

Afin de traiter les données de manière efficace, Jiang (2003b) utilise une méthodologie basée sur la correction des erreurs en plusieurs étapes (Figure 3.1). Dans un premier temps, les données extraites du procédé sont analysées. Cette première étape permet d'identifier les anormalités présentes dans les mesures. Une fois identifiés, les points anormaux sont corrigés en remplaçant la valeur supprimée par une valeur corrigée. Cette valeur peut être prédite à l'aide des données précédentes, peut être corrigée par une technique de symétrie (favorisé pour les variables contrôlées) ou encore par une valeur moyenne. Par la suite, afin d'éliminer les erreurs aléatoires, les données normalisées sont filtrées. Théoriquement, la série de données issue de ces étapes ne devrait plus contenir d'anormalités et d'erreurs aléatoires. Le seul type d'erreur présent dans cette série devrait être des erreurs brutes. Par conséquent, les dernières étapes de la méthodologie utilisée par Jiang (2003b) permettent de réconcilier les données. Pour se faire, une étape de détection du régime permanent permet de choisir les données qui seront réconciliées.

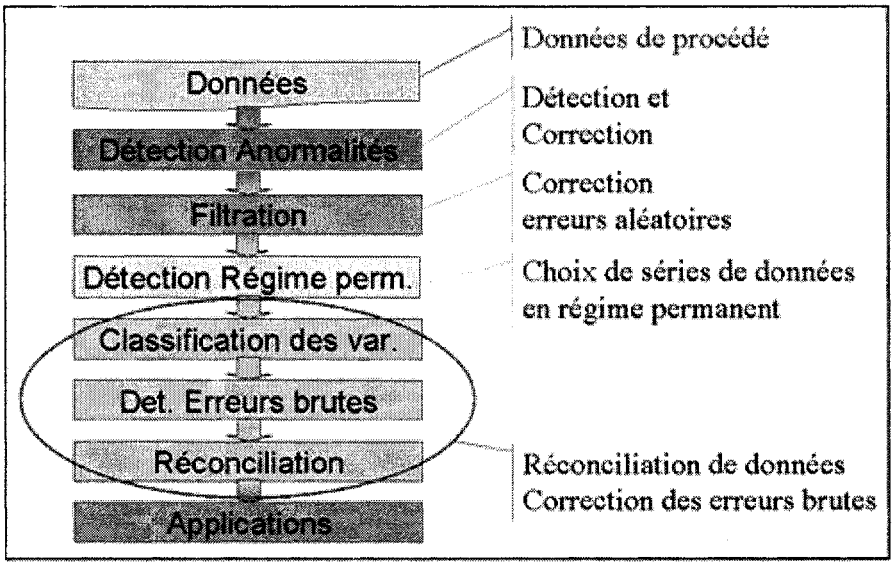


Figure 3. 1 : Méthodologie utilisée pour traiter les différents types d'erreurs

3.2 Techniques utiles pour la correction des erreurs en ligne

Pour chacune des étapes de la méthodologie décrite à la section précédente, plusieurs techniques spécifiques ont été développées. Certaines de ces techniques sont applicables en ligne. Les plus importantes et les plus utilisées seront présentées dans cette section.

3.2.1 La théorie des ondelettes

Les ondelettes sont très utilisées dans le domaine du traitement des signaux et des images. Elles permettent de représenter le signal ou l'image de manière détaillée.

3.2.1.1 *Découper le signal en niveau*

En pratique, les mesures prises sur un système contiennent des caractéristiques présentes dans plusieurs domaines de fréquences. Ces caractéristiques sont causées par des perturbations ou sont inhérentes au procédé. Pour bien représenter ces caractéristiques, on utilise les ondelettes. L'idée principale, à la base des ondelettes, est de représenter la

série de mesures en une suite d'approximations successives de données appelées : niveaux.

$$f(t) = f_0 = \sum_{i\varepsilon I_0} C_{0,i}\varphi_{0,i} = \sum_{i\varepsilon I_1} C_{1,i}\varphi_{1,i} + \sum_{k\varepsilon K_1} d_{1,k}\psi_{1,k}$$
$$= \sum_{i\varepsilon I_2} C_{2,i}\varphi_{2,i} + (\sum_{k\varepsilon K_1} d_{1,k}\psi_{1,k} + \sum_{k\varepsilon K_2} d_{2,k}\psi_{2,k})$$
$$= \ldots$$
$$= \sum_{i\varepsilon I_j} C_{j,i}\varphi_{j,i} + (\sum_{k\varepsilon K_j} d_{j,k}\psi_{j,k} + \sum_{L=1}^{j}\sum_{k\varepsilon K_L} d_{L,k}\psi_{L,k})$$
$$= \sum_{i\varepsilon I_J} C_{J,i}\varphi_{J,i} + (\sum_{j=1}^{J}\sum_{k\varepsilon K_L} d_{j,k}\psi_{j,k})$$

(Équation 5 : tirée de Jiang 2003a)

Où $\varphi_{j,i}$ est la fonction dyadique discrétisée de niveau j et $\psi_{j,k}$ est la fonction d'ondelettes dyadique discrétisée de ce niveau. Le $\sum_{i\varepsilon I_j} C_{j,i}\varphi_{j,i}$ est nommé : signal lisse et contient les éléments de faibles fréquences présents au niveau j. Tandis que $\sum_{k\varepsilon K_L} d_{j,k}\psi_{j,k}$ porte le nom de signal détaillé et contient les éléments de hautes fréquences présents à ce niveau.

Les niveaux sont utiles pour représenter les caractéristiques du signal, car ils ne contiennent que les mesures présentent sous une fréquence spécifique. Ainsi, les différences entre les signaux lisses de deux niveaux successifs permettent d'identifier la contribution de différentes sources à la distribution du signal (Bakshi et Stephanopoulos 1995).

En résumé, les ondelettes permettent la décomposition du signal en plusieurs niveaux. Elles sont associées à une représentation des oscillations contenues dans un petit espace et sur une courte période de temps. Ce qui provoque une représentation du signal permettant de bien localiser ses différentes caractéristiques, autant dans le temps que dans les fréquences (Mitra et Acharya 2003) (Figure 3.2).

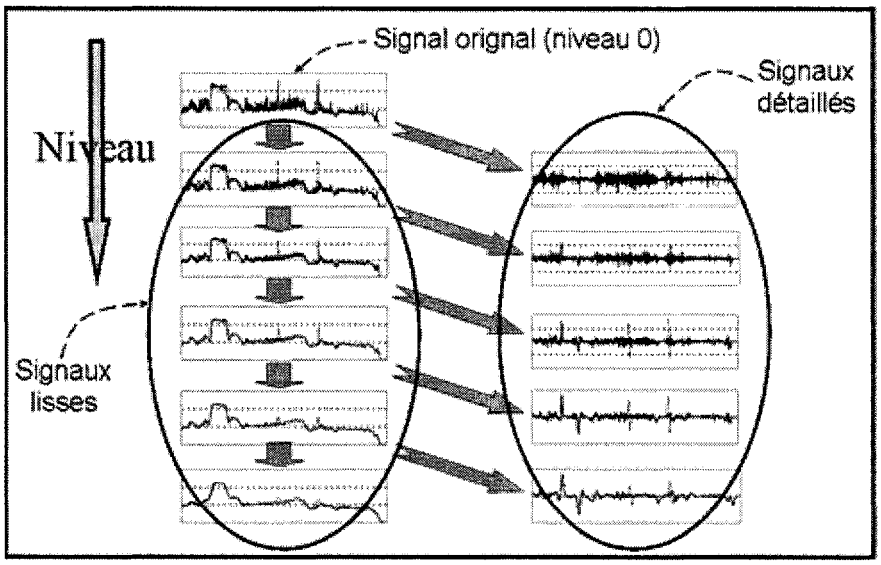


Figure 3. 2 : Représentation de la décomposition d'un signal à l'aide des ondelettes

3.2.1.2 *Transformée d'Ondelettes de premier et second ordre*

Lorsque $\psi_{j,k}$ est dite de premier ordre, elle représente la première dérivée de la fonction de niveau j (Mallat et Hwang 1992). Par conséquent, la transformée d'ondelettes de premier ordre d'une fonction donnée est proportionnelle à la dérivée de cette fonction lissée par φ_j (7).

$$\psi_j(t) = \frac{d\varphi_j(t)}{dt} \qquad (6)$$

$$W_j f(t) = f * \varphi_j(t) = f * (2^j \frac{d\varphi_j}{dt})(t) = 2^j \frac{d}{dt}(f * \varphi_j)(t) \qquad (7)$$

La transformée d'ondelettes d'ordre 2 représente la transformée d'ondelettes de la transformée de premier ordre. Elle est donc proportionnelle à la seconde dérivée du signal :

$$WW_j f(t) = W_j f * \psi_j(t) = 2^{2j} \frac{d^2}{dt^2}(f * \varphi_j)(t) \qquad (8)$$

3.2.2 Détection des anormalités

Une des tâches importantes du traitement de données est de détecter et de remplacer, les anormalités présentes dans une série de mesures effectuées, par des données fiables. En effet, la présence d'un ou deux points aberrants pourraient corrompre les résultats obtenus à l'étape de réconciliation des données (Chen et Romagnoli 1997). Afin de détecter ces anormalités, plusieurs techniques différentes ont été mises au point. Cependant, ces techniques ne sont pas toutes applicables en ligne. Celles qui le sont, se divisent en deux grandes classes : les techniques directement basées sur les statistiques reliées à la distribution normale des erreurs aléatoires et celles qui exploitent la théorie des ondelettes.

3.2.2.1 Détection par test statistique

La majorité des méthodes permettant la détection des anormalités sont basées sur l'hypothèse de distribution normale, indépendante et identique des données. Puisque, selon cette supposition, 99.72 % des points aléatoires se situent à plus ou moins 3 écart-type (σ) de la moyenne (μ) et que les points anormaux ne font pas parties de cette distribution, il est possible de détecter ces points aberrants en identifiant les mesures qui n'appartiennent pas à l'intervalle $[\mu-3\ \sigma,\ \mu+3\ \sigma]$. Cependant, le calcul de la moyenne, peu être grandement influencé par la présence d'une (ou plusieurs) anormalité(s).

Afin de permettre de mesurer la robustesse d'un estimateur face à la présence d'anormalités, Hampel (1971) a proposé le concept du point de rupture. Ce concept représente le plus petit pourcentage d'anormalités devant être présent dans un échantillon pour provoquer une erreur substantielle de l'estimateur. Plus le point de rupture est élevé, plus l'estimateur utilisé est robuste. Puisque la présence d'une seule anormalité peu avoir une grande influence sur la moyenne calculée, celle-ci possède un point de rupture de 1/n.

Afin d'améliorer la détection des anormalités basée sur la moyenne et l'écart-type d'un échantillon, Hampel (1974) a proposé d'utiliser la médiane ($\tilde{x}$) et la médiane de la déviation absolue (MDA) de cette dernière. En effet, la médiane est une mesure de tendance centrale beaucoup plus résistante à la présence de points aberrants que la moyenne (Turner et McElrath 1996). Selon Hampel, cet estimateur possède un point de rupture de 50%. Ce qui confirme que la médiane est beaucoup plus robuste à la présence d'anormalités que la moyenne.

$$\tilde{x} = \begin{cases} x_i, i = \dfrac{n}{2} + 1 ....lorsque.n..est.pair \\ \dfrac{x_i + x_{i+1}}{2}, i = \dfrac{n}{2} ....lorsque.n..est..impair \end{cases} \qquad (9)$$

$$\text{MDA = médiane } (|x_1 - \tilde{x}|, |x_2 - \tilde{x}|, ..., |x_n - \tilde{x}|) \qquad (10)$$

Pour identifier la présence de points aberrants, on peu utiliser le test suivant :

$$\text{Si } (|x_i - \tilde{x}| \geq g(N, \alpha_N) * MDA), \text{ alors } x_i \text{ est un point aberrant} \qquad (11)$$

Ici la fonction $g(N, \alpha_N)$ est reliée au nombre de points et à la quantité d'erreurs de type I (pour plus de détails, voir Turner et McElrath, 1996). Davies et Gather (1993) ont effectués des travaux sur cette fonction. Ils ont simulé des essais standardisés sur le taux d'erreurs de type I, à l'aide du critère suivant :

$$P(OR(x_N, \alpha_N)) \subset outside(\alpha_N, \mu, \sigma^2) = 1 - \alpha \qquad (12)$$

$$\text{où } OR(x_N, \alpha_N) = [-\infty, L(x_N, \alpha_N)] \cup [R(x_N, \alpha_N)] \qquad (13)$$

Les résultats obtenus ont permis de déterminer la valeur de la fonction $g(N, \alpha_N)$ pour plusieurs échantillons de volumes différents en supposant une quantité d'erreurs de type I déterminée (5%).

Tableau 3. 1 : Résultats des simulations pour les essais standardisés de Davies et Gather

	N=20	N=50	N=100
Valeur de la fonction g(N,alpha)	5.67	4.31	3.81

Puisque la méthode proposée par Hampel (1974) est appliquée sur une série de données déterminées, une fenêtre de temps déroulante doit être couplée à cette technique afin de permettre de l'utiliser en ligne. L'utilisation de cette fenêtre permet de baser la détection des points aberrants sur les comportements récents du procédé. Malheureusement, il n'existe aucune règle universelle permettant de déterminer la dimension de cette fenêtre. La longueur choisie est un paramètre de réglage qui doit préalablement être défini à l'aide de données historiques (Cao et Rhinehart 1995). Si la fenêtre est trop longue, un décalage sera introduit et les points appartenant à un comportement antérieur seront pris en compte. Par contre, si la fenêtre de temps est trop courte, certaines anormalités seront tolérées (Chen and Romagnoli 1997).

3.2.2.2 Détection à l'aide des ondelettes

Contrairement aux techniques, basées sur la détection des points aberrants, présentées à la section précédente, les techniques de détection à l'aide d'ondelettes sont associées à une détection des changements anormaux dans une série de mesures. Pour ce faire, les méthodes conventionnelles, basées sur les séries de Fourier, ne sont pas adéquates pour analyser les procédés non stationnaires, si les changements à identifier sont rapides. Afin d'identifier ces changements, l'exploitation des caractéristiques uniques aux ondelettes s'avère une idée intéressante (Jiang et al. 2000).

Afin de détecter les anormalités à partir des ondelettes, Jiang et al. (2003a) se sont basé sur l'approche proposée par Bakshi et Stephanopoulos (1995) qui permet d'identifier les différentes caractéristiques d'une séries de données en s'appuyant sur l'information contenue dans les différents niveaux du signal traité par la méthode des ondelettes. À

l'aide des caractéristiques retrouvées dans des niveaux successifs, Jiang a proposé une méthode permettant d'identifier les changements anormaux dans la suite de mesure étudiée.

Les extremums de la dérivée d'un signal correspondent aux points d'inflexions de ce signal. En détectant les maximums et minimums locaux de la série de données, on identifie les changements soudains qui y sont reliés (Jiang et al. 2000). Selon l'équation (7), aux différents niveaux j, les extremums de $W_j f(t)$ correspondent aux points d'inflexion du signal lissé. Par conséquent, puisque $W_j f(t)$ est directement proportionnel à la première dérivée du signal, ces changements anormaux peuvent être détectées. En effet, en identifiant les différents couples d'extremums de grandes amplitudes et de signes opposés de $W_j f(t)$, on identifie les changements de haute intensité (Figure 2.3). Les anormalités sont détectées si les deux critères suivants sont atteints:

$$\begin{gathered} Wf(p_1) \text{ et } Wf(p_2) \geq T_1 \\ signe[Wf(p_1)] * signe[Wf(p_2)] = -1 \\ p_2 - p_1 \leq t_p \end{gathered}$$

où t_p représente la limite supérieure de durée d'une anormalité et T_1 est le seuil de tolérance de $Wf(t)$ pour l'identification d'une anormalité (Jiang et al. 2003a). Cette valeur seuil est donnée par :

$$T_1 = 3\lambda_1 w \qquad (14)$$

où w représente l'écart-type de $Wf(t)$ et λ_1 est un paramètre d'ajustement. Lorsque ce paramètre est égal à 1, on admet donc un quantité d'anormalité d'environ 0.3% (voir section 2.1.1).

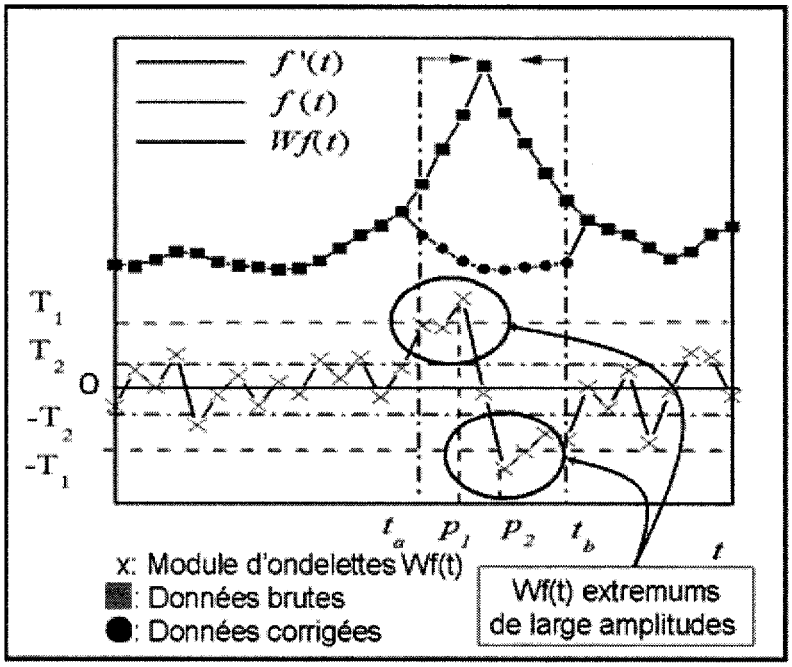


Figure 3. 3 : Représentation de la détection des anormalité à l'aide des ondelettes

Suite à l'identification de la présence d'anormalités dans une série de données, on détermine la durée [t_a t_b] de celle-ci. Pour se faire, on identifie t_a comme étant le point situé le plus près à gauche de p_1 et t_b comme étant le point situé le plus près à droite de p_2 de façon à satisfaire les exigences suivantes :

$$|Wf(t_a)| \text{ et } |Wf(t_b)| \geq T_2$$
$$|Wf(t_a - 1)| \text{ et } |Wf(t_b + 1)| < T_2$$

où T_2 est le seuil de tolérance de $Wf(t)$ pour l'identification de la durée d'une anormalité (Jiang et al. 2003a). Cette valeur seuil est donnée par :

$$T_2 = w \quad (15)$$

Afin d'identifier ces erreurs en ligne, Jiang et al. (2003a) proposent de suivre les trois étapes suivantes :

1 Déterminer les valeurs seuils (w, T_1 et t_p) hors-ligne à l'aide de données historiques.
2 Effectuer un traitement du signal à l'aide des ondelettes pour les niveaux 1 et 2.
3 Détecter les anormalités à l'aide de la réponse obtenue au niveau 2 et corriger les points aberrants à l'aide de la technique de symétrie périodique.

3.2.3 Filtration de données

Afin d'éliminer, en ligne, les erreurs aléatoires présentes dans une série de mesures, plusieurs techniques différentes ont été élaborées. Les filtres passes-bas (filtre LPF), digitaux et linéaires, comme les filtres basés sur une moyenne déroulante ou à lissage exponentiel, sont les méthodes les plus utilisées dans l'industrie chimique. Leur facilité d'utilisation en ligne et leur simplicité les rendent attrayants. Ces techniques rectifient le signal à l'aide d'une somme pondérée basée sur les mesures précédant la donnée à traiter.

Malheureusement, les méthodes de filtration linéaire représentent les données en un seul niveau de résolution. Puisque les séries de mesures sont généralement caractérisées de bruits appartenants à plusieurs fréquences différentes, l'utilisation de ces filtres force à faire un compromis entre la qualité de la rectification sur une longue période (distorsion du signal) et celle effectuée sur les caractéristiques locales du signal. Par conséquent, les filtres linéaires ne sont pas efficaces afin de filtrer les signaux contenant plusieurs caractéristiques localisées à différentes positions dans le domaine du temps et/ou de la fréquence (Bakshi et Stephanopoulos 1995), (Nounou et Bakshi 1999).

Les problèmes d'efficacité des filtres linéaires ont provoqué l'apparition des méthodes de filtration à plusieurs niveaux. Conceptuellement, ces méthodes décomposent la série

de données en plusieurs étapes. Pour se faire, elles utilisent des filtres possédant des caractéristiques différentes. Ce qui permet d'analyser la totalité des fréquences du bruit imprégné dans le signal à traiter. Dans cette catégorie de filtres à plusieurs niveaux, il y existe deux grandes classes de méthodes : celles basées sur les séries de Fourier, les autres utilisant les ondelettes. Selon Bakshi et Stephanopoulos (1995), les faiblesses de localisation temporelle des transformées de Fourier les rendent inadéquates pour effectuer des analyses à plusieurs niveaux. De l'autre côté l'utilisation d'ondelettes est efficace et permet de décomposer, à l'aide d'une dilatation de 2^j entre chaque niveau, n'importe quel signal qui est quadratiquement intégrable.

Comme, il a été démontré précédemment (section 3.1), les ondelettes représentent la série de données étudiées à l'aide d'une somme entre un signal lisse et la sommation de bruits de hautes fréquences, présent dans les niveaux inférieurs :

$$f(t) = \sum_{i \varepsilon I_J} C_{J,i} \varphi_{J,i} + (\sum_{j=1}^{J} \sum_{k \varepsilon K_L} d_{j,k} \psi_{j,k}) \quad (16)$$

Cette représentation permet d'identifier adéquatement les caractéristiques du signal et de diminuer le bruit présent dans la série de données, tout en minimisant la distorsion (Nounou et Bakshi 1999). En se basant sur la supposition de distribution Gaussienne du bruit, on peu traiter les mesures en trois étapes :

1 Décomposer le signal brut en un signal détaillé à l'aide des transformées d'ondelettes;
2 Éliminer les données qui possèdent une transformée d'ondelettes au-dessus d'une valeur seuil;
3 Construire un signal lisse et répéter la séquence jusqu'à ce que les changements présents dans le signal lisse soient jugés équivalent à ceux associés à l'opération du procédé.

Afin d'appliquer la technique d'ondelettes en ligne, la méthode propose d'effectuer les étapes précédentes sur une fenêtre déroulante dyadique de dimension 2^j (figure 3.4). À chacun des niveaux, ce type de fenêtre s'adapte afin de représenter les caractéristiques importantes du signal. Cette variation de la résolution est importante, car elle permet à la transformée d'ondelettes de zoomer sur les irrégularités d'une série de données et de les caractériser. Une fenêtre de courte longueur est automatiquement sélectionnée afin de représenter les changements rapides, alors qu'une fenêtre plus longue est favorisée pour identifier les changements plus lents.

En terminant, les travaux de Daubechies (1990), Bakshi et Stephanopoulos (1995), Nounou et Bakshi (1999) et Jiang (2000) concluent, sur des bases mathématiques et expérimentales, que l'utilisation d'ondelettes est plus efficace que l'utilisation des séries de Fourier et des filtres linéaires pour traiter les signaux en ligne.

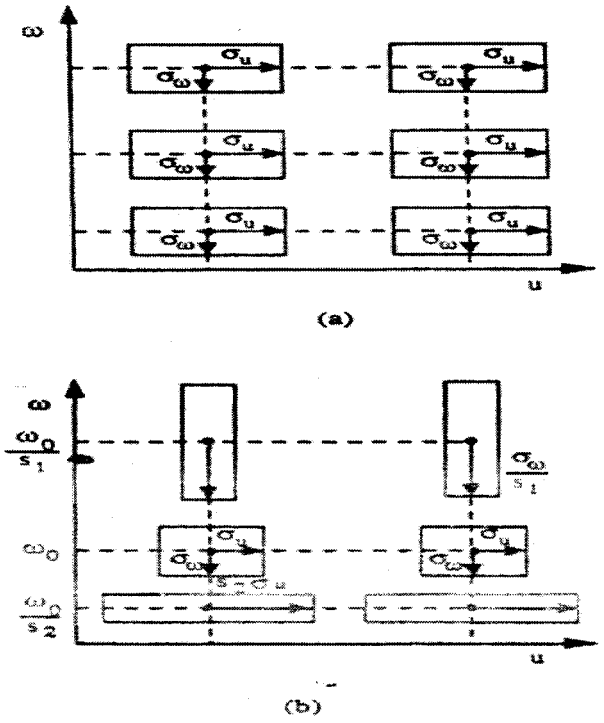


Figure 3. 4 : Différence de résolution entre a) La fenêtre utilisée dans la méthode par transformée de Fourier b) La fenêtre utilisée par les méthodes basée sur les transformées d'ondelettes

Figure tirée de Bakshi et Stephanopoulos (1995)

3.2.4 Détection du régime permanent

L'étape de détection du régime permanent proposée dans la méthodologie de Jiang (Figure 2.1) permet de sélectionner un groupe de données qui seront utilisées dans l'étape de réconciliation. Pour se faire, plusieurs techniques statistiques, très performantes, ont été mises au point. Les trois méthodes les plus utilisées sont : le test statistique composé (CST) (Narasimhan 1986), le test d'évidence mathématique (MTE) (Narasimhan 1987) et une méthode efficace proposée par Cao et Rhinehart (1995). Par contre, puisque ces méthodes (à l'exception de celle de Cao et Rhinehart) sont effectuées sur des périodes de temps spécifiques, elles ne sont pas applicables en ligne (Jiang 2003a).

Pour une application en ligne, il existe deux grandes catégories de techniques. La première est basée sur l'hypothèse de distribution normale des mesures ainsi que sur les théories statistiques qui en découlent. L'autre s'appuie sur les informations contenues dans l'utilisation des ondelettes lors de l'étape de filtration des données.

Dans l'un ou l'autre des cas, la première étape à franchir, avant d'entreprendre la détection elle-même, est celle de la sélection des variables de références. En effet, à cause du grand nombre de variables présentes dans la plupart des systèmes étudiés et des restrictions au niveau de leur échantillonnage, il est impossible, à cette étape, de tenir compte de toutes les variables associées au système étudié. Il faut donc faire un choix sur les variables qui seront analysées. En règle générale, chacune des variables choisies est testée individuellement. Par la suite, à l'aide des résultats obtenus, un index représentant le niveau de stabilité du système est calculé.

3.2.4.1 Tests statistiques –Approche de Cao et Rhinehart

Les tests, basés sur l'hypothèse de normalité des mesures et sur les théories statistiques, permettant de détecter les régimes permanents en ligne sont nombreux. Premièrement, il est possible d'identifier ces régimes à l'aide d'une approche dite directe. Pour cette méthode, on effectue une régression linéaire à l'intérieur d'une fenêtre de temps. Il faut donc préalablement choisir la longueur de la fenêtre de temps à l'aide de données historiques (voir section 3.2.2.1).On utilise ensuite un test de Student sur la pente de cette régression afin de déterminer si la variable est dans un état permanent. Si la pente est égale 0 le test est rejeté et on conclue que la variable est en régime permanent.

En plus de la méthode directe, une méthode comparant la moyenne calculée à l'intérieur d'une fenêtre déroulante à une valeur de référence basée sur des données historiques, est appliquée en industrie. Pour ce faire, on utilise encore une fois un test de student. Cependant, cette méthode n'est pas idéale. En effet, l'hypothèse de présence d'un régime permanent et celle d'égalité de deux moyennes sont différentes. Si le procédé oscille et que la moyenne des données contenues dans la fenêtre est égale à une valeur de référence, la variable est considérée en régime permanent, alors qu'elle ne l'est pas (Cao et Rhinehart 1995).

En plus des méthodes basées sur la pente de la régression linéaire et sur la moyenne, les autres tests statistiques utilisent la variance du signal afin de détecter les périodes permanentes. La méthode développée par Cao et Rhinehart (1995) est la plus populaire de celles-ci. Elle ne nécessite pas un grand espace d'entreposage et une grande exigence informatique (Jiang 2003a). Cette méthode utilise premièrement un filtre linéaire conventionnel, filtre pondéré à lissage exponentielle, pour filtrer les données. Cette approche permet à la méthode de fonctionner en ligne sans utiliser de fenêtre déroulante.

$$x_{f_i} = \rho_1 x_i + (1-\rho_1) x_{f_{i-1}} \qquad (17)$$

où ρ_1 est un paramètre de pondération attribué à la valeur mesurée ($0 < \rho_1 \leq 1$)

Par la suite, les auteurs proposent de comparer deux méthodes d'estimation de la déviation des mesures : l'estimation de la déviation des mesures par rapport à la moyenne filtrée et l'estimation de la déviation entre deux points successifs. Puisque le calcul de la déviation des mesures par rapport à la moyenne (18) est exigeant au point de vue informatique, les auteurs proposent de calculer une déviation filtrée par rapport à la moyenne (19).

$$v^2 = E((x_i - x_{f_{i-1}})^2) \quad (18)$$

$$v_{f_i}{}^2 = \rho_2 (x_i - x_{f_i})^2 + (1-\rho_2) v_{f\,i-1}{}^2 \quad (19)$$

où ρ_2 est la pondération attribuée à la déviation par rapport à la moyenne filtrée au moment de la mesure ($0 < \rho_2 \leq 1$).

Le même raisonnement est appliqué au calcul de la variance entre deux points successifs.

$$\delta^2 = E((x_i - x_{i-1})^2) \quad (20)$$

$$\delta_{f_i}{}^2 = \rho_3 (x_i - x_{i-1})^2 + (1-\rho_3) \delta_{f\,i-1}{}^2 \quad (21)$$

où ρ_3 est la pondération attribuée à la différence entre 2 données successives au moment de la mesure ($0 < \rho_3 \leq 1$).

Afin de comparer les deux méthodes de calcul de la déviation, Cao et Rhinehart (1995), utilise des estimées des variances :

$$S_{1_i}{}^2 = \frac{2-\rho_1}{2} v_{f\,i}{}^2 \qquad (22)$$

$$S_{2i}{}^2 = \frac{\delta_{f\,i}{}^2}{2} \quad (23)$$

Le Rapport de ces estimé est finalement calculé :

$$R_i = \frac{S_{1_i}{}^2}{S_{2_i}{}^2} \qquad (24)$$

Le résultat de R_i peu être influencé par trois différents types de comportement :

1 Si le procédé est en régime permanent, alors $R \approx 1$

2 Si la moyenne des données change, alors R>1. Dans ce cas, les deux estimations de la déviation seront influencées. Cependant, puisque la première est comparée à une moyenne filtrée et que la deuxième est comparée à la donnée précédente, l'effet du changement sera plus long sur le premier calcul.

3 Si les données mesurées successivement alternent entre des valeurs extrêmes, alors R<1. Dans ce cas-ci, la première estimation sera moins influencée par les variations que la deuxième, ce qui entraîne une diminution de la valeur calculée de R.

Cao et Rhinehart (1995) jugent que la troisième situation est inhabituelle dans les procédés chimiques. Ils proposent donc de baser la détection du régime permanent sur le test suivant : $R>R_{Crit}$. En résumé, lorsque R dépasse une valeur seuil R_{Crit}, on peu conclure que le procédé n'est pas en régime permanent.

Afin de profiter de cette méthode de manière optimale, Bhat et Saraf (2004) ont proposé une méthodologie permettant de faire un choix judicieux des paramètres utilisés pour la pondération. Cette méthode propose de fixer ces paramètres afin de minimiser les erreurs de types I et II (pour plus de détails sur ces types d'erreurs, voir Turner et McElrath 1996) ainsi que les délais de détection du régime permanent. Voici les étapes proposées afin d'atteindre ces buts :

1 Choisir une valeur de départ de ρ_3 (0,01 est proposée)

2 Choisir une valeur de départ de ρ_2 (0,05 est proposé)

3 Choisir une valeur de ρ_1 (0.02 est proposé). Pour ces valeurs de ρ_1, ρ_2 et ρ_3, calculer la valeur de R_{Crit}, la quantité d'erreurs de type II et le délai de détection du régime permanent.

4 Augmenter la valeur de ρ_1 de 0.02 jusqu'à ce que la limite souhaitée d'erreur de type II soit atteint.
5 Augmenter ρ_2 de 0.05 et retourner à l'étape 3. Répéter jusqu'à ce que le délai de détection soit minimisé.
6 Augmenter ρ_3 de 0.01 et retourner à l'étape 2 jusqu'à ce que la méthode donne les résultats optimaux.

Une fois la détection du régime permanent effectuée pour chacune des variables de références, on détermine si le système étudié est en régime permanent : si toutes les variables de références sont en régime permanent, alors le système est considéré en régime permanent. Dès qu'une variable se situe à l'extérieur des bornes de permanences, l'opération du système sera considérée non permanente.

3.2.4.2 Technique basée sur les ondelettes

Jiang et al. (2003a) proposent une technique de détection du régime permanent basée sur les caractéristiques des ondelettes. Cette technique utilise les transformées d'ondelettes. La transformée de premier ordre est directement proportionnelle à la première dérivée de la série de données (voir section 3.2.1). Par conséquent, une valeur positive de $Wf(t)$ permet d'identifier un accroissement des mesures appartenant à f(t), alors qu'une valeur négative de $Wf(t)$ est associée à une diminution des données. Finalement, lorsque $Wf(t) = 0$, on peut être en présence de deux phénomènes différents :

1 La donnée appartient à un régime permanent
2 La donnée est un extremum local

Dans le deuxième cas, le point est appelé : point croisant le 0. Afin d'identifier adéquatement le régime permanent, il est important de différencier ces points de ceux appartenant à la première catégorie. Pour se faire, Jiang et al. (2003a) ont proposés

d'utiliser la transformée d'ondelettes de deuxième ordre du signal (8). Puisque cette transformée est proportionnelle à la seconde dérivée du signal, il est possible de différencier les point croisant le 0 et les points appartenants à un régime permanent :

Si $WW_j f(t) = 0$, alors le point fait parti d'un régime permanent

Si $WW_j f(t) \neq 0$, alors c'est un point croisant le 0

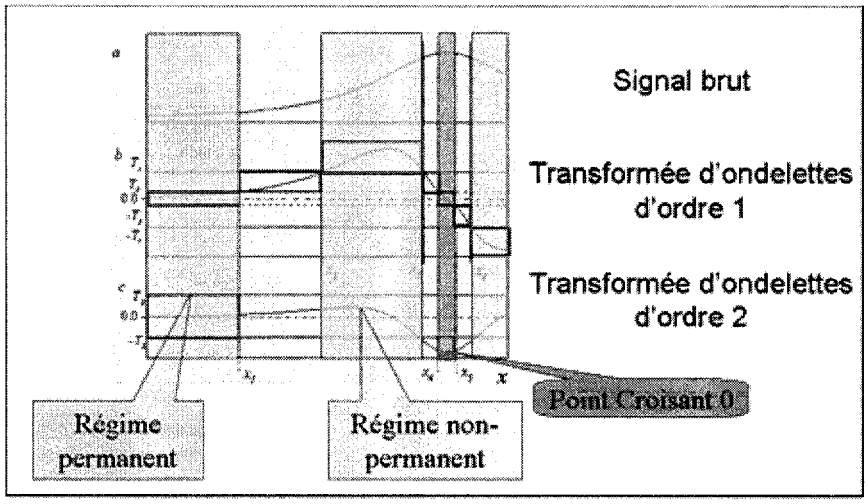


Figure 3. 5 : Représentation graphique de la méthode de détection du régime permanent à l'aide des transformées d'ondelettes.

Afin de permettre la classification des différents régimes permanent, non permanent, quasi-permanent, Jiang et al (2003a) introduisent un index (β) permettant de mesurer le degré de stabilité de la variable analysée :

- Si $\beta(t) = 0$, alors la variable est instable
- Si $\beta(t) = 1$, alors la variable est en régime permanent
- Si $0 < \beta(t) < 1$, alors la variable est situé entre un régime permanent et un statut instable

Le calcul de l'index s'effectue en trois étapes :

1 Si $|W_j f(t)| > T_u ...alors. \beta(t) = 0$

où $T_u = 3\lambda_2 \sigma_{W_j}$

où λ_2 est un paramètre à réglé et où σ_{W_j} représente la variance de la transformée d'ondelettes de premier ordre.

2 Si $|W_j f(t)| < T_s$ et $|WW_j f(t)| < T_W$, alors $\beta(t) = 0$

où $T_s = \sigma_{W_j}$ et $T_W = \sigma_{WW_j}$

où σ_{WW_j} représente la variance de la transformée d'ondelettes d'ordre 2.

3 Dans les autres cas, $\beta(t) = \xi[\theta(t)]$

où $\theta(t)$ est un facteur représentant l'influence des transformées d'ondelettes d'ordres 1 et 2 sur la stabilité de f(t).

$$\theta(t) = |W_j f(t)| + \gamma |WW_j f(t)| \qquad (25)$$

$$\text{où } \gamma = \begin{cases} 0,...|WW_j f(t)| \le T_W \\ \dfrac{(|WW_j f(t)| - T_W)}{2T_W},...|WW_j f(t)|.\varepsilon.(T_W, 3T_W) \\ 1,...|WW_j f(t)| \ge 3T_W \end{cases}$$

$\xi[x]$ est une fonction de transfert lisse comprise entre 0 et 1 et donnée par :

$$\xi[x] = \frac{1}{2}\left[\cos\left(\frac{x - T_S}{T_u - T_S}\pi\right) + 1\right] \qquad (26)$$

En se basant sur les différentes équations données et sur la Figure 2.5, on peut résumer les calculs permettant d'obtenir l'index de stabilité ainsi :

$$\beta(t) = \begin{cases} 0, \ldots\ldots\ldots\ldots \theta(t) \geq T_u \\ \xi[\theta(t)], \ldots T_S < \theta(t) < T_u \\ 1, \ldots\ldots\ldots\ldots \theta(t) \leq T_S \end{cases} \qquad (27)$$

Une fois l'indice de stabilité calculé pour chacune des variables de référence, on calcule l'index de stabilité du système à l'aide de la règle de combinaison de Dempster. Pour N variables, on a:

$$B(t) = \prod_{i=1}^{N} \beta_i(t)^{\frac{wi}{\Sigma wi}} \qquad (28)$$

où w_i représente la pondération attribué à la ième variable.

3.3 Réconciliation de données

Selon la méthodologie utilisée par Jiang (Figure 3.1), cette étape est séparée en trois sous-étapes distinctes. Premièrement, les différentes variables du système étudié sont classifiées. C'est à cette étape que l'on détermine quelles variables peuvent être réconciliées. Par la suite, les erreurs brutes présentes dans les mesures sont identifiées. Finalement, à l'aide des contraintes du système (bilan de masse, d'énergie, contrainte physiques, ...), les données sont réconciliées. Ce qui permet de minimiser les erreurs présentes dans la série de données.

3.3.1 Classification des variables

Lors de cette première étape, les différentes variables présentes dans les systèmes étudiés sont classifiées. Cette classification permettra par la suite de déterminer quels sont les variables qui pourront être réconciliées, celles pouvant être estimées et celles dont l'exactitude ne peut pas être améliorée à l'aide de l'étape de réconciliation.

Crowe (1996) et Bagajewicz (2001) ont présentés une nomenclature permettant de classer les variables de manière à tirer un maximum d'informations de celles-ci à l'étape de réconciliation. Selon cette nomenclature, il existe deux grandes classes de variables qui sont à leur tour subdivisées en deux sous-classes :

- Variable mesurée
  - Redondante
  - Non redondante
- Variable non mesurée
  - Observable
  - Inobservable

Lorsqu'une variable est estimée par un instrument quelconque, elle est dite mesurée. Une variable mesurée est redondante s'il est possible, en utilisant les autres variables mesurées, de la calculer d'au moins une autre façon (Bagajewicz 2001). Il est important de différencier les variables redondantes de celles qui ne le sont pas. En effet, puisque la réconciliation (voir section 3.3.3) est un problème d'optimisation minimisant l'erreur présente dans les mesures, il est important que la série de mesures étudiée soit surspécifiée (redondante). Une série de mesures est redondante si toutes les mesures qui la composent sont redondantes (Bagajewicz 2001).

Une variable non mesurée est dite observable lorsqu'elle peu être calculée à l'aide des contraintes issues d'un procédé (Crowe 1996). Il est important de mentionner que ces contraintes peuvent aussi bien être des mesures que des informations quantitatives reliées à l'opération du procédé.

3.3.2 Détection des erreurs brutes

La détection des erreurs brutes est une étape cruciale permettant d'avoir une réconciliation des données efficace. En effet, en détectant les erreurs brutes de manière

adéquate, il est possible d'ajuster l'équation de minimisation afin de tenir compte de cette information. Si les erreurs brutes ne sont pas détectées, elles influencent la réconciliation de manière négative (Bodington 1995).

Selon Bagajewicz (2003), cette étape, reconnue comme étant l'étape limitante de la réconciliation de données, comporte deux parties distinctes : la détection de l'erreur et l'estimation de sa dimension. Selon lui, les défis importants de cette étape du traitement de données sont :

- L'identification de l'existence des erreurs brutes
- L'identification du lieu exacte des erreurs brutes
- L'identification du type d'erreurs brutes.
- La détermination de la dimension de l'erreur.

Plusieurs travaux ont été effectués dans ce domaine. Reilly et Carpani ont proposé le test global, Mah a proposé le test sur les contraintes. Plus tard, au même moment, Mah et Crowe ont proposés indépendamment le test sur les mesures. Plus récemment, des méthodes basées sur l'analyse des composantes principales (PCA) et sur les réseaux de neurones ont été appliquées. Cependant, aucun de ces tests sur la détection des erreurs brutes ne permet de toutes les identifiées (Références tirées de Crowe 1996).

Le test sur les mesures est basé sur une supposition de distribution normale des erreurs et utilise le vecteur de correction des mesures (29). La moyenne des mesures des corrections devaient donc être 0 et η_i suivre une distribution normale N(0,1) (Bagajewicz 1999).

$$a = F - \tilde{F} \qquad (29)$$

$$\eta_i = \frac{a_i}{\sqrt{Q_{ii}}} \qquad (30)$$

Une fois que les erreurs brutes sont identifiées, il est possible de réagir de deux manières différentes avant de passer à l'étape de réconciliation:

- Éliminer l'erreur et performer la réconciliation sans la variable affectée
- Corriger la donnée ou le modèle (cas d'une fuite ou d'un dépôt) et effectuer la réconciliation.

La première alternative est privilégiée lorsque l'on est en présence d'un système qui conservera de la redondance même si la donnée affectée est supprimée. Dans ce cas, la mesure sera observable et pourra être recalculée par la suite. La seconde option est appliquée lorsque l'élimination de la variable affectée provoque une non redondance du système étudié. Dans ce cas, afin de pouvoir effectuer l'étape de réconciliation, il faut corriger la donnée au meilleur de ses connaissances. Pour se faire, on peut utiliser la méthode compensatoire en série (Narasimhan et Mah 1987) couplée au test sur les mesures. Cette méthode itérative identifie l'erreur brute et estime sa grandeur jusqu'à ce qu'il n'y est plus d'erreurs identifiées.

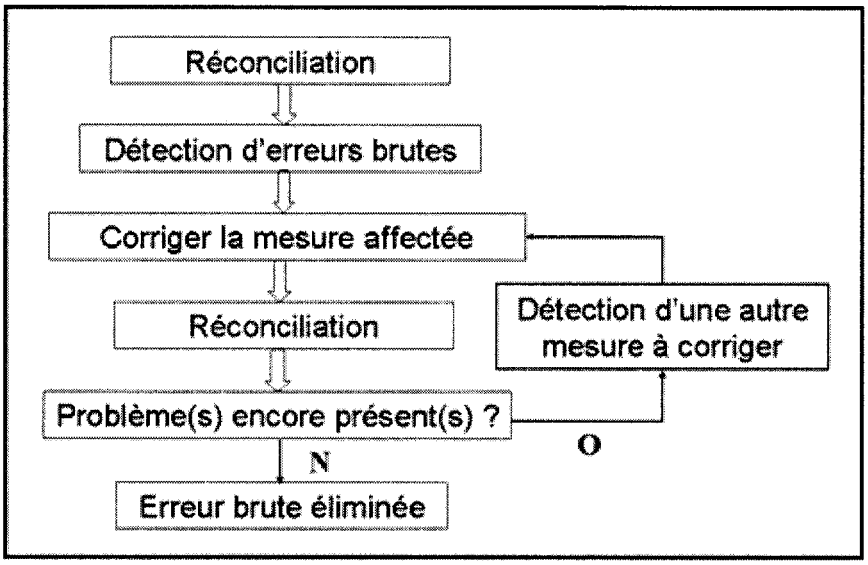


Figure 3. 6 : Étapes de résolution de la méthode compensatoire

3.3.3 Réconciliation en régime permanent

Les méthodes utilisées pour améliorer la qualité des données à l'aide de modèles physiques sont appelées : méthode de filtration à base d'un modèle, ajustement à l'aide de bilan massique ou plus généralement, techniques de réconciliation de données (Hodouin et Makni 1996). La réconciliation en régime permanent permet donc de corriger les mesures tout en les validant à l'aide de modèles qui représentent le procédé.

La réconciliation dynamique des données et les méthodes de détection des erreurs brutes qui y sont associées n'ont pas encore été étudiées de manière approfondie. De plus, Bagajewicz (2000) a démontré, mathématiquement et à l'aide d'exemples, que pour un système ne contenant pas de cuvier, la réconciliation en régime permanent basée sur la moyenne des mesures donne les mêmes résultats qu'une réconciliation dynamique des données, si les poids attribuées à chacune des mesures sont adéquats (voir section 3.3.3.1). Bagajewics (2001) a également démontré que les changements linéaires associés à un cuvier peuvent être modélisés par un courant entrant (ou sortant) d'un nœud (voir figure 3.7). Par conséquent, la réconciliation en régime permanent d'une série de données comportant un changement linéaire dans un cuvier devrait donner des résultats semblables à ceux d'une réconciliation dynamique.

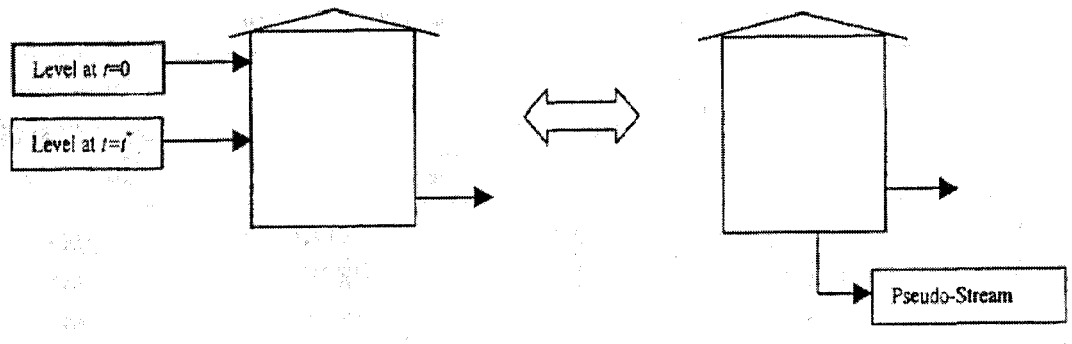


Figure 3. 7 : Modélisation d'un cuvier à l'aide d'un courant entrant/sortant d'un nœud

Figure tirée de Bagajewicz 2001

3.3.3.1 Problème de minimisation

Tous les problèmes de réconciliation de données sont basés sur la supposition de distribution normale des erreurs, il est donc très important que les anormalités soient éliminées avant cette étape (Bodington 1995). La réconciliation est un problème de minimisation sous contraintes. Ces contraintes peuvent aussi bien être linéaires que non linéaires. La fonction objective de ce problème d'optimisation tente normalement de minimiser le carré des erreurs entre les valeurs mesurées et ciblées à l'aide d'une pondération associée à chacune des mesures effectuées (Crowe 1996). Afin d'attribuer une pondération adéquate, on utilise l'équation suivante :

$$w_i = \frac{1}{\sigma_i} \qquad (30)$$

Le problème de réconciliation linéaire en régime permanent est donc donné par :

$$\left.\begin{array}{l} Min\left[F-\tilde{F}\right]^T Q^{-1}\left[F-\tilde{F}\right] \\ s.t.: D\tilde{F}=0 \end{array}\right\} \qquad (31)$$

où D est la matrice des contraintes associées au procédé
(Bagajewiz 2001)

Le problème de réconciliation non linéaire en régime permanent est donné par :

$$\left.\begin{array}{l} Min\left[Z-\tilde{Z}\right]^T Q^{-1}\left[Z-\tilde{Z}\right] \\ s.t.: g(\tilde{Z})=0 \end{array}\right\} \qquad (32)$$

où $g(\tilde{Z})$ inclus les contraintes reliées à l'opération des unités et aux instruments
(Bagakewicz 2001)

3.3.3.2 *Sigmafine*

Sigmafine d'OSI est un logiciel commercial, intégrable à PI permettant d'effectuer la réconciliation des données en régime permanent. La réconciliation des données basée sur les bilans de masses globaux et sur les différents composés est inclue dans ce logiciel. Cette réconciliation est basée sur une supposition de distribution normale des données. Puisque le logiciel tient compte des bilans sur les différents composés en plus des bilans globaux, son algorithme de résolution est basé sur des équations bilinéaires. Les contraintes de ce problème sont donc :

$$\begin{aligned} \sum_i D_{i,k}\tilde{F}_i &= 0 \\ \sum_i D_{i,k}\tilde{F}_i C_{i,i} &= 0 \end{aligned} \qquad (33)$$

où C est la matrice des composés réconciliés

La technique de résolution du logiciel est basé sur le filtre de Kalman(1960) et ne requière aucune inversion matricielle. Il peut permettre de travailler sur de gros système sans problème de convergence. Ce qui le rend très pratique pour les applications industrielles (Jiang et al. 2003b).

3.3.3.3 *Effet des mesures indirectes sur la supposition de normalisation*

La supposition de distribution normale des erreurs est très importante dans le domaine du traitement des données. Bagajewicz (1995) a démontré que cette supposition est effectivement valide dans le cas des mesures qui sont prises directement. Dans le cas où il y a une transformation non linéaire entre la mesure prise et la lecture donnée par l'appareil, cette supposition n'est plus applicable. Par conséquent, toutes les lectures de débits effectués à l'aide d'un différentiel de pression ne sont pas distribuées normalement. En effet, ce type de lectures fait appel à une fonction relative à la racine carrée du différentiel de pression enregistré.

Chapitre 4. Synthèse

La qualité des données a une grande importance dans le domaine du génie chimique et tout particulièrement dans celui des pâtes et papier. En effet, depuis quelques années, les systèmes de gestions de l'information et des données font une entrée en force dans ce secteur de l'industrie. À l'aide de ces systèmes, les différents acteurs espère d'améliorer les conditions associés à la gestion et à la production de leur procédé. Malheureusement, présentement, les données recueillies en temps réels sont analysées de manière impromptue. Ce type d'analyse ne permet malheureusement pas de tirer totalement profit des systèmes d'informations mis en place. Par conséquent, l'implantation d'une méthodologie permettant la correction des erreurs de mesures en ligne amélioreraient l'efficacité de ces systèmes.

4.1 Retour sur les concepts importants

4.1.1 La qualité des données

La qualité des données est un terme général qui est souvent utilisé pour faire référence à une liste de propriétés que devraient détenir une série de données idéales. Selon Bagajewicz (2001) ces propriétés peuvent être regroupées selon les trois caractéristiques suivantes : l'exactitude, la précision (reproductibilité) et la robustesse (confiance). Par conséquent, une série de mesures de haute qualité devrait être exacte, précise et robuste. Cependant, une série de mesures prises dans l'industrie contient plusieurs types d'erreurs que l'on peut regrouper sous trois principales catégories : anormalités, erreurs aléatoires erreurs brutes.

4.1.2 La méthodologie générale

Afin de traiter ces mesures de manière efficace, Jiang et al. (2003a) utilisent une méthodologie basée sur la correction des erreurs en plusieurs étapes. La première étape proposée permet d'identifier les anormalités présentes dans les mesures. Une fois

identifiés, les points anormaux sont corrigés en remplaçant la valeur supprimée par une valeur corrigée. Par la suite, afin d'éliminer les erreurs aléatoires, les données normalisées sont filtrées. L'étape suivante permet de détecter les séries de données en régime permanent. Ce qui permet de choisir les données qui seront réconciliées. Finalement, après une classification des variables et une détection des erreurs brutes, les données sont réconciliés.

4.1.3 La détection des anormalités

Dans le domaine de la détection des anormalités pour des applications en ligne, il existe deux grands types de tests qui peuvent être utilisés : les tests statistiques et ceux basés sur l'utilisation des caractéristiques propres aux transformées d'ondelettes. Dans la catégorie des tests statistiques, il a été démontré par Hampel (1971 et 1974) que l'utilisation de la médiane comme estimateur est une méthode plus robuste à la présence d'anormalités que l'utilisation d'une moyenne. Par conséquent, l'utilisation de la médiane et de la médiane des déviations absolues est plus efficace pour détecter des anormalités que l'utilisation de la moyenne et de son écart-type. Jiang et al. (2000) ont utilisés les caractéristiques présentent dans les différents niveaux des transformées d'ondelettes afin d'identifier les anormalités du signal étudié.

4.1.4 Filtration de données

Plusieurs travaux ont été effectués dans le domaine de la filtration des données. Des filtres de plusieurs espèces ont été élaborés lors de ces travaux. Les filtres passes-bas n'ont pas la capacité d'éliminer le bruit présent à différentes fréquences et provoquent une distorsion du signal traité. Les faiblesses de ces filtres ont forcé le développement des filtres à plusieurs niveaux. Dans cette catégorie de filtres, l'utilisation des ondelettes est privilégiée sur celle des séries de Fourier à cause des faiblesses de localisation temporelle de ces dernières. Cependant, l'utilisation des ondelettes nécessite le choix du

niveau de coupure du signal. En effet, présentement la seule façon de choisir ce niveau de coupure est de juger si le signal lisse du niveau contient seulement les changements associés à l'opération du procédé. Puisqu'il n'y a pas de techniques permettant d'optimiser le choix de ce niveau de coupure, il doit être préalablement être ajusté à l'aide de données historiques.

4.1.5 Détection du régime permanent

La détection du régime permanent peu être effectuée à l'aide de techniques basées sur les théories statistiques ou à l'aide de celle utilisant les caractéristiques des transformées d'ondelettes. Puisque les techniques statistiques sont effectuées sur des périodes de temps spécifiques, à l'exception de celle proposée par Cao et Rhinehart (1995), elles ne sont pas utiles en ligne.

La première étape de la méthode proposée par Cao et Rhinehart (1995) filtre les données à l'aide d'un filtre passe-bas. Ce type de filtre n'est pas efficace pour éliminer le bruit présent à différentes fréquences et amplitude. Il cause des distorsions dans le signal (voir section 2.2.3). Or son utilisation dans les moments où la fréquence du signal varie rapidement devrait être peu efficace (figure 3.1).

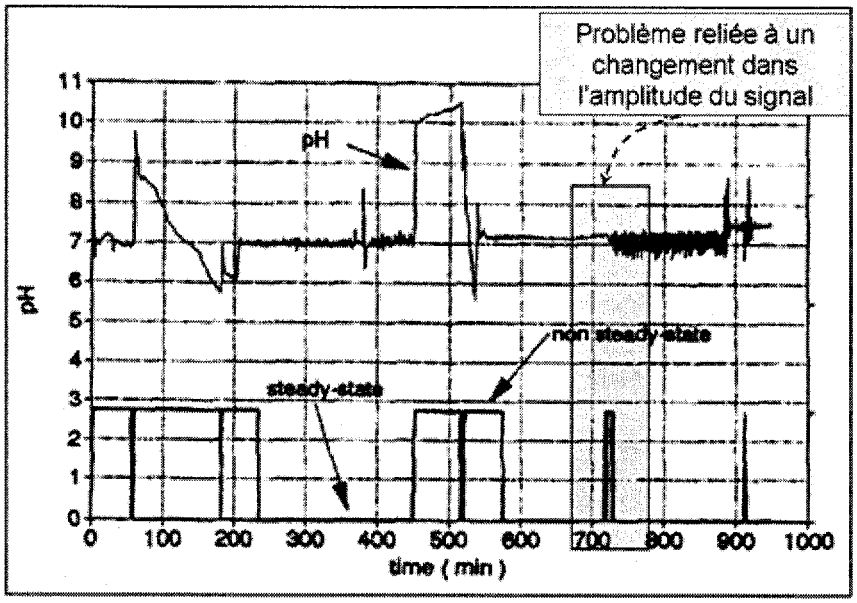


Figure 4. 1 : Exemple d'utilisation de la méthode de Cao et Rhinehart

Figure tirée de Cao et Rhinehart (1995)

La méthodologie de Jiang (voir section 2.1) précède l'étape de détection du régime permanent d'une filtration du signal. Comme il en a été discuté à la section précédente, l'utilisation des ondelettes est la méthode la plus efficace pour filtrer les données en ligne. Cependant, on doit préalablement choisir le niveau de coupure à l'aide de données historiques. Puisque les erreurs aléatoires associées à un signal donné ne possèdent pas la même fréquence à tous les instants, cette restriction crée un problème au niveau de l'élimination du bruit aléatoire et pourrait causé la présence d'erreurs aléatoires de faibles intensités à cette étape. Puisque le second type de techniques se base sur l'utilisation de la transformée d'ondelettes (proportionnelle à la dérivée du signal) pour prendre une décision, la détection pourrait être faussée. De plus, afin de différencier les points croisant le 0 des points appartenant à un régime permanent, cette méthode fait appel à la seconde transformée d'ondelettes du signal. En théorie, le signal ne devrait contenir que les changements associés à sa vraie nature suite à son premier passage dans la transformée. Donc, l'utilisation d'une seconde découpure du signal en plusieurs niveaux devrait automatiquement éliminer une partie des changements propres au signal traité dans le signal détaillé.

4.1.6 La réconciliation des données

Selon les travaux de Bagajewicz (2001), la réconciliation, en régime permanent, d'une série de données comportant un changement linéaire dans un cuvier devrait donner des résultats semblables à ceux d'une réconciliation dynamique.

La réconciliation des données s'effectue en plusieurs étapes. La première d'entre elles est la classification des variables. Elle permet d'identifier quelles sont les variables redondantes dans le système et par conséquent quelles seront les données qui pourront être réconciliés. Cette étape est donc cruciale dans l'élaboration d'un projet et doit être effectuée avant l'implantation d'une méthodologie de traitement des mesures englobant une réconciliation des données.

Tous les problèmes de réconciliation de données sont basés sur la supposition de distribution normale des erreurs, cependant il a été prouvé que les mesures indirectes ne suivent pas une distribution normale (Bagajewicz 1995). Par conséquent, l'utilisation de ce type de mesures ne peut pas être associé à une supposition de distribution normale des lectures.

Sigmafine d'OSI est un logiciel commercial, intégrable à PI permettant d'effectuer la réconciliation des données en régime permanent. Il peut permettre de travailler sur de gros systèmes sans problème de convergence. Ce qui le rend très pratique pour les applications industrielles.

4.2 Problèmes reliés au traitement des données en ligne

4.2.1 Relié à la méthodologie générale

La méthodologie générale de traitement de données utilisées par Jiang (2003a) est très intéressante pour traiter les mesures hors-ligne, car elle permet de bien identifier les différents types d'erreurs reliés aux mesures tout en incluant des étapes adéquates qui permettent leur correction. Cependant, cette méthode nécessite beaucoup de travail relié à la mise au point des différents paramètres devant être ajustés à chacune des étapes du traitement des données. Afin d'utiliser cette méthodologie en ligne, il serait important d'inclure une étape permettant d'automatiser les réglages associés à ces différents paramètres. Cette étape pourrait être construite à l'aide des données historiques du procédé étudié.

4.2.2 La détection des anormalités

Deux grandes méthodes sont utilisées afin de détecter les anormalités présentent dans une série de mesures : la méthode d'Hampel(1974) qui est basée sur l'utilisation de la médiane et sur les théories statistiques qui en découlent et la méthode basées sur l'exploitation des propriétés des transformées d'ondelettes. Afin d'être implémenté en ligne, ces deux méthodes nécessitent l'utilisation d'une fenêtre déroulante qui permet d'utiliser les caractéristiques récentes du signal afin de détecter ses anormalités. Les deux méthodes ont prouvées leur efficacité. Par contre, jusqu'à maintenant aucune étude n'a comparé leur efficacité globale.

4.2.3 La détection du régime permanent

La détection des régimes permanents en ligne peut être effectuée de plusieurs manières différentes. Cependant, chacune des façons comporte des lacunes qui ont été soulignées dans les sections 3.2.4 et 4.1.5. En effet, les méthodes statistiques n'exploitent pas le plein potentiel des différentes techniques de filtration à plusieurs niveaux et se confinent dans une détection totale du régime permanent (test totalement positif ou totalement négatif), tandis que celles basées sur l'exploitation des caractéristiques des transformées d'ondelettes ne tiennent pas compte des problèmes associés à la présence probable de bruit dans le signal traité et à la création de distorsion associée au second découpage d'une série de données. Le développement d'une méthode plus efficace que les précédentes serait donc une alternative à envisager.

4.2.4 Relié au domaine des pâtes et papier

Les systèmes de gestion des données sont de plus en plus utilisés dans le domaine des pâtes et papiers. Cependant, les méthodes d'analyses actuelles ne permettent pas de profiter pleinement des informations recueillies dans les mesures prises en temps réels. Le développement d'une méthodologie de traitement automatique des données de procédés permettrait d'en tirer d'avantage de profits.

Chapitre 5. Problématique et méthodologie de recherche

5.1 Problématique

Une méthodologie permettant de réconcilier les données en régime permanent a été élaborée (section 3.1). Elle suggère un prétraitement des erreurs afin de permettre une identification des régimes permanents plus efficace et, par conséquent, une amélioration de la qualité des données après une réconciliation. Cette méthode permet de bien identifier les différents types d'erreurs reliées aux mesures tout en y incluant des étapes adéquates qui permettent leur correction. Cependant, elle nécessite beaucoup de travail relié à la mise au point des différents paramètres devant être ajustés à chacune des étapes du traitement des données. Ces paramètres sont grandement influencés par la fréquence et l'amplitude des mesures prises sur le procédé. En ligne, les données en temps réel possèdent des caractéristiques (amplitude et fréquence) qui varient dans le temps et selon les conditions d'opérations. Cette méthodologie n'est donc pas directement applicable en ligne.

La première étape de la méthodologie, proposée afin de réconcilier les données en régime permanent, suggère une détection des anormalités à l'aide des caractéristiques propres aux ondelettes. Cependant, des tests basés sur les théories statistiques peuvent également être utilisés pour détecter les points anormaux présents dans une série de données. Le plus efficace d'entre eux est le test d'Hampel qui se base sur l'utilisation de la médiane. Ce test peut également être utilisé en ligne. Par contre, son efficacité n'a jamais été comparée à celle de la méthode basée sur les ondelettes.

Plusieurs techniques de détection des anormalités existent. Cependant, seulement que deux d'entre-elles semblent se démarquer (section 3.2) : la méthode de Cao et Rhinehart et la méthode basée sur l'utilisation des ondelettes. Ces deux techniques possèdent des faiblesses qui leurs sont propres. La première utilise des filtres passe-bas afin d'approximer la valeur moyenne des mesures et dans la filtration les calculs de déviations (utilisés afin de juger du statut du procédé). L'utilisation de ces filtres

engendre la création de distorsions et de délais dans les données traitées. De son côté, la méthode basée sur l'utilisation des ondelettes possède des problèmes reliés au choix optimal du niveau de coupure utilisé lors de l'étape de filtration des données. Puisque chacune de ces méthodes possède des limitations, une méthode plus efficace que les deux premières serait donc souhaitable pour les implantations en ligne.

En plus des problèmes méthodologiques et spécifiques à la détection des anormalité et du régime permanent, il faut souligner le problème relié à l'émergence de plus en plus grandissante des systèmes de gestion de données dans les industries chimiques et plus particulièrement dans le secteur des pâte et papiers. Ces systèmes de gestions des données permettent une accessibilité grandissante des mesures, en temps réel, effectuées sur les procédés. Cependant, ces données sont présentement utilisées suite à une analyse impromptue. Ce type d'analyse ne permet pas de détecter systématiquement tous les types d'erreurs présents dans les mesures prises. Par conséquent, l'utilisation des données qui sont entreposées en temps réels, dans les systèmes de gestion, n'est pas adéquate.

5.2 Hypothèses de recherche

À partir des informations résumées dans la section précédentes, quatre hypothèses de recherche ont été formulées :

1. En utilisant les données historiques et les connaissances reliées à un système, la méthodologie générale proposée par Jiang et al. (2003b) peut être adaptée, en ligne, afin d'améliorer en régime permanent, de manière robuste, l'exactitude et la précision, des mesures effectuées.
2. La détection des anormalités en ligne, à l'aide des transformées d'ondelettes est plus performante que la méthode de détection basée sur l'utilisation de la médiane.

3. Une méthode, couplant l'utilisation des caractéristiques propres aux transformées d'ondelettes et les théories statistiques est plus efficace pour identifier, en ligne, les régimes permanents, que les méthodes proposées jusqu'à maintenant.
4. Une méthodologie systématique d'analyse des données peut être appliquée sur les données d'un procédé industriel et permettre d'améliorer la qualité de celles-ci de manière significative.

5.3 Suppositions

Afin de permettre d'effectuer les travaux permettant de tester les hypothèses de recherche présentée à la section 5.2, il est essentiel de faire deux suppositions reliées aux mesures effectuées en temps réel sur le procédé : les erreurs aléatoires associées aux mesures sont normales et ces dernières sont effectuées de manière directe. Les détails relatifs à ces deux suppositions sont disponibles dans les deux sous sections qui suivent.

5.3.1 Normalité des erreurs aléatoires

Il est supposé dans ce travail que les erreurs aléatoires présentes dans les différentes mesures effectuées en temps réel sont normales et centrées à une moyenne de 0.

L'importance de cette supposition a été démontrée indirectement dans la revue de la littérature. En effet, les différentes méthodes utilisées pour détecter les anormalités présentes dans une série de données sont basées sur la présomption de normalités des autres mesures effectuées. De plus, les techniques permettant de filtrer les données aléatoires sont aussi basées sur cette supposition. Cette constatation est aussi valable dans le cas des étapes de détection du régime permanent et de la réconciliation des données (voir le chapitre 3, pour plus de détails).

5.3.2 Mesures effectuées directement sur les variables mesurées

À la section 3.3.3.3, il a été précisé que les mesures nécessitant une transformation non linéaire (par exemple : débit mesuré par un différentiel de pression) des valeurs prises ne répondent plus à la supposition de normalité des mesures après leur transformation. Cette transformation a un impact direct sur la qualité des données issues de la réconciliation. Des fonctions objectives spécifiques ont été élaborées pour ces cas particuliers de réconciliation (Bagajewicz 1995). Cependant, dans le cas présent, les données sont réconciliées à l'aide du logiciel Sigmafine 3 de OsiSoft Inc. Ce logiciel n'offre pas la possibilité d'utiliser ces fonctions. La supposition, selon laquelle les mesures sont effectuées de manière directe, permet de conserver les propriétés normales des mesures prises et par conséquent, elle permet l'utilisation du logiciel.

5.4 Méthodologie de recherche

Afin de valider les hypothèses de recherche présentées précédemment (section 5.2), une méthodologie a été mise en place. Puisque la première hypothèse présentée est reliée à la méthodologie générale du traitement de données en ligne, elle englobe les autres hypothèses et ne peut être vérifiée qu'après que les trois autres eurent été analysées.

Afin de corréler adéquatement la méthodologie implantée et la vérification des hypothèses de recherche, la méthodologie générale, permettant de répondre à la première hypothèse de recherche sera premièrement présentée. Par la suite, chacune des autres sections présentera la méthodologie utilisée pour vérifier les trois hypothèses spécifiques.

5.4.1 Méthodologie exacte, précise et robuste

Cette section présente la méthodologie générale utilisée afin de vérifier les hypothèses présentées à la section 5.2. Elle permettra donc de tester la première hypothèse : " En

utilisant les données historiques et les connaissances reliées à un système, la méthodologie générale proposée par Jiang (2003b) peut être adaptée, en ligne, afin d'améliorer en régime permanent, de manière robuste, l'exactitude et la précision, des mesures effectuées en régime permanent"

Afin de pouvoir juger l'amélioration de l'exactitude et de la précision des données mesurées à l'aide de l'implantation d'une méthodologie systématique, il est nécessaire de connaître la vraie nature des mesures de procédé. Puisque les mesures, effectuées en temps réel sur un procédé, possèdent toujours plusieurs types d'erreurs (anormalités, erreur aléatoire et erreur brute), il est impossible d'être totalement certain de la valeur réelle de ce qui est mesuré. L'utilisation directe de ces données afin de prouver l'amélioration de l'exactitude et de la précision des mesures est donc inutile. Pour juger de ces deux paramètres, des données expérimentales, générées à partir d'une investigation du procédé étudié, seront donc utilisées. Pour se faire, suite à l'analyse, une série de vraies données sera premièrement construite. Par la suite, des erreurs de tous les types y seront ajoutées. Cette façon d'effectuer la génération permettra de connaître la valeur ciblée par les mesures expérimentales et la nature exacte des erreurs qui y sont ajoutées.

Une fois que les données expérimentales seront créées, les techniques les plus pertinentes aux applications en ligne seront comparées. En ce qui concerne la détection des anormalités, l'utilisation des caractéristiques associées aux transformées d'ondelettes sera comparée à l'implantation du test de Hampel en ligne. À l'aide des conclusions tirées par la revue de la littérature, la filtration sera effectuée à l'aide d'une décomposition des fréquences par les ondelettes. Ensuite, un nouveau test de détection, en ligne, du régime permanent sera comparé à l'utilisation de la méthode proposée par Cao et Rhinehart et à l'utilisation des caractéristiques propres aux ondelettes. Finalement, l'impact de l'utilisation de chacune de ces techniques sur l'exactitude et la

précision des résultats sera analysé en comparant l'utilisation des différentes combinaisons possibles entre les techniques.

Une fois que l'effet de la méthodologie, sur l'exactitude et la précision des mesures effectuées en régime permanent, sera quantifié, sa robustesse sera testée. Pour se faire, cette dernière sera implantée sur des données mesurées en temps réels.

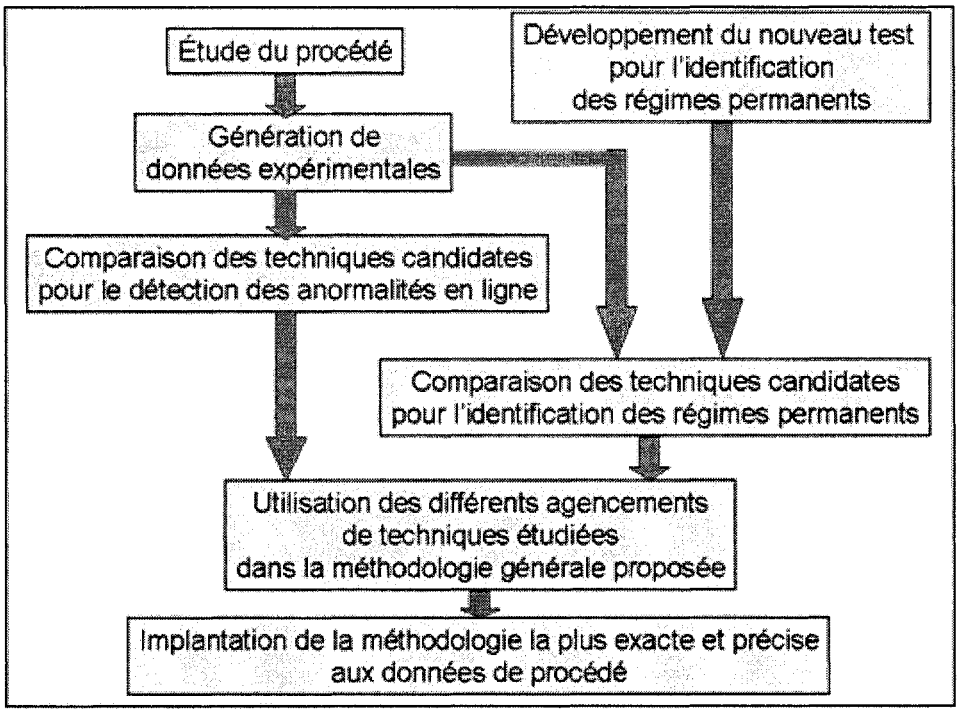


Figure 5. 1 : Méthodologie générale pour juger de la robustesse de la méthode implantée et son impact sur l'exactitude et la précision des résultats.

5.4.2 Efficacité de la détection des anormalités

Cette section effectue le lien entre la partie de la méthodologie qui est pertinente à la détection des anormalités et la vérification de la seconde hypothèse : « La détection des anormalités en ligne, à l'aide des transformées d'ondelettes est plus performante que la méthode de détection basée sur l'utilisation de la médiane. ».

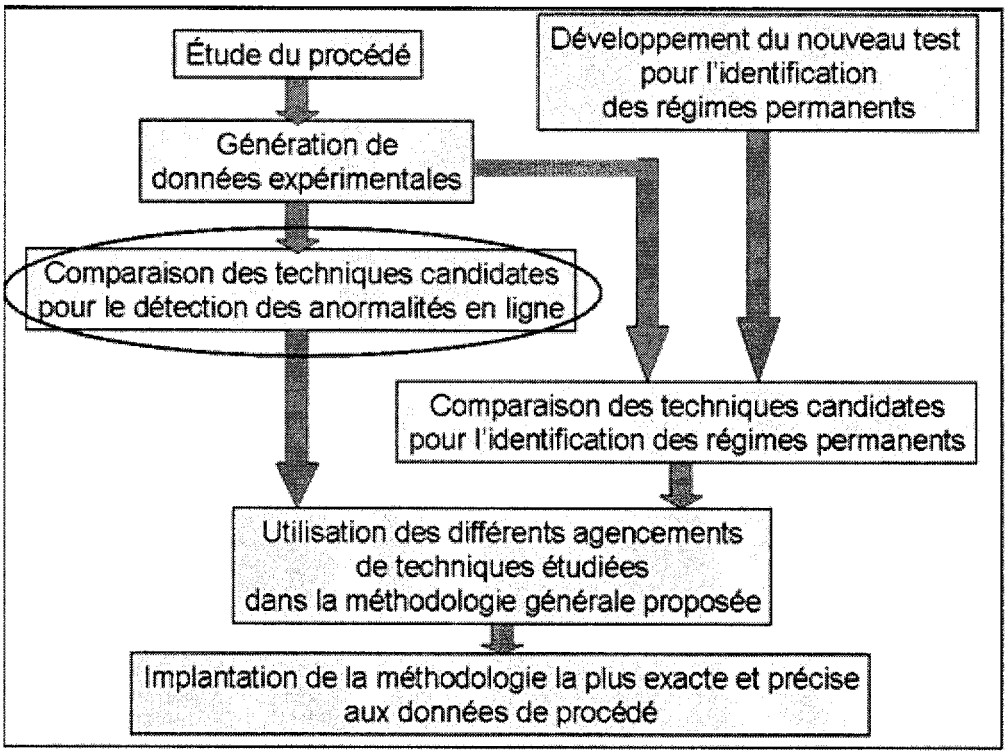


Figure 5. 2 : Partie de la méthodologie générale reliée à la deuxième hypothèse

Ici, les deux techniques les plus adéquates (selon la revue de la littérature) pour la détection des anormalités en ligne seront comparées : techniques basées sur les ondelettes et technique de Hampel. Les deux techniques seront utilisées pour détecter les mesures anormales ajoutées dans les données expérimentales. Pour bien délimiter les forces et les faiblesses des deux techniques, quatre séries de test seront effectuées. La figure 5.3 démontre que l'efficacité de détection des deux techniques sera comparée selon quatre catégories d'anormalités. Ces quatre catégories permettront de juger la performance des techniques selon l'amplitude de l'erreur et sa durée. Afin de juger adéquatement de l'efficacité des techniques, celles-ci doivent être évaluées selon leur habilité à détecter les mesures anormales de manière efficace. Pour se faire, les taux d'erreur de détection de type I (détection d'une mesure anormale manquée) et de type II (fausse détection) de détection seront calculés.

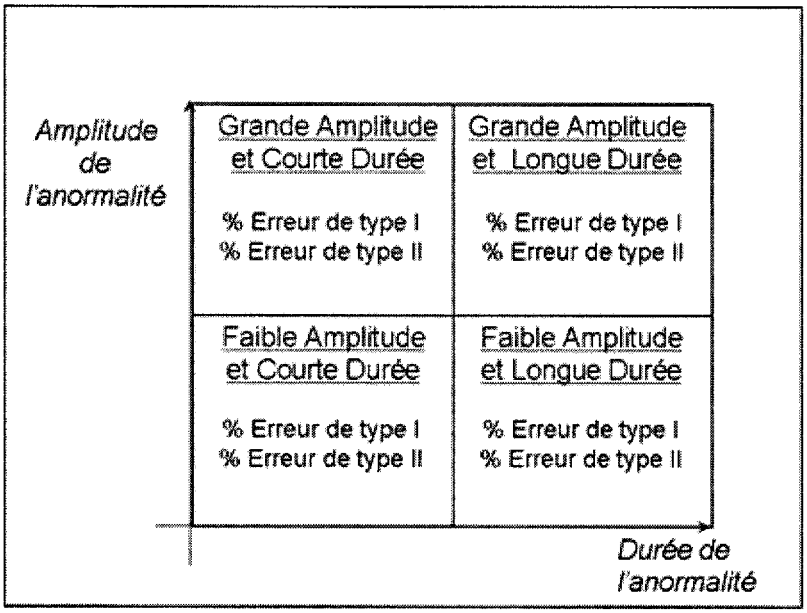


Figure 5. 3 : Les quatre tests utilisés pour comparer les deux techniques étudiées

5.4.3 Détection du régime permanent

Cette section permet de faire le lien entre la partie de la méthodologie qui est pertinente à la détection des régimes permanents et la vérification de la troisième hypothèse : « Une méthode, couplant l'utilisation des caractéristiques propres aux transformées d'ondelettes et les théories statistiques est plus efficace pour détecter, en ligne, les régimes permanents, que les méthodes proposées jusqu'à maintenant. ».

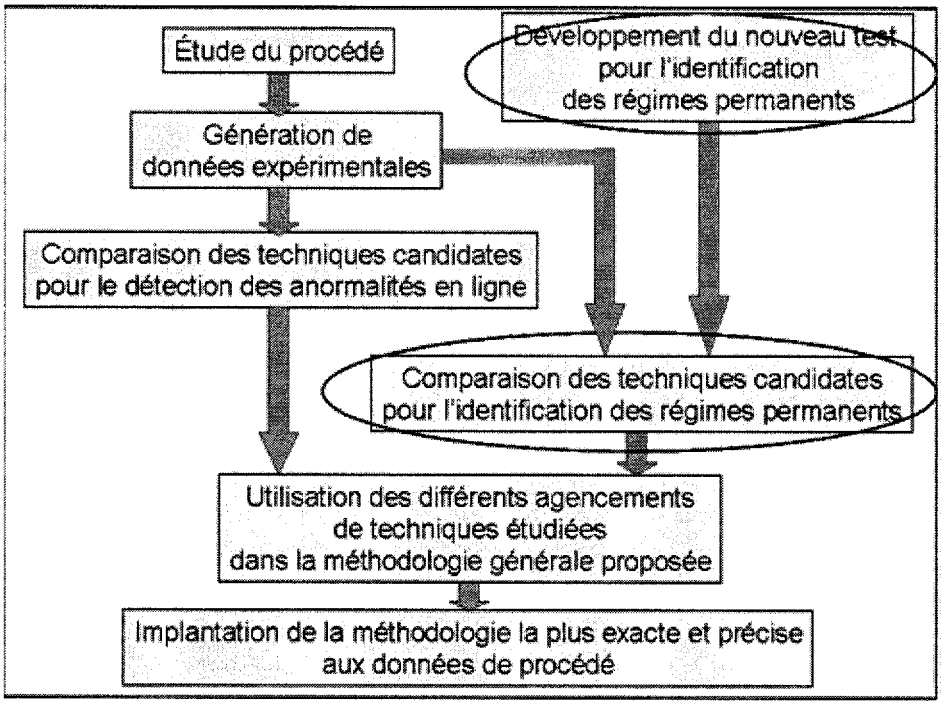


Figure 5. 4 : Parties de la méthodologie générale reliée à la troisième hypothèse

Dans cette section, une nouvelle technique couplant l'utilisation des caractéristiques associées aux transformées d'ondelettes et les théories statistiques sera développée. Son efficacité sera ensuite comparée aux techniques proposées par Cao and Rhinehart (1995) et Jiang (2003a).

Afin de juger de l'efficacité de la détection du régime permanent et de son impact sur la réconciliation, en ligne, des données en régime permanent, les taux d'erreurs de type I et de type II occasionnés par les trois méthodes proposées seront calculés et comparés. De plus, l'influence de chacune de ces méthodes sur l'amélioration de l'exactitude et de la précision des mesures prises lors de régimes permanents sera également comparée.

5.4.4 Application de la méthodologie sur des données industrielles

5.4.4.1 Méthodologie utilisée

Afin de vérifier la quatrième hypothèse de recherche reliée à la possibilité d'implanter la méthodologie proposée sur des mesures prises en temps réels dans l'industrie, la section permettant de prouver la robustesse de la méthodologie sera utilisée.

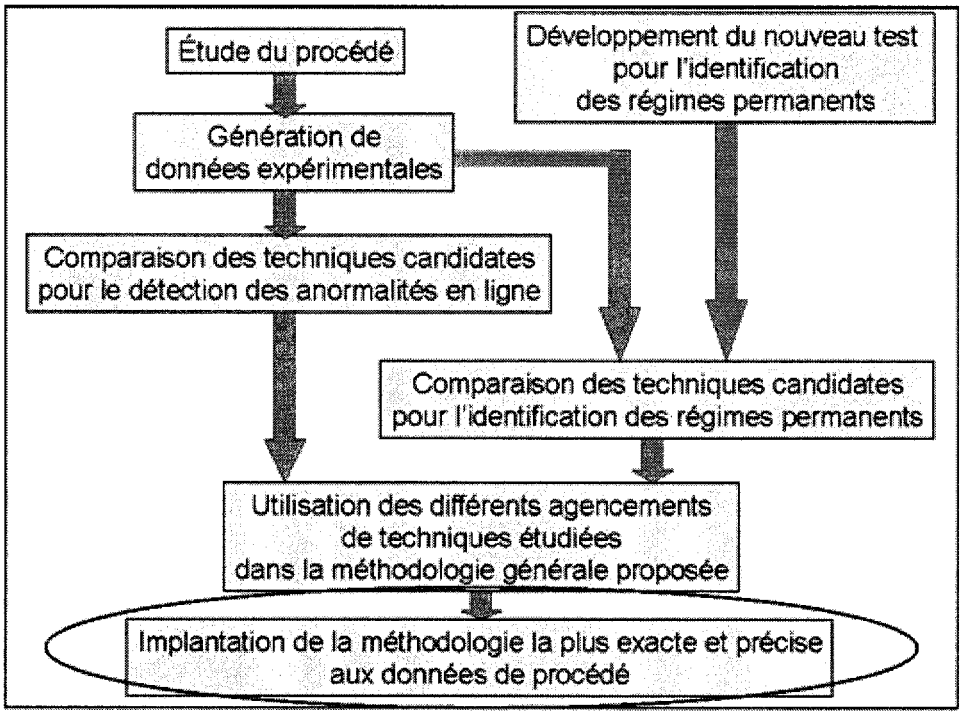


Figure 5. 5 : Partie de la méthodologie générale reliée à la quatrième hypothèse

Cette partie des expérimentations sera effectuée sans apporter de modification à la méthodologie proposée. Les mesures recueillies seront traitées directement. Ainsi la robustesse de la méthode sera analysée en même temps que sa possibilité d'implantation.

5.4.4.2 Description du procédé

Afin de tester la robustesse et la possibilité d'application industrielle de la méthode proposée, les données en temps réel d'une usine intégrée de pâte et papiers seront

étudiées. Plus particulièrement, le système d'approche d'une machine à papier, produisant deux grades de papiers journal différents, sera abordé.

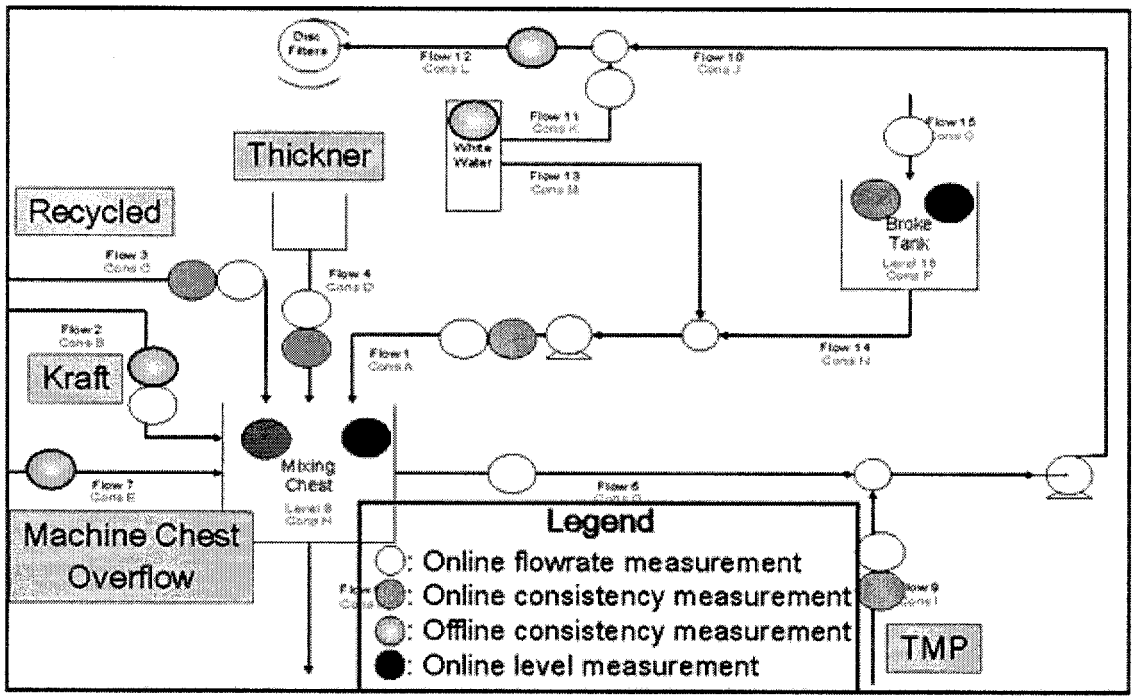


Figure 5. 6 : Système à l'étude

L'utilisation de ce système, pour prouver les possibilités d'application industrielle et la robustesse de la méthode, est très intéressante. Premièrement, le système possède suffisamment d'informations pour permettre la réconciliation des données. De plus, les données, mesurées en temps réel, possèdent tous les types d'erreurs présentés dans la partie : revue de la littérature. Aussi, le système possède plusieurs régimes d'opérations qui sont influencés par le grade de papier produit ainsi que par les caractéristiques intrinsèques au procédé. Finalement, l'amplitude des signaux propres aux différentes variables mesurées varient beaucoup selon la valeur nominale mesurée.

Chapitre 6. Méthodologie pour la réconciliation de données en ligne

Afin de permettre une réconciliation de données en ligne pour les périodes en régime permanent, il est nécessaire de mettre en place une méthodologie efficace permettant de détecter et de corriger les mesures anormales, de filtrer les erreurs aléatoires et finalement de détecter les périodes de régimes permanents.

Les différentes étapes de la préparation à une réconciliation efficace nécessitent l'utilisation de paramètres propres à l'amplitude et à la fréquence des mesures prises. Puisque ces mesures peuvent varier selon les différentes conditions d'opérations, il est nécessaire d'ajouter une étape de caractérisation des paramètres lorsque l'on veut appliquer ces étapes en ligne.

La méthodologie développée doit être robuste et permettre de réconcilier les données de manière exacte et précise. Puisque la vraie nature des données du procédé doit être connue afin de pouvoir calculer l'exactitude et la précision des données réconciliées, des données expérimentales ont du être créées.

Ces différents points ont été étudiés et soumis pour publication dans une revue scientifique. Les détails reliés à cet article et à l'élaboration de la méthodologie sont présentés dans ce chapitre.

6.1 Présentation de la méthodologie

La méthodologie proposée dans ces travaux est basée sur la méthodologie appliquée par Jiang (2003b). Cette section présente une brève récapitulation de la méthodologie proposée par Jiang et des problèmes reliés à son application directe en ligne. Dans un

deuxième temps, la méthodologie proposée dans le cadre de cette maîtrise sera présentée.

6.1.1 Méthodologie de Jiang et problèmes pour les applications en ligne

Afin de traiter les données de manière efficace, Jiang (2003b) utilise une méthodologie basée sur la correction des erreurs en plusieurs étapes (Figure 3.1). Dans un premier temps, les données extraites du procédé sont analysées. Cette première étape permet d'identifier les anormalités présentent dans les mesures. Une fois identifiés, les points anormaux sont corrigés en remplaçant la valeur supprimée par une valeur corrigée. Par la suite, afin d'éliminer les erreurs aléatoires, les données sont filtrées. Enfin, les séries de données en régime permanent sont identifiées. Jiang propose d'utiliser les caractéristiques propres aux transformées d'ondelettes dans ces étapes de préparation des données. Finalement, les données sont réconciliées et peuvent être utilisées dans différents secteurs afin d'améliorer la prise de décision.

L'utilisation de cette méthodologie nécessite une étude détaillée des séries de mesures effectuées. En effet, elle nécessite l'utilisation de plusieurs paramètres qui sont influencés par l'amplitude et la fréquence des mesures prises. Afin d'implanter cette méthode en ligne, il est donc nécessaire d'automatiser la sélection de ces paramètres.

6.1.2 Méthodologie proposée

La figure 6.1 représente la méthodologie générale proposée pour réconcilier, en ligne, les données associées à un régime permanent. Celle-ci est inspirée de la méthodologie proposée par Jiang (2003b). En effet, les étapes de détections de anormalités, de filtration, d'identification du régime permanent et de réconciliation des données y sont présentes. Notons que la présence des deux premières étapes est primordiale afin de préparer les données qui seront utilisés lors de l'étape d'identification du régime

permanent. Une détection des anormalités inefficace et/ou la présence d'erreurs aléatoires dans les données issues de la filtration engendre une identification fragmentée des régimes permanents (voir le chapitre 7 pour les détails relatifs aux anormalités et le chapitre 8 pour ceux relatifs à la filtration).

En plus des étapes relatives à l'élimination des différents types d'erreurs présentes dans les mesures effectuées, la méthodologie proposée inclue une étape d'identification du régime d'opération du système étudié. À cette étape, l'état d'opération du système est évalué grâce à un suivi de l'évolution de variables représentants fidèlement les opérations du système. Une fois que l'évaluation et l'identification du régime d'opération sont effectuées, cette étape joue deux rôles importants. Premièrement, les paramètres, contenues dans une librairie basée sur les données historiques, qui sont utilisés par les techniques permettant la correction des différents types d'erreurs sont sélectionnés ou ajustés en ligne. Par la suite, ces données sont acheminées vers les différentes étapes du traitement de données. En procédant de cette manière, les paramètres nécessaires aux différentes étapes sont ajustés en fonction de l'amplitude et de la fréquence propre à chaque régime d'opération.

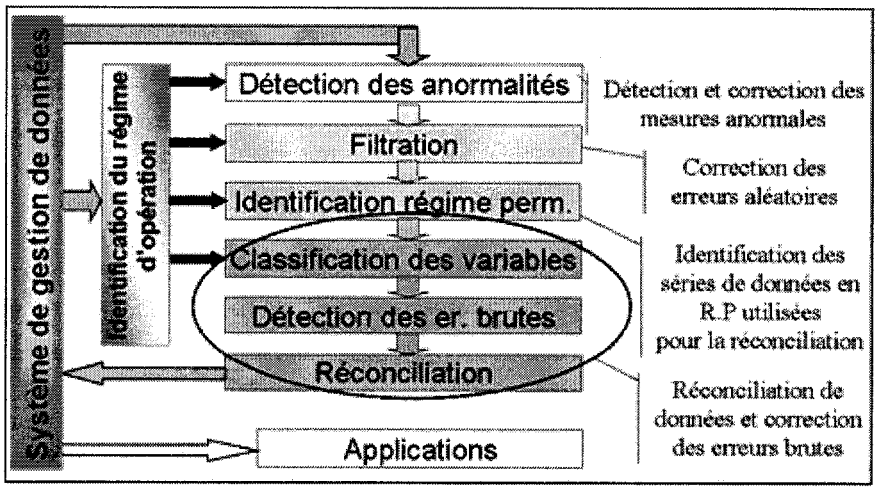


Figure 6.1 : Méthodologie proposée

6.2 Présentation du premier article

Les travaux de recherche reliés à la méthodologie décrite à la section précédente et à son implantation dans un contexte industriel ont été soumis pour publication dans une revue scientifique : Pulp and Paper Canada. La version soumise de l'article se trouve à l'Annexe VII. Les résultats intermédiaires reliés à cet article sont également disponibles en annexe :

- Détails sur la composition des différents régimes d'opération : Annexe I
- Données expérimentales créées Annexe II
- Détails des paramètres utilisés aux différentes étapes de la méthodologie
  - Détection des anormalités Annexe III
  - Filtration des données Annexe IV
  - Détection du régime permanent Annexe V
- Résultats de la réconciliation des données Annexe VI
- Résultats des calculs reliés à l'exactitude et à la précision Annexe VI

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

6.3.1 Abstract

In order to increase the value of real-time data from the process data management systems implemented at many pulp and paper mills in recent years, on-line process data treatment techniques have been evaluated in this study. The goal of applying data processing and reconciliation techniques is to correct for various sources of errors including random noise, abnormalities and inconsistencies – and thus have better quality data for decision-making. The application of these techniques on-line can potentially assist mills to quickly identify instrument maintenance requirements, to troubleshoot

operating problems, and to improve decision-making related to process activities. This paper presents a robust and practical methodology for on-line data processing and steady state reconciliation that improves the accuracy and precision of measurements when the process is operating at near steady state conditions.

6.3.2 Introduction

A variety of decisions such as planning and scheduling are generally made based on steady state data sets. To better be able to take such decisions in real time, pulp and paper mills capture real time process data in their information management systems. Janssen *et al* (2004a) showed that mill personnel recognize that the implementation of data management systems improve the accessibility to real time data but until now however, the data have generally been used for impromptu process analysis. Only in a few cases have data been used for process-wide analysis or optimization, e.g. for plant-wide multivariate analysis of processes (Harrison *et al*, 2003), for plant-wide process optimization (Dabros *et al*, 2005), or for plant-wide business modeling (Janssen *et al*, 2004b). Part of the reason for this may be attributable to data quality. Measurements are inherently corrupted with various sources of error including instrument miscalibration or malfunction, power supply fluctuation, wiring and process noise. Unless they are carefully treated, using process data to make process decisions in real time can lead to misinterpretations caused by data quality.

Figure 6.2 shows a set of flowrate measurements containing the 3 different types of error present in data. Random errors are mostly concentrated in high frequencies and can be associated with power supply fluctuation and signal conversion noise. An Abnormality is a supernatural random error. Ths high frequency and amplitude measurement can be caused by electric voltage surges and wiring problems. Gross errors are associated to bias and drifting in measurements, complete failure and precision degradation of instruments (Narasimhan and Jordache 2000), and are related to instrument wear,

corrosion, miscalibration or malfunction and inconsistencies in a set of measurements around a system, e.g. process leaks or accumulation.

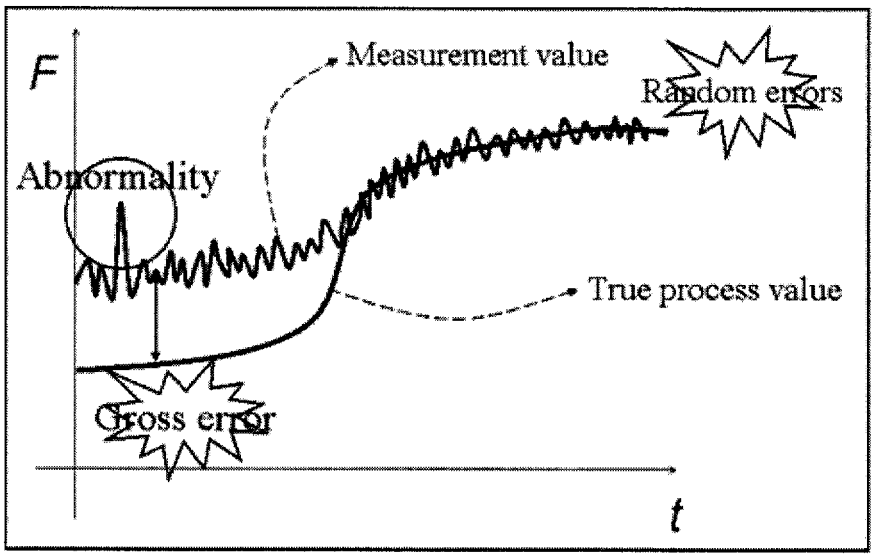


Figure 6.2 : Types of errors in online measurements

Data processing and reconciliation techniques have been applied in the pulp and paper industry for various **off-line** applications. For example Jacob and Paris (2003a and 2003b) 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. This paper proposes an on-line methodology for improving steady state data quality in **real time**.

6.3.3 Methodology

Figure 6.3 illustrates the systematic methodology used in this study based on Jiang *et al* (2003b) to reduce the various types of errors included in a set of measurements. First, abnormal measurements were detected and corrected, followed by random error correction. To correct gross errors, steady state data reconciliation was performed. For

this, near steady state data sets must be carefully identified, which was done using a novel technique based on wavelet transform and statistical theories.

Critically for on-line application, the methodology included a *system status identification*. This step has two main purposes, as follows:

- to select data processing parameters from a library which were developed based on historical data
- to identify data tuning parameters to be used in each of the different steps of the data treatment methodology, based on the measurement sources and their link to the system operation regimes.

Selection of data processing parameters and identification of tuning parameters are based on the association of unique combinations of measurements values and offline properties investigation for optimal results in the data treatment procedure.

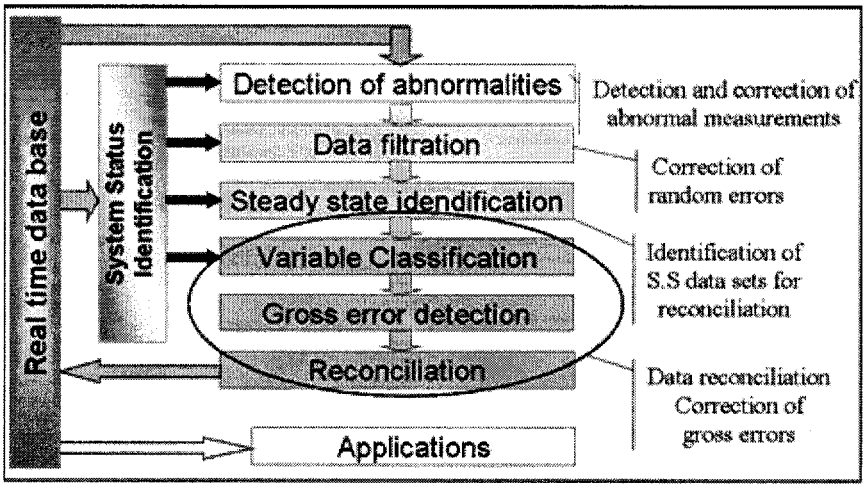


Figure 6.3 : General methodology applied in this study to improve data quality

For the first two steps in the methodology shown in Figure 6.3, data treatment using wavelet transform was compared to a) utilization of statistical theory based on the median and its absolute deviation to detect abnormal features, and b) low pass filtering to correct random errors. Due to their ability to detect abnormalities over short or long

time periods and their power to avoid distortion in the corrected signal, wavelet transforms were preferred over the alternatives (Bellec *et al*, 2004).

Data reconciliation corrects the gross errors in redundant measurements around a system by comparing their reliability using mass and energy balance relationships. Due to the importance of redundancy in data reconciliation, the first task to accomplish in data reconciliation is to classify the variables under study (Bagajewicz 2001). If the measurements are redundant, then gross error detection and reconciliation can be performed. To do so, errors between measurements and real process values are minimized by considering physical constraints and using a least square equation. Figure 6.4 illustrates this in a hypothetical case by treating measured values (100, 61, 38) to arrive at reconciled values (100, 60.3, 39.7).

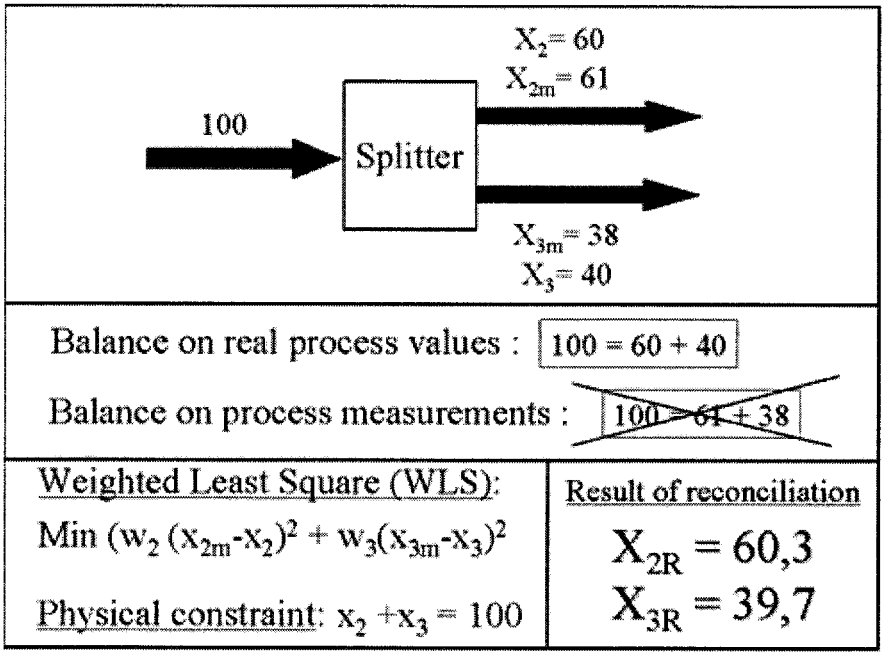


Figure 6.4 : Example of reconciliation problem around a splitter

6.3.4 Case study

The applicability and performance of the proposed methodology for on-line processing and steady state data reconciliation was tested using real time data from the stock approach system at an integrated newsprint paper mill (Figure 6.5). The mill has a process data management system (PI system supplied by OsiSoft Inc.). In order to assess and evaluate the methodology, the system was investigated in 3 steps: 1) process data analysis and creation of a hypothetical "measured" data set from a "true" data set, 2) systematic analysis of the data processing and reconciliation methodology using the hypothetical data set, and finally 3) application of the methodology on process real time measurements.

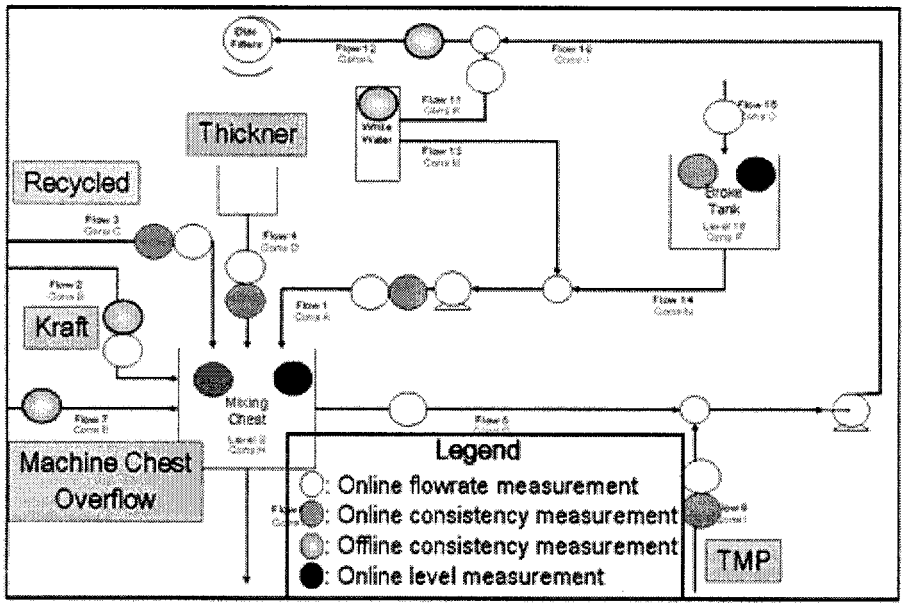


Figure 6.5 : Paper machine stock approach system studied

Before starting the detailed study around the system, a redundancy analysis was performed. By using engineering judgment to couple on-line measurements with other data, it was possible to improve the measurement redundancy and therefore the possibility for reconciling data around the system.

6.3.5 Process data analysis and creation of hypothetical "true" and "measured" data sets

To be able to characterize the accuracy and precision of the methodology, treated data must be compared with the true value of the process. Since on-line measurements are always corrupted with errors, the true value is never exactly known. Therefore, in order to investigate the accuracy of this methodology, hypothetical data were created.

In order to create hypothetical data that are representative of actual process operations, a 3 step study of the system was performed as follows.

6.3.5.1 Signal processing and near steady state operation regime identification

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 coupled to identify overall system steady state regimes. Twenty-seven different operating regimes were grouped into four system status conditions. The system status conditions were: production of paper grade A, production of paper grade B, maintenance stops, and during the occurrence of a paper machine break.

6.3.5.2 Steady state detection measurement correction through mass balances

Once near steady state operating regimes had been identified, measurements were reconciled to fit mass balances for flow and fiber content in the different process streams. After correction, the probability of occurrence, length and delays, relative to each status and regime of operation were compiled based on historical data. Figure 6.6 summarizes the twenty-seven operation regimes grouped in the four system status

conditions of the system. An example demonstrating the differences between two regimes of operation within the same status is also shown.

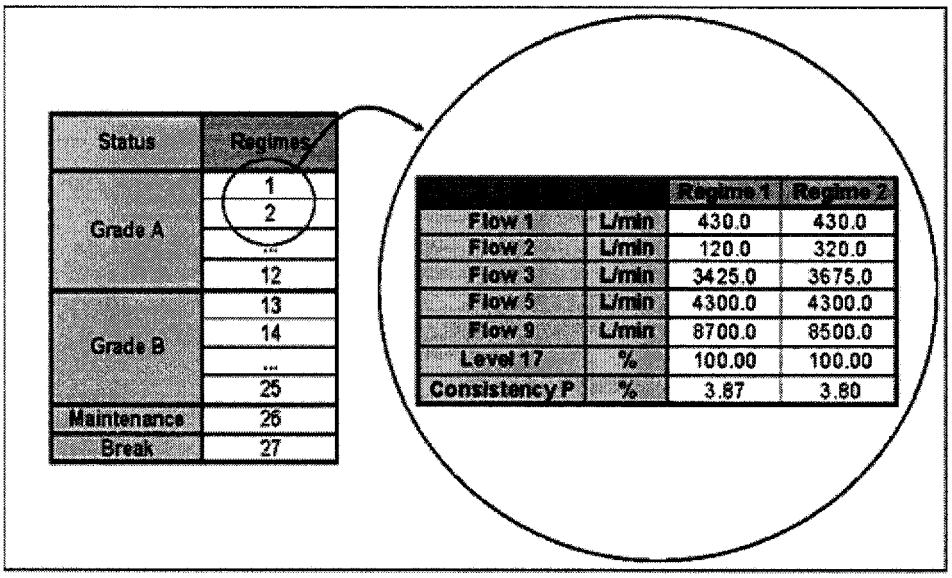


Figure 6.6 : System status and regimes of operation identified

6.3.5.3 Creation of hypothetical data with known errors

Based on the system analysis and statistics associated with each signal, hypothetical data where created. In order to generate the corrupt hypothetical set of measurements, errors were then applied to the true value. Abnormalities and random errors where added to the signals according to the standard deviation and accounting for the probability of abnormal measurement presence in the historical data.

To demonstrate how well the hypothetical data represented the real process operation, operation regimes of the system where analyzed over a one week period and reproduced by the hypothetical data. Figure 6.7 shows the result of that investigation for a particular flow rate measurement in time. The general characteristics of the two signals is the same, but one can notice that the hypothetical data can contain more or less noise in particular regions, and that part of the signal has been corrected for overall mass balances in the system.

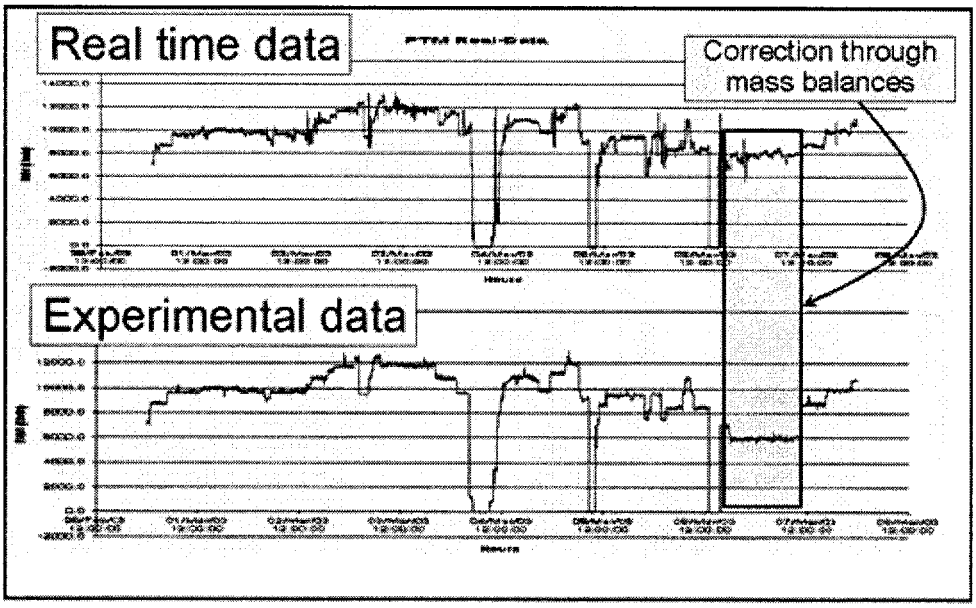


Figure 6.7 : Comparison between hypothetical and real time data for the same operation regimes

6.3.6 Investigation of Data Processing and Reconciliation Methodology Using the Hypothetical Data Sets

The various methodologies for on-line steady state data reconciliation were applied using the hypothetical data set. Figure 6.8 shows an example of the results obtained, for a particular flow rate measurement. At each step of the methodology, the raw measurements in the process being corrected are compared to the targeted value ("true" hypothetical data). In the first two steps, abnormal measurements and random errors are corrected using wavelet transform. Steady state identification is performed and steady state data sets are selected. Finally, the selected data are reconciled through mass balances on the overall system. One can notice the improvement in the accuracy of the corrected steady state measurement compare to the raw data.

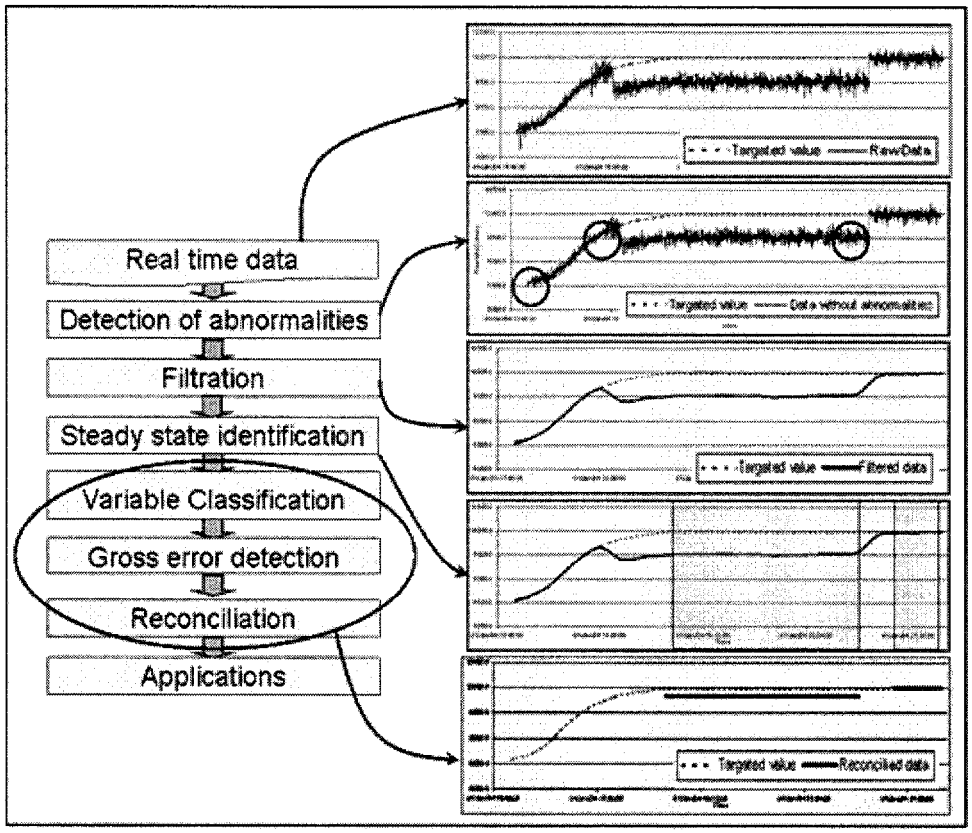


Figure 6.8 : Step by step results from the application of on-line processing and reconciliation methodology

Figures 6.9 and 6.10 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. More specifically, Figure 8 shows the relative error improvement for the different measured variables, while Figure 9 reflects the effect of the methodology on the standard deviation of measurements.

Accuracy is defined as the average of relative error (%RE) from the targeted value (Equation 1) while precision is shown as the standard deviation ($SD_{\%RE}$) of such errors (Equation 2). Therefore, lower values indicate improvements in accuracy and precision. Improvements in the accuracy underline the overall improvement of steady state data

due to the methodology, while the improvement of precision underlines the efficiency of the method for correcting gross errors.

$$\%\mathrm{RE} = 100\% * \frac{|\mathrm{Targeted.value} - \mathrm{measurement}|}{\mathrm{Targeted.value}} \quad (1)$$

$$SD_{\%RE} = \sqrt{\frac{\sum_{i=1}^{n}\left[100\% * \left(average(RE) - RE_i\right)^2\right]}{n-1}} \quad (2)$$

Through the application of the proposed methodology, it can be concluded that overall accuracy and precision has been improved. In general, the methodology has a larger effect on flowrate than on measurements of consistency. This can be explained by the fact that flowrate measurements are associated with two levels of mass balance constraints (overall and fiber mass balances), and that consistency measurements are only implicated only in the fiber mass balances.

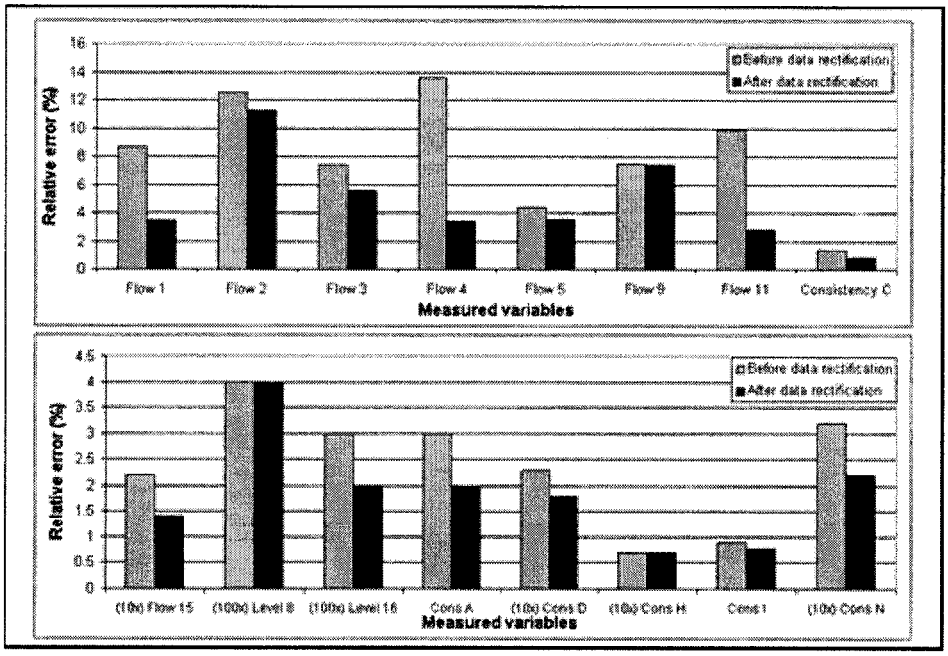


Figure 6.9 : Summary of improvement in data accuracy from the application of the data treatment methodology

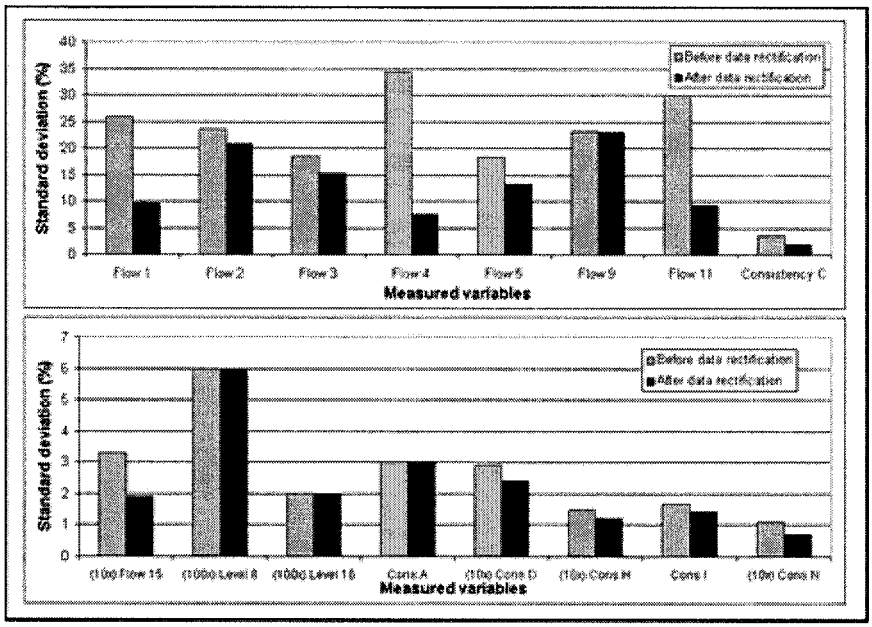


Figure 6.10 : Summary of improvement in data precision from the application of the data treatment methodology

6.3.7 Evaluation of Methodology Robustness Using Real Time Process Measurements

After proving that by applying the proposed methodology, overall accuracy and precision of hypothetical process data can be improved, real time data were treated for longer time periods. The method under investigation was found to be robust for treating real time process data using 10 flow and 8 consistency signals over a time period of 3 days. In order to ensure that process time lags were taken into account, only steady state periods over a certain time threshold were considered.

Figure 6.11 shows 3 different examples of data quality improvement by the application of the methodology. Abnormality and random noise have been corrected, and data under near steady state conditions were efficiently corrected using data reconciliation (Figure 6.11a). Instrument offset (Figure 6.11b) and total failures (Figure 6.11c) were also identified and this data corrected.

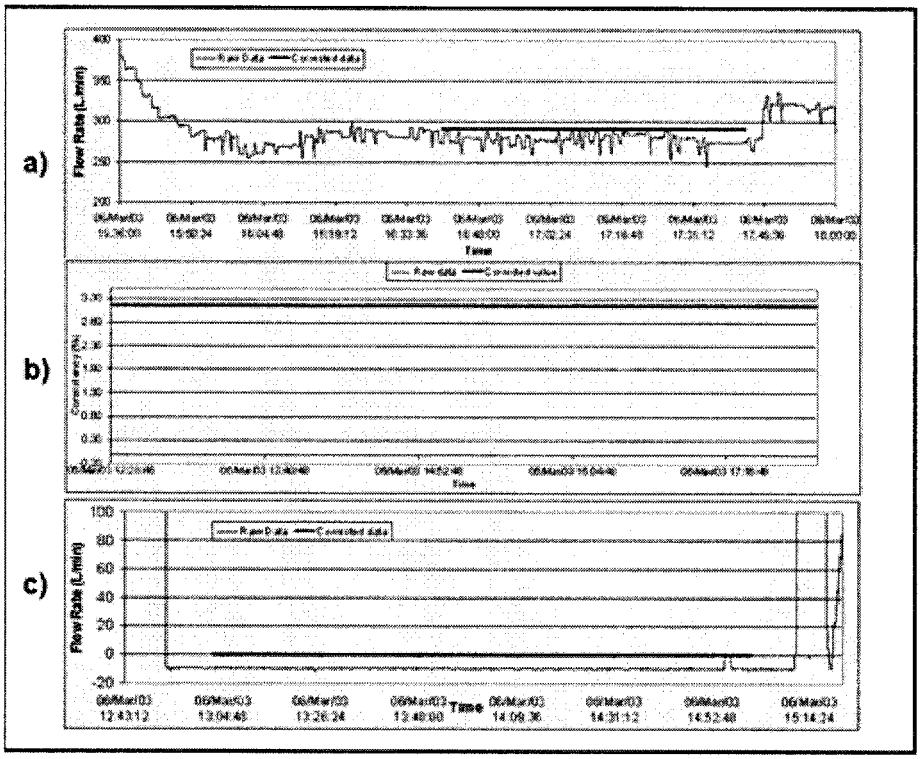


Figure 6.11 : Examples of corrections required to real time data:

a) Accuracy improvement
b) Total failure correction c) Offset correction

One should be aware that, as the system under study becomes larger, the amount of operating regimes quickly increases, and information needed for online data reconciliation will augment considerably. Therefore, process operating regimes must be carefully and logically combined in order to minimize the different operating regimes.

6.3.8 Implications of results

A novel methodology for data processing and reconciliation has been illustrated using real-time data from an integrated newsprint mill. 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 were identified, inconsistencies

were removed using data reconciliation. By applying this data treatment methodology, data accuracy and reliability were improved.

From this activity, confidence in data quality for well-defined process operating regimes can be achieved (called "direct benefits" in Figure 6.12), and this alone can likely justify the effort and expense of implementing such a system. In this section, the implications and **further benefits** of implementing such an on-line data treatment system are outlined.

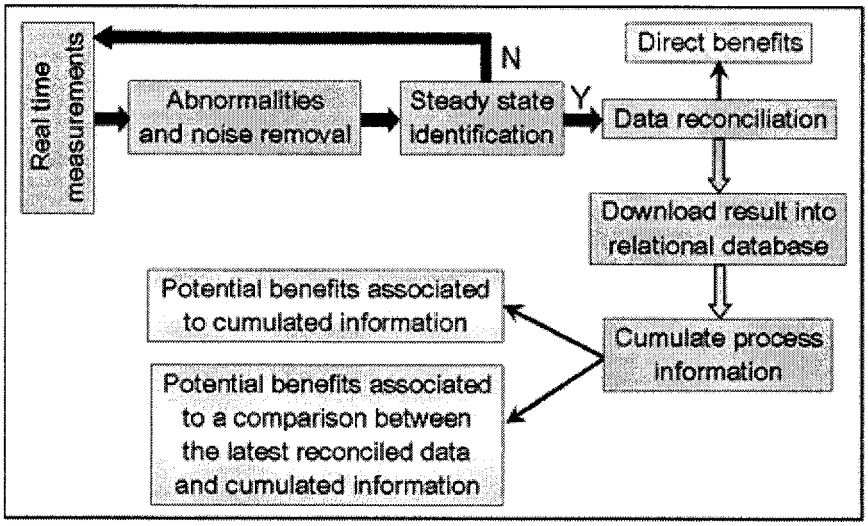


Figure 6. 12 : Types of Benefits Associated with the Application of the Proposed Methodology

By compiling on-line data reconciliation results for near steady state operating regimes, mills can develop relational databases of accurate process data. This data might be expressed as the probability of occurrence of different operating parameters. Figure 6.13 shows the example of the probability of occurrence of a flow rate measured at a process point, during a specified operating regime.

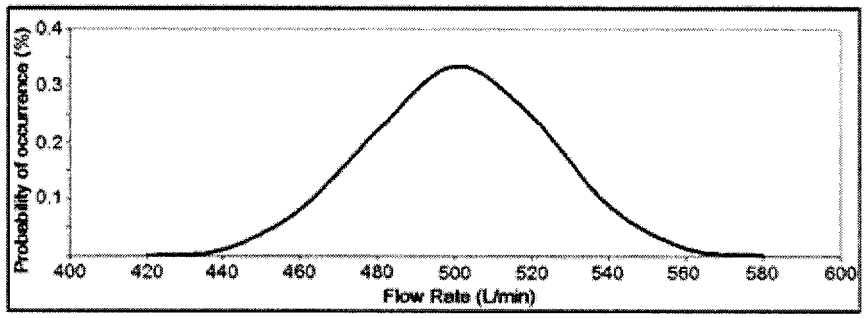


Figure 6. 13: Example of Cumulated Process Information from Successive Online Data Reconciliation Results

This could potentially lead to multiple benefits related to process operation analysis and troubleshooting. Depending on the preliminary steps necessary to diagnose potential benefits, those can be classified in two major categories (Figure 6.12):

- Benefits obtained as a consequence of changes in the data reconciliation outcome compared to many previous on-line analyses, and
- 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 addressed:

- By comparing gross error detection and correction results for the current data set with previous data sets, one can quickly identify out-of-calibration instruments and process leaks.
- One can perform real-time yield accounting (performance analysis) for near steady-state operating conditions.

In the second case, information, issued from the compiled data reconciliation results (illustrated in Figure 6.13) can be used to localize, assess and analyze process operations. The better-resolved data at different process operating regimes improves

understanding of process during near steady state operating conditions. This definition of different process operations can be used in combination with process simulation models to examine "what-if" scenarios.

The compiled data from different operating regimes issued from the reconciliation methodology can also enable more precise conditions for optimization techniques to enhance process planning (i.e. scheduling and plant-wide optimization). 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.

6.3.9 Conclusions

A robust methodology to improve steady state data accuracy and precision was assessed in this study. Improvements in data precision and accuracy were demonstrated by generating hypothetical "true" and "measured" data sets based on actual process data from an integrated newsprint mill.

The application of this methodology on-line could yield significant benefits including rapid instrument failure/process leak detection, improved process yield calculations, and better relational data for overall process optimization.

Chapitre 7. Détection des anormalités

La détection des anormalités représente la première étape permettant la préparation des données de procédé à une réconciliation en régime permanent. Cette étape est primordiale et son efficacité a un impact direct sur la détection du régime permanent et la réconciliation des données. La présence de points anormaux fausse l'allure des données associées au procédé et engendre des problèmes de détection des régimes permanents (figure 7.1). Puisque les mesures identifiées par cette détection sont utilisées dans l'étape de réconciliation, cette problématique a un impact direct sur la qualité des données issues du processus proposé. Il est donc important de départager les techniques candidates à la détection des points anormaux selon leur efficacité.

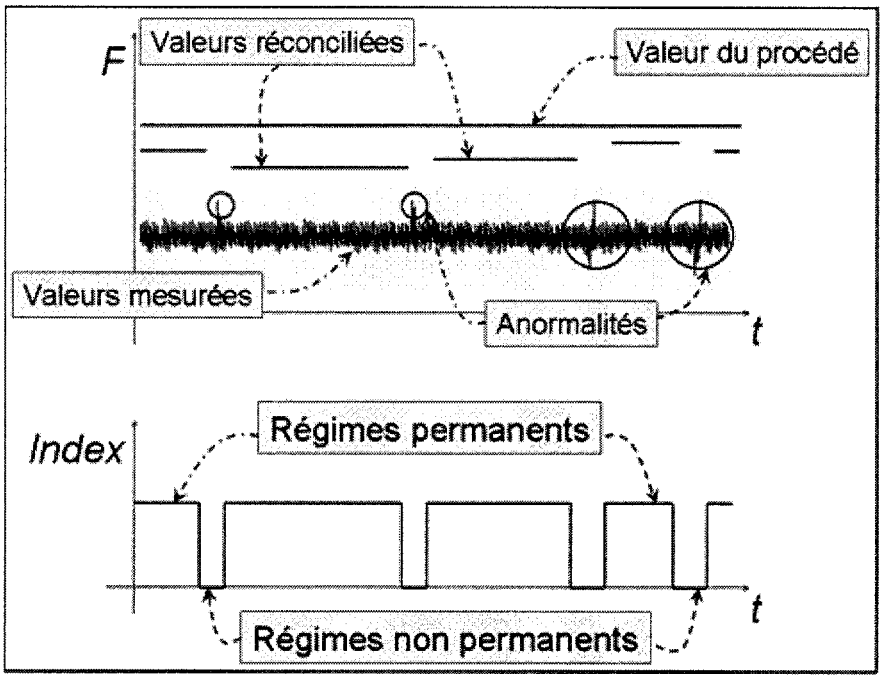


Figure 7. 1 : Effet de la présence d'anormalités sur la détection du régime permanent et sur le résultat de la réconciliation

Dans ce chapitre, deux méthodes applicables en ligne sont comparées. Dans la revue de la littérature, il a été démontré que la méthode d'Hampel est la plus efficace des

méthodes statistiques permettant de détecter les points anormaux propres à une série. De l'autre côté, Jiang et al. (2000) proposent une méthode basée sur la détection des changements anormaux présents dans une série de données. L'efficacité de ces deux méthodes n'a jamais été comparée jusqu'à maintenant.

7.1 Méthodologie utilisée pour démontrer l'efficacité des méthodes étudiées

Afin de comparer l'efficacité des deux méthodes proposées, les données expérimentales créées à partir de l'étude du système (voir chapitre 6) seront utilisées. Pour permettre d'avoir un meilleur aperçu des forces et des faiblesses relatives à chaque méthode, les anormalités ajoutées sur les données expérimentales ont été divisées en quatre catégories. Il est à noter que les anormalités de grande amplitude regroupent celles situées à 10 écart-types de la moyenne ou plus et que les mesures anormales de longue durée regroupent celles de 2 minutes et plus (au moins douze points d'échantillonnage).

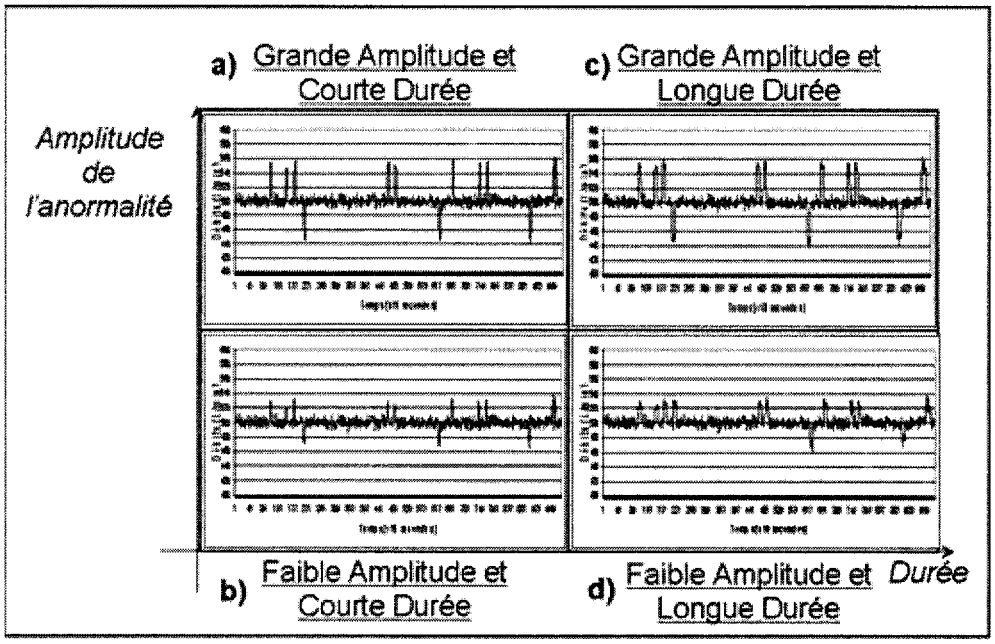


Figure 7. 2 : Quatre types d'anormalités

a) grande amplitude et courte durée, b) petite amplitude et courte durée, c) grande amplitude et longue durée d) petite amplitude et longue durée

Afin de comparer l'efficacité des méthodes, les erreurs de types I et II reliées à la détection seront compilées. Les erreurs de type I représentent le taux de mesures anormales qui n'ont pas été détectée tandis que les erreurs de type II sont associées à une fausse détection de ces mesures.

7.2 Retour sur les méthodes comparées

Les deux méthodes comparées dans cette étude sont applicables en ligne. La méthode de Hampel se base sur l'utilisation des théories statistiques reliées à la médiane pour détecter les points anormaux appartenant à une série de données. De l'autre côté, la méthode basée sur l'utilisation des ondelettes utilise les caractéristiques de ces dernières afin de détecter les changements anormaux présents dans une série de mesures. Les deux méthodes sont brièvement résumées plus bas (pour plus de détails, consulter les sections 3.2.2.1 et 3.2.2.2).

7.2.1 Méthode de Hampel

Afin d'effectuer une détection robuste des anormalités, Hampel (1974) propose d'utiliser la médiane ($\tilde{x}$) et la médiane de la déviation absolue (MDA) de cette dernière pour détecter les anormalités présentes dans une série de données. Selon lui, pour identifier la présence de points aberrants, on peu utiliser le test suivant :

$$\text{Si } (|x_i - \tilde{x}| \geq g(N, \alpha_N) * MDA), \text{ alors } x_i \text{ est un point aberrant}$$

où $g(N, \alpha_N)$ est une fonction influencée par le nombre de données utilisées (N) dans la détection et le niveau de confiance ($1-\alpha_N$) accordé aux résultas.

Davies et Gather (1993) ont effectué des travaux sur cette fonction. Ils ont simulé des essais standardisés sur le nombre d'erreur de type I. Ces essais permettent d'attribuer

une valeur de 5.67 à la fonction $g(N,\alpha_N)$ lorsque, pour un niveau de confiance de 95%, 20 données sont utilisées.

7.2.2 Méthode basée sur les caractéristiques propres aux ondelettes

La méthode de détection des mesures anormales à l'aide des d'ondelettes exploite le fait que leur transformée de premier ordre est directement proportionnelle à la première dérivée du signal représentant les données mesurées. La méthode utilise les extremums de la dérivée du signal qui correspondent aux points d'inflexions de ce dernier. En détectant les maximums et minimums locaux de la série de données, on identifie les changements soudains qui y sont reliés. En identifiant les différents couples d'extremums de grandes amplitudes et de signes opposés de $W_j f(t)$, les changements de haute intensité sont identifiés (figure 2.3). Les anormalités sont donc détectées si les deux critères suivants sont atteints:

$$Wf(p_1) \text{ et } Wf(p_2) \geq T_1$$
$$signe[Wf(p_1)] * signe[Wf(p_2)] = -1$$
$$p_2 - p_1 \leq t_p$$

où t_p représente la limite supérieure de durée d'une anormalité et T_1 est le seuil de tolérance de $Wf(t)$ pour l'identification d'une anormalité

(voir section 3.2.2.2 pour plus de détails).

7.3 Résultats et discussion

Comme il a été mentionné précédemment, l'efficacité des deux méthodes, applicables en ligne, a été testée sur les données générées (voir chapitre 6). Les erreurs de types I et II ont été compilées pour chacune des techniques. Les résultats obtenus sont résumés dans cette section. Afin de ne pas privilégier l'une ou l'autre des méthodes, la sélection de la taille de l'échantillon utilisé pour les tests a été fixée. Cet échantillon possède une

fenêtre de 20 mesures (critère N pour le test de Hampel, et t_p pour la méthode associée aux ondelettes).

Il est à noter que les résultats intermédiaires (pour les 32 variables et 43200 mesures par variable) permettant d'obtenir les résultats présentés dans cette section sont disponibles à l'annexe III. De plus, les résultats de l'analyse de sensibilité permettant de déterminer le paramètre λ_1 utilisé dans la détection à l'aide des ondelettes sont également disponibles (voir l'annexe III). La valeur issue de cette analyse et qui est utilisée pour la comparaison est : $\lambda_1 = 0.95$.

7.3.1 Erreur de type I

Une erreur de type I est effectuée lorsqu'une mesure anormale est présente dans l'échantillon testé, mais qu'elle n'est pas détectée. On peut obtenir le taux d'erreur de type I à l'aide de l'équation suivante :

$$\tau_{TypeI} = 100\% * \frac{n_{anormalités\_non\_détectés}}{n_{anormalités\_présentes}} \quad (34)$$

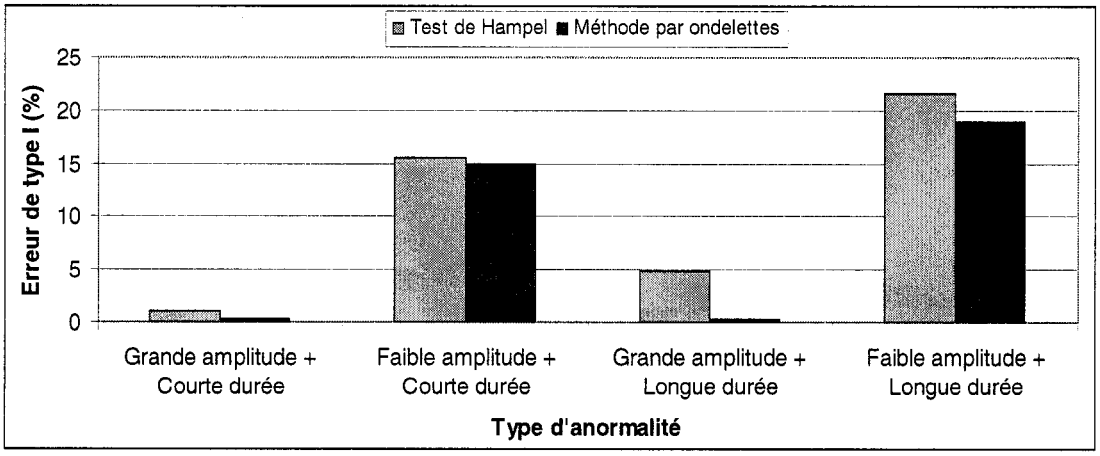


Figure 7. 3 : Taux d'erreurs de types I associé aux techniques de détection des anormalités

À l'aide de la figure 7.3, on remarque que les deux tests ont des performances semblables lorsqu'ils sont utilisés pour détecter des périodes anormales relativement courtes. Cependant le test de Hampel éprouve plus de problèmes que la méthode basée sur les ondelettes à détecter les longues périodes anormales. Cette constatation peut être expliquée par un point de brisure, associé à l'utilisation de la médiane dans le test de Hampel, d'une valeur de 50% (voir section 3.2.2.1). Cette affirmation signifie que lorsque plus de 50% des données contenues dans l'échantillon testé son anormales, la médiane est affectée par la présence des ces points. Ce phénomène provoque un déplacement de la médiane vers les points anormaux et rend leur détection inefficace (figure 7.4). Puisque dans ce cas-ci, les anormalités sont jugées de longues durées lorsqu'elles sont d'une durée minimale de 12 mesures, le point de rupture est passé et le test n'est plus robuste.

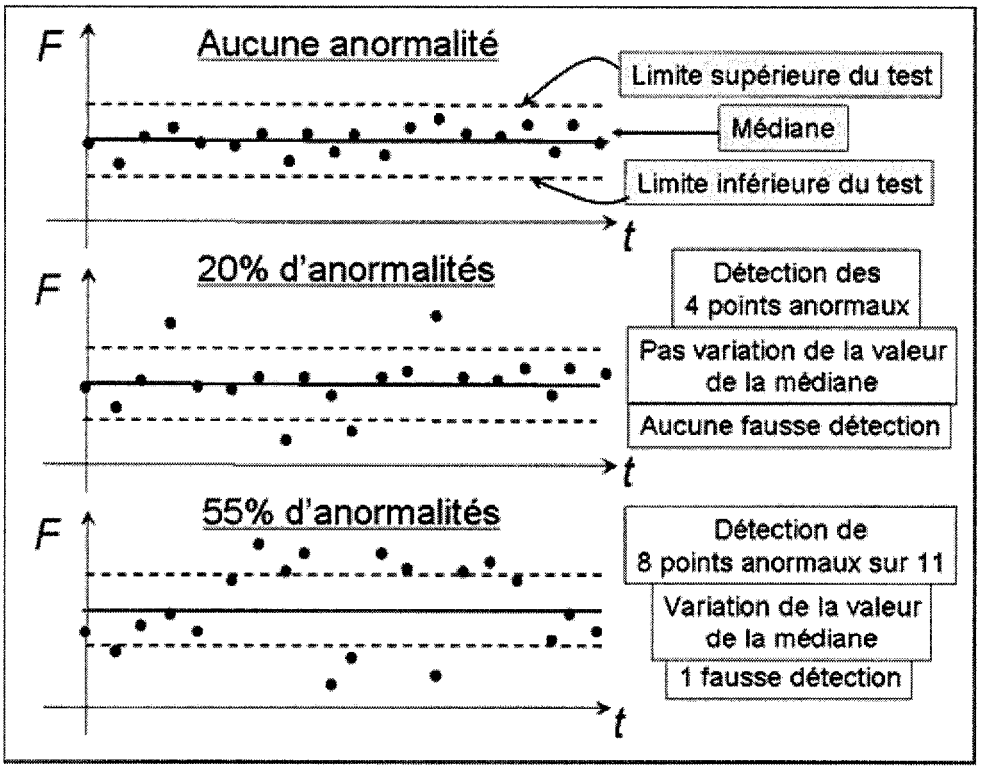


Figure 7. 4 : Effet de la présence de plus de 50% d'anormalités dans l'échantillon étudiée sur le test d'Hampel.

7.3.2 Erreur de type II

Une erreur de type II est effectuée lorsqu'une mesure anormale est détectée, mais quelle est appartient en fait à une distribution normale. On peut obtenir le taux d'erreurs de type II à l'aide de l'équation suivante :

$$\tau_{TypeII} = 100\% * \frac{n_{fausses\_détections}}{n_{points\_distribués\_normalement}} \quad (35)$$

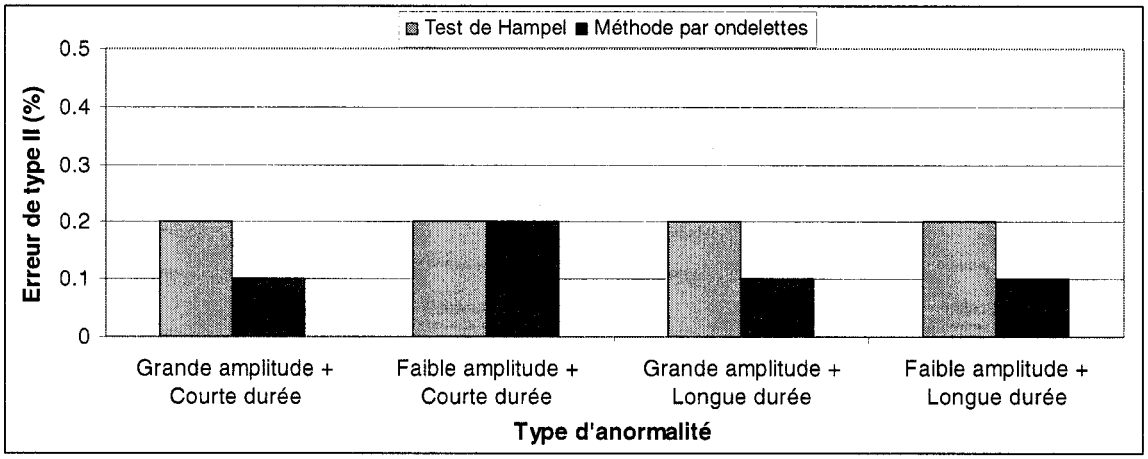


Figure 7. 5 : Taux d'erreurs de types II associé aux techniques de détection des anormalités

La figure 7.5 indique quel le taux d'erreurs de type II reliées à la détection des anormalités à l'aide de l'une ou de l'autre des techniques varient entre 0.1 et 0.2%. À l'aide de ces résultats, on peut affirmer que les deux techniques étudiées n'effectuent pas de fausse détection en grande quantité.

Chapitre 8. Identification des périodes de régime permanent

L'étape d'identification des données en régime permanent précède les étapes relatives à la réconciliation des données. Cette étape permet de sélectionner les données qui seront utilisées afin de réconcilier les mesures et, par le fait même, d'éliminer les erreurs brutes présentes dans la données en temps réel.

8.1 Présentation du second article

L'article inclus dans cette section présente une nouvelle méthodologie d'identification des régimes permanents. Cette identification est basée sur l'utilisation de filtre et des caractéristiques propres aux ondelettes. La version soumise de cet article à Computers & Chemical Engineering est disponible à l'Annexe VII. Les résultats intermédiaires reliés à cet article sont également disponibles en annexe :

- Données expérimentales créées Annexe II
- Détails des paramètres utilisés aux différentes étapes de la méthodologie
  - Détection des anormalités Annexe III
  - Filtration des données Annexe IV
  - Détection du régime permanent Annexe V
- Résultats des calculs reliés aux erreurs de type I et II associés à l'identification des régimes permanents Annexe V
- Résultats de la réconciliation des données Annexe VI
- Résultats des calculs reliés à l'exactitude et à la précision Annexe VI

8.2 Steady state identification for online data reconciliation based on wavelet transform and filtering

8.2.1 Abstract

In this paper a process steady state detection technique suitable for on-line application is developed, based on wavelet transform (WT) and filtering. First, measurement abnormalities and high frequency components are removed by using a wavelet based multi-scale processing technique. Then steady state is detected in 3 steps, as follows. 1) The start point of the steady state period is detected by using the first order WT and its first derivative. 2) Steady state duration is approximate by using an f-test to compare the ratio of the filtered square deviation from the mean and the filtered squared difference of successive data. 3) Finally, the end point of the steady state period is identified by using first order WT features. Compared to other methods tested, this method is efficient at identifying steady state periods, improved steady state data reconciliation results, and is suitable for on-line implementation. A case study based on data from the stock approach system of a paper machine is presented to illustrate the method.

8.2.2 Introduction

In recent years, the emergence of plant-wide data management systems has improved accessibility to real time data and consequently, the possibility of enhancing value extracted from data by its analysis, e.g. for plant-wide multivariate analysis of processes (Harrison *et al*, 2003), for plant-wide process optimization (Dabros *et al*, 2005), or for plant-wide business modeling (Janssen *et al*, 2004b). Unfortunately, measurements are inherently corrupted with various sources of errors, e.g. instrument miscalibration or malfunction, power supply fluctuation, wiring and process noise. Therefore, real time data must be processed and reconciled before being used in these

applications. There are still additional benefits possible from *on-line* data processing and reconciliation. In the present study, methods for on-line determination of near steady state process conditions are addressed.

False detection of the process steady state can lead to misinterpretation of true process features, especially if the incorrect steady state data are subsequently reconciled. Under-estimating the true process steady state periods can result in only partial correction of gross errors (Figure 8.1a), while over-estimating steady state periods can lead to false input to data reconciliation (Figure 8.1b).

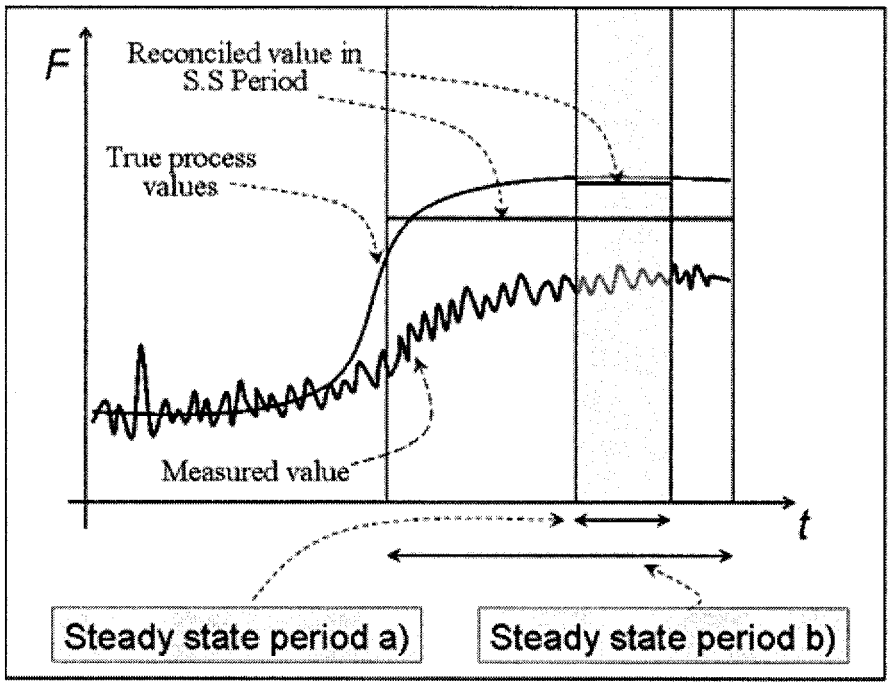


Figure 8. 1 : Inaccurate Estimation of Steady State Periods

a) Under-Estimation b) Over-Estimation

A variety of techniques for on-line process status identification have been proposed in the literature. Bakshi and Stephanopoulos (1995) developed a geometric approach for the description of process trend. Cao and Rhinehart (1995) proposed a steady state identification technique based on low pass filtering technique and statistical theory. Flehmig *et al* (1998) used wavelet transform features to approximate process

measurements by a polynomial of limited degree and identify trend in process. Nounou and Bakshi (1999) used wavelet features to identify and remove random and gross errors. More recently Jiang *et al* (2003a) used 1st and 2nd wavelet transform features to correct process measurements and extract process trends. In the so-called direct approach, linear regression of the measured value is calculated over a data window, and a t-test is performed on the regression slope. This approach is executed over a specified time period, which is not ideal when dealing with real time measurements.

Cao and Rhinehart (1995) proposed a method based on the comparison of data variances calculated in different ways. In this method, a weighted moving average is used to filter the sample mean. Then, the filtered mean square deviation from the new mean is compared with the filtered squared difference of successive data. This method uses a low pass filter to estimate the mean value. On the one hand, the computational requirements and storage are significantly reduced. On the other hand, low pass filters are less sensitive to abnormalities, and cause time delays in transient periods. Furthermore, using a weighted average to filter the calculated variances creates a delay in the characterization of process measurement frequency. These delays cause problems of detection in periods where the frequency varies in real time measurements. Figure 8.2 shows the effect of a sudden change in signal amplitude on steady state identification when using a low pass filter to estimate the signal variance. One can notice that the presence of frequency variation causes a problem of identification at the moment of such variation in the measurement signal.

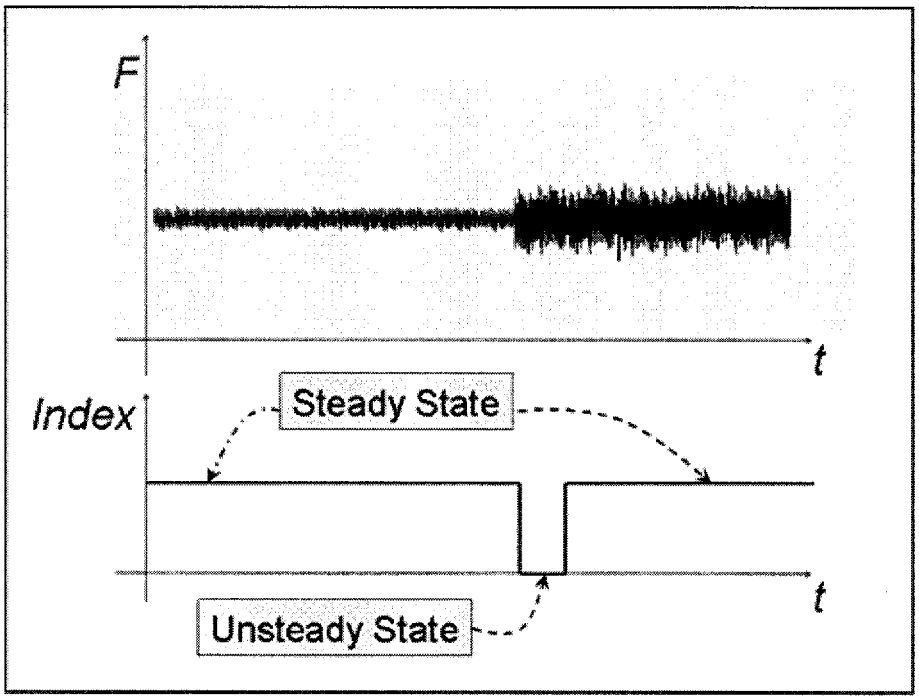


Figure 8. 2 : Steady State Detection Problems Related to Method of Filtering Variances by a Low Pass Filter

Jiang *et al* (2003a) proposed a wavelet transform (WT) based method for the detection of near steady state periods. The wavelet based multi-scale data processing technique is used to eliminate random noise and abnormalities. Then, the process status is analyzed according to the modulus of the first and second order wavelet transforms. The method has the ability to analyze the high frequency components and abnormalities accurately. When applying the multi-scale method, the accurate choice of scale is critical. When selecting a scale that is too low, the WT will be corrupted by high frequency noise (Figure 8.3a), i.e. process status identification is corrupted by temporal features. If a too high is selected, then process measurements are excessively smoothed which creates distortion (Figure 8.3b). This creates a deviation from the true process trend and leads to an incorrect reflection of process status. Jiang *et al* (2003a) proposes to select the optimal scale by taking in consideration the response time constants and sampling intervals. This criterion is adequate for off-line purposes, but not practical for on-line treatment of real time data because on-line measurements can be corrupted with

different high frequency features over time. Therefore, the scale choice must be known *a priori* for on-line wavelet based treatment of real time data. Furthermore, this method uses the second order WT of the signal to distinguish zero-crossing points from steady state periods. The second WT is directly proportional to the second derivative of the smoothed signal at the sample cutting scale. It is adequate to represent process trend but requires high computational storage and requirements.

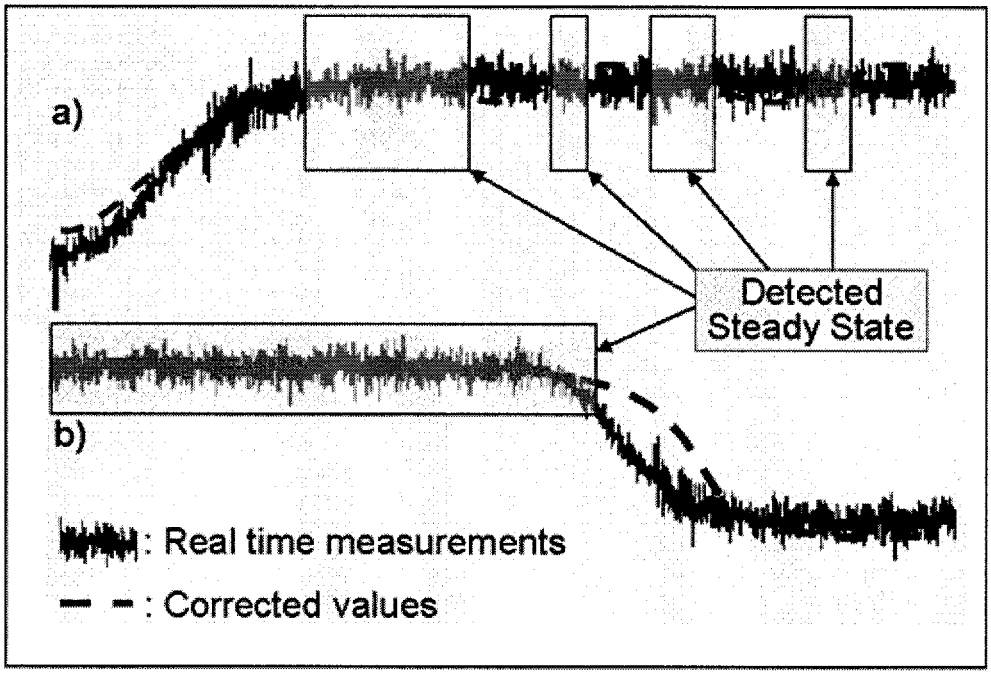


Figure 8. 3 : Scaling Choice for On-Line Data Treatment

a) Impact of Under-Scaling on Steady State Identification
b) Impact of Over-Scaling on Steady State Identification

In this paper, a steady state detection technique for on-line estimation of process status is proposed. First, by using multi-scale wavelet data processing (coupled with historical process data analysis in order to select appropriate wavelet cutting scale), random noise and abnormalities are eliminated. Then, the process status is evaluated by a 3-step method based on wavelet transform and statistical theory. The identification of the steady state period start point is identified using wavelet transform and its first derivative. Then the steady state duration is approximated by coupling an f-test with

filtration. Finally the end point of the period is identified by using wavelet transform features.

8.2.3 Elimination of random noise and abnormalities

High frequency process measurement features corrupt the process trend, and have a direct impact on process status identification. Jiang *et al* (2000) proposed a multi-scale wavelet method for processing measurements which is effective for removing noise and detecting abnormalities in real time data sets. The approach for steady state detection presented in this study builds on this multi-scale wavelet methodology.

8.2.3.1 Representing the process measurements features by wavelet multi-scale approach

The objective of the multi-scale wavelet technique is to represent a real time measurement as a limit successive approximation (Figure 8.4). First, the process measurement signal ($\Sigma C_{0,i}\varphi_{0,i}$) is decomposed into smoothed ($\Sigma C_{1,i}\varphi_{1,i}$) and detailed signals ($\Sigma D_{1,k}\Psi_{1,k}$). $\Sigma C_{1,i}\varphi_{0,i}$ represents the low frequency feature of the process measurements, while $\Sigma D_{1,k}\Psi_{1,k}$ represents the higher one. In subsequent scales(j), the smoothed signal of the previous scale ($\Sigma C_{j-1,i}\varphi_{j-1,i}$) is also decomposed in a smoother signal ($\Sigma C_{j,i}\varphi_{j,i}$) and a lower frequency detailed signal ($\Sigma D_{j,k}\Psi_{j,k}$). Multi-scale wavelet cuts the real time measurements signal into different frequency components represented at each scale. Therefore it is critical to identify the proper scale(j) that allows noise elimination in the detailed signals, and does not remove the true trend (Mallat and Hwang 1992).

$$f(t) = f_0 = \sum_{i \in I_0} C_{0,i}\varphi_{0,i} = \sum_{i \in I_1} C_{1,i}\varphi_{1,i} + \sum_{k \in K_1} d_{1,k}\psi_{1,k}$$
$$= \sum_{i \in I_2} C_{2,i}\varphi_{2,i} + (\sum_{k \in K_1} d_{1,k}\psi_{1,k} + \sum_{k \in K_2} d_{2,k}\psi_{2,k})$$
$$= \cdots$$
$$= \sum_{i \in I_j} C_{j,i}\varphi_{j,i} + (\sum_{k \in K_j} d_{j,k}\psi_{j,k} + \sum_{L=1}^{j} \sum_{k \in K_L} d_{L,k}\psi_{L,k})$$

Equation 1 [from Jiang *et al* (2000)]

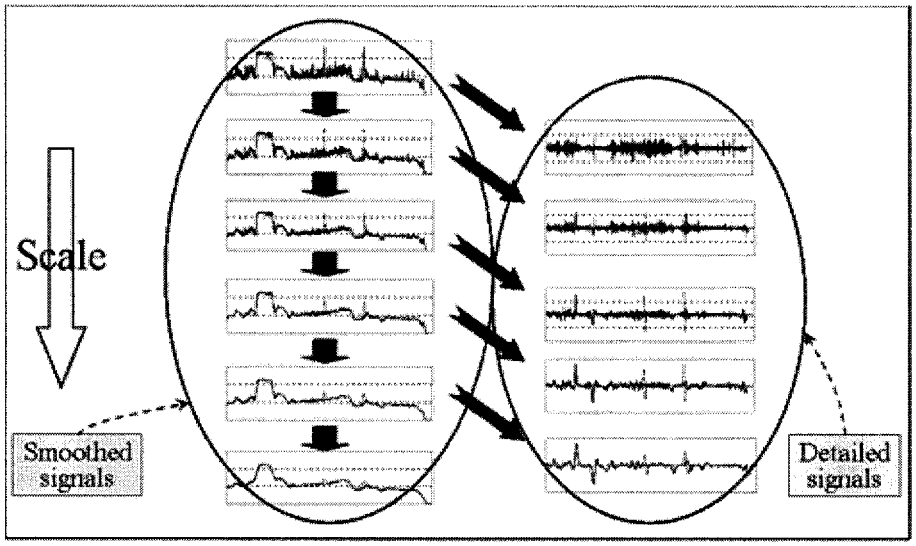


Figure 8. 4 : Multi-Scale Wavelet Representation of Real-Time Measurements

8.2.3.2 Identification of abnormal measurements

Abnormalities are defined as supernatural peeks of short duration. In other words, they are errors represented as large changes at high frequency. Such changes in real time measurements can be detected by using the first order WT, which is proportional to the first derivative of the smoothed signal (Equations 2 and 3). Since the extrema of the first derivative indicates fast changes in the function under study, one can detect fast changes in a set of measurements using the first order WT (Jiang *et al* 2003a).

$$\text{for first order WT: } \psi_j(t) = \frac{d\varphi_j(t)}{dt} \quad (2)$$

$$WT_j f(t) = f * \varphi_j(t) = f * (2^j \frac{d\varphi_j}{dt})(t) = 2^j \frac{d}{dt}(f * \varphi_j)(t) \quad (3)$$

8.2.3.3 Denoising based on wavelet multi-scale features and historical process data.

Real time measurements contain noise at higher frequency than the true process trend. Therefore multi-scale wavelet decomposition of process measurements is effective for denoising if the correct scale is employed. Optimal scale choice for signal denoising is based on the response time, which cannot be predicted in real time and thus is not useful for real time data application. The technique proposed in this paper bases the scale choice on historical data, and assumes that the scale choice is time invariant.

As shown previously, choosing a WT scale that is too high creates a distortion of process measurements, and leads to an inaccurate reflection of process trend. On the other hand, choosing a scale that is too low will leave the smoothed signal with some random noise. This second possibility does not affect the true process trend. At the chosen scale, the process trend is still available but corrupted with higher frequency measurements. Therefore, in a subsequent step the corrupted smoothed signal can be refiltered to isolate the process trend from higher frequency perturbations. This opportunity is not possible in the case where scale is over-estimated.

One should choose a scale that does not affect the process trend. By studying historical process measurements, one can investigate the optimal cutting criterion for different process operation. The scale employed for on-line implementation can be selected in such a way that high frequency features are mostly deleted, and the true process trend is not affected by signal distortion.

8.2.4 Steady state detection based on WT features, filtering and statistical theories

8.2.4.1 3 Steps methodology to perform steady state detection

Once abnormal measurements and high frequency features are removed, real time data steady state detection can be performed on-line. To do so precisely, a 3-step methodology based on wavelet transform and statistical theory is proposed, as follows:

1. The start point of the steady state period is detected by using WT characteristics and its first derivative,
2. High frequency features of the measured signal, which were not eliminated in the first step, are removed by filtering,
3. Finally, steady state period end point is detected through WT features analysis.

8.2.4.1.1 Detecting steady state starting point with WT features and the first derivative

The first order WT is proportional to the first derivative of the smoothed signal. Hence the WT measures variation in the smoothed signal and can be used to represent process variation (Jiang *et al* 2003a). WT extrema indicate fast changes in the data set, while near zero values indicate the presence of a slow change or a zero crossing point of the WT.

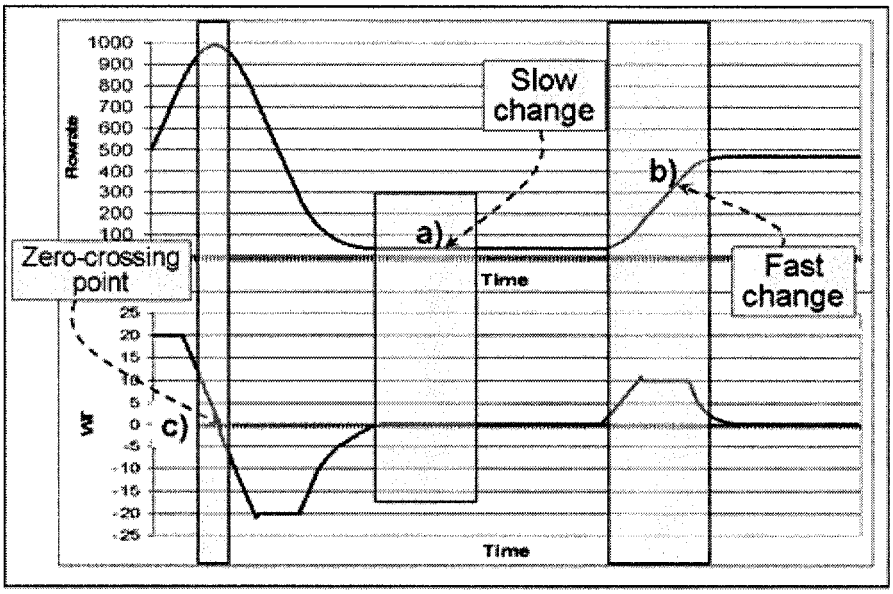


Figure 8. 5 : Using Wavelet Transform Features to Represent Process Trend

a) Slow change in process variation: 1st WT an its slope are near 0 values
b) Fast change in process variation: 1st WT value is not near 0
c) Zero-crossing point: 1st wavelet value is near zero but not the slope value

As shown in Figure 8.5, a zero crossing point corresponds to a peak in the process trend. Therefore the first derivative slope at that point is not equal to zero, as opposed to the case of steady state data. Based on the above information, the steady state starting point can be detected when the following equations are verified for the first time following a transient period:

$$\begin{cases} |WT(f)| < \alpha_1 \\ \left|\frac{dWT(f)}{dt}\right| < \alpha_2 \end{cases} \quad (4)$$

Figure 8.6 shows the methodology used to detect steady state starting point by taking advantage of wavelet transform features. Starting point is identified when steady is detected when the first order WT and its first derivative are simultaneously at near zero values. One can also notice that zero-crossing points are discarded by using this approach.

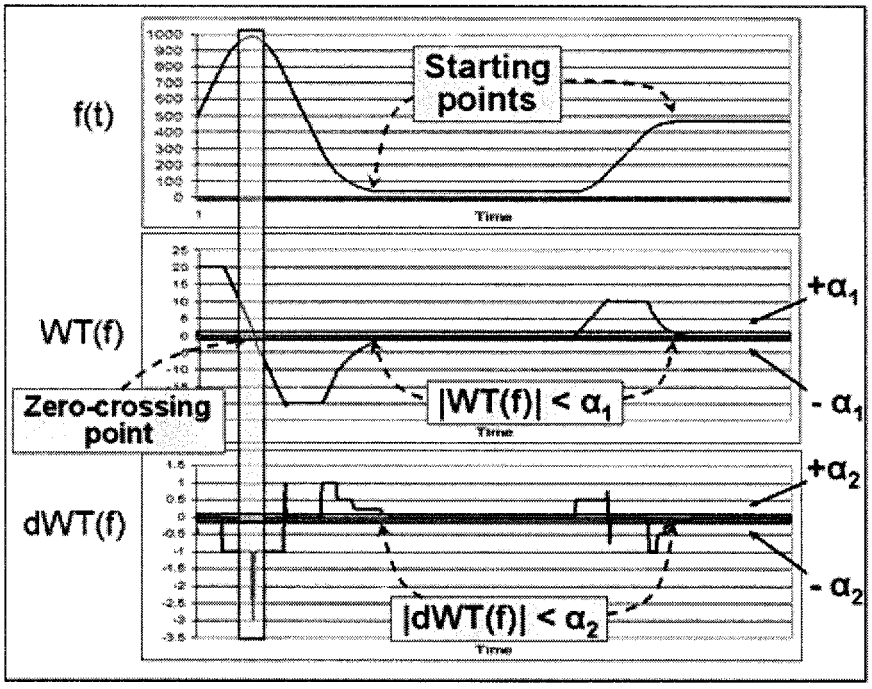


Figure 8. 6 : Detection of Steady State Starting Point Using WT and First Derivative

8.2.4.1.2 Filtering steady state data

As discuss previously, over a long time period, the WT scale choice for high frequency features removal must be done carefully. Since the on-line scaling choice was done in order to insure that distortion is not incorporated in the signal, at certain points in time, some high frequency component will not be removed from process measurements. The oscillation created by this phenomenon can lead to false detection by the WT method. Figure 8.7 shows that when high frequency residuals exceed the minimal threshold acceptance, steady state identification is partial. Therefore, using the method described in the previous section is not suitable for on-line application over a long period when selecting the scale as proposed.

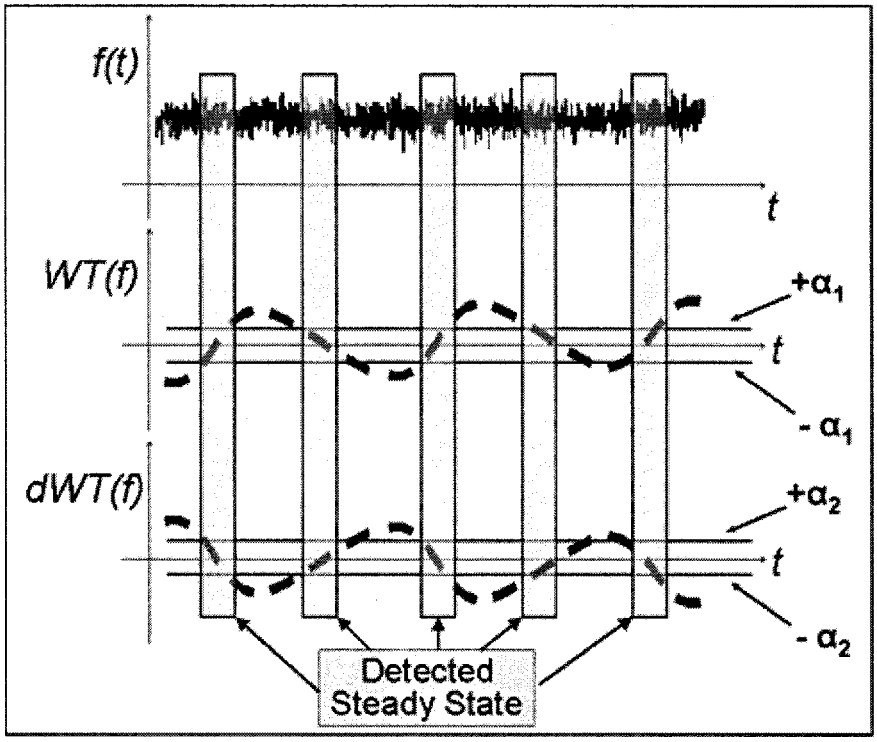


Figure 8. 7 : Problems Related to Detection of Slow Changes Trend over a Long Period

To overcome this problem, the steady state period can be approximated through a statistical test. An f-test on the ratio of the filtered squared deviation from the mean and the filtered squared difference of successive data is proposed. Before performing the test, an average value is approximated at each point in time by using a conventional moving average filter (Equation 5). Since this can create time delays and distortion in the corrected signal, such filters are not suitable for transient period correction, but are useful over steady state analysis.

$$\bar{x}_i = \beta_1 x_i + (1-\beta_1)\bar{x}_{i-1} \quad (5)$$

Figure 8.8 shows the basic concept behind the f-test proposed by Cao and Rhinehart (1995). "*A*" represents the deviation from the mean, while "*B*" represents the difference between two successive data. To avoid corruption by high frequency features that remains in the signal, the two estimators are filtered (Equations 6 and 7) and their ratio is calculated and compared to the threshold value of 1 (Equation 8). At this step of the

analysis, real time measurements have already been treated through a wavelet processing method. Therefore, all features with frequencies higher than a certain threshold have been eliminated, and utilization of a low pass filter on the estimators does not affect the steady state method.

$$D_{1,i} = \beta_2 A_i + (1-\beta_2) D_{1,i-1} \quad \text{where} \quad A_i = x_i - \bar{x}_{i-1} \quad (6)$$

$$D_{2,i} = \beta_3 B_i + (1-\beta_3) D_{21,i-1} \quad \text{where} \quad B_i = x_i - x_{i-1} \quad (7)$$

$$R_i = \frac{D_{1,i}}{D_{2,i}} < 1$$

if , then the process is near steady state (8)

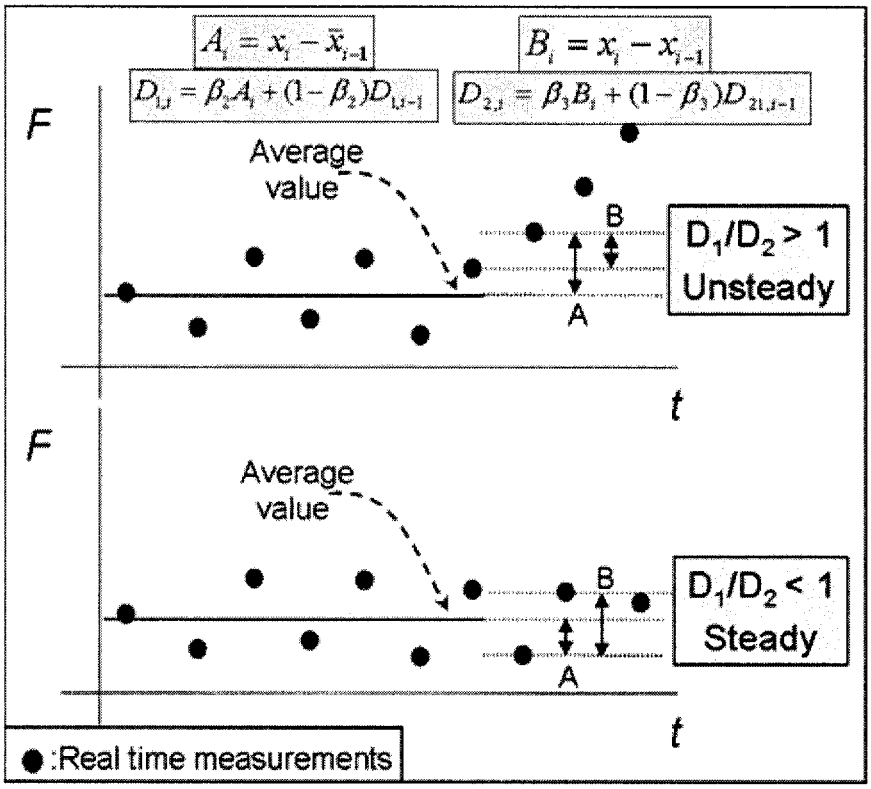


Figure 8. 8 : Representation of Deviation from Mean and Difference in Successive Data Points

Figure 8.9 shows, that as transient period evolves, distance between the measured value and filtered average (A_i) is increasing considerably. A low value of β_2 is selected in order to over-estimate the steady state period. In this way, D_1 and R variation will be

minimized, even in the beginning of the transient period. In the third step of this methodology, the selected the over-estimation of the steady state period is corrected.

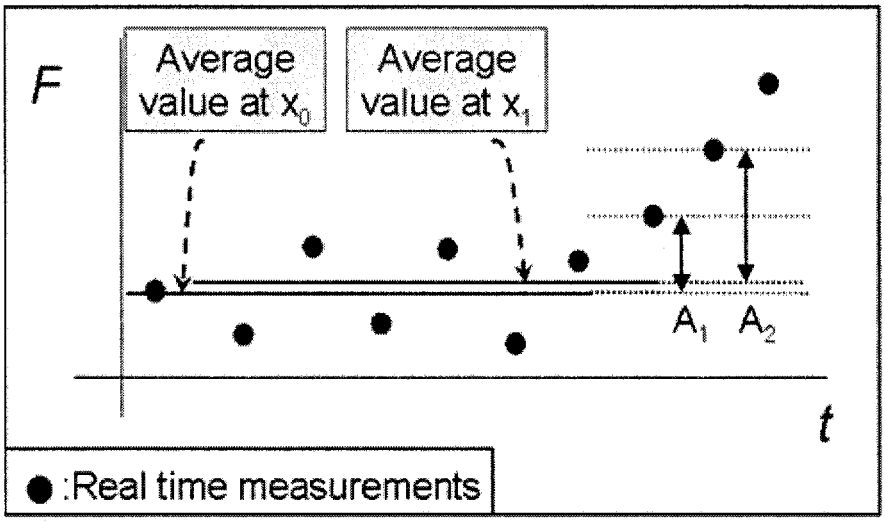


Figure 8. 9 : Evolution of Deviation from Average Value (A) During Transient Periods

8.2.4.1.3 Detecting steady state ending point with WT features

The final step of the methodology is to identify the end point of the steady state period. To do so, the last portion of the period identified in the previous step is analyzed. The WT of first order is use to detect the last moment before process status becomes transient. Since the WT value is proportional to the rate of change in the measurements one can detect, by analyzing the data backwards in time, the ending point of the steady state period by selecting the first value of WT under a certain threshold:

$$|WT(f)| < \alpha_3 \quad (9)$$

Since the selected set of data is already at steady state, no analysis related to the zero-crossing is needed here.

8.2.4.2 Selecting the threshold values and filtering parameters for steady state detection

In order to take advantage of the proposed methodology, an accurate selection of appropriate threshold values and filtering parameters are critical. Since the selection of these parameters is done based on the degree of fluctuation of process measurements, threshold values and filtering parameters can be selected *a priori* based on historic measurements.

8.2.4.2.1 Selection of threshold values (α_1, α_2 and α_3)

α_1 and α_2 are used to identify the steady starting point while α_3 is used in the detection of the period's ending point. More specifically, α_1 and α_2 are respectively used to ensure that WT modulus and its first derivative are at near zero values while α_3 is helpful to characterize process features in the WT modulus at the end of the steady state period.

The first two threshold values are used to detect steady state starting point. These are selected according to the degree of fluctuation in first order WT and its first derivative. To choose α_1 effectively, select adequate successive measurements under steady state, perform the first order WT and compute the standard deviation of the WT modulus (σ_{WT}) Jiang *et al* (2003a). The selection of α_2 is made to differentiate zero-crossing points from true steady state values. When comparing the true steady state WT modulus to the zero crossing point, one can notice (Figure 8.5) that the steady state values are associated with a slow change in the WT modulus. Therefore, one can perform the first derivative on the WT and compute its standard deviation ($\sigma_{\frac{dWT}{dt}}$). Then the threshold values are selected as follows:

$$\alpha_1 = \sigma_{WT} \quad (10)$$

$$\alpha_2 = \sigma_{\frac{dWT}{dt}} \quad (11)$$

Selection of α_3 is based on the WT features to identify significant changes in the process trend. To choose α_3 effectively, multiple sets of different historic steady state process measurements are selected, and the first order WT and its standard deviation are computed. Starting with a λ value of 1, calculate the following threshold value and judge on the efficiency of the method on the selected data sets:

$$\alpha_3 = \lambda \sigma_{WT} \quad (12)$$

If the ending point of the steady state period is detected to late, then decrease the λ value (around 0.1) and repeat. If it is detected to early, increase λ value. That process is repeated until a suitable solution for all steady state set is achieved.

8.2.4.2.2 Selection of filtering parameters (β_1, β_2 and β_3)

Filtering parameters have a significant impact on the performance of the second step of the steady state detection. To avoid, as much as possible, undesired results due to unexpected on-line events, the investigation of the optimal filtering parameter must be performed simultaneously on different historic steady state periods. Optimal filtering parameters are obtained through the following steps (inspired by Bhat and Saraf 2004):

1. Select a value of β_3 (around 0.01)
2. Select a low value of β_2 (around 0.005)
3. Select a high value of β_1 (around 0.25)

4. Increment β_1 by .05 while the filtered average is affected by the high frequency component
5. Increment β_2 by .005 while every steady state period is over-estimate
6. Increment β_3 by 0.005 and go back to step 5 while all steady state period is detected.

8.2.5. Case study : Application to real time papermachine data

The efficiency of the method described above was investigated on real time plant data from a papermachine stock approach system in an integrated pulp and paper mill. The plant is equipped with a process data management system (PI system from OsiSoft inc.) that samples tags approximately every 10 seconds. The papermachine under study has multiple process operation regimes associated with different characteristics in the process measurements and high frequency features of the signal.

8.2.5.1 Performance indices

In order to investigate the efficiency of the proposed methodology, it as been compared to two other methods used to detect steady state properly. The comparison was done by calculating, for each method, the type I (steady state is not detected) and type II (false detection) errors related to steady state detection.

The accuracy and precision of the steady state detection technique for on-line steady state data reconciliation was also evaluated. Accuracy is calculated for each steady state period, as the relative error between the true process value and the reconciled measurement. Precision is defined as the standard deviation of such errors. Therefore, low values for those two indices indicate high accuracy and precision.

Hypothetical real time true process signals were systematically corrupted with various types of errors. Since the performance indices used to investigate the efficiency of the proposed methodology are based on the true process data value, it was impossible to use directly process data to asses this task. Therefore, based on a systematic study of the process operations, experimental data were created to investigate the methodology performance.

8.2.5.2 Efficiency of the online steady state detection

The proposed methodology for on-line steady state detection has been compared to the method based on first and second order WT and Cao's method. Optimal parameters were used for the WT based and Cao's method. WT based parameters were selected according to Jiang *et al*'s description (2003a), while optimum parameters for Cao's method were selected based on the method proposed by Bhat and Saraf (2004).

Figure 8.10 shows the quantity of type I error related to the identification of true steady state periods. One can notice that, for the system under study, the application of the proposed method reduces considerably the amount of type I error. That can be explained by the presence of a large amount of different high frequency features in the measurements:

1. Filtering the raw measurement variance by a low pass filter can lead to misidentification of process trend when high frequency features varies suddenly (figure 8.2).
2. Utilization of one the wavelet features in a conventional way, using a single cutting scale per variable, can lead to under-scaling or over-scaling problems when dealing with signal associated to multiple frequency features (figure 8.3).

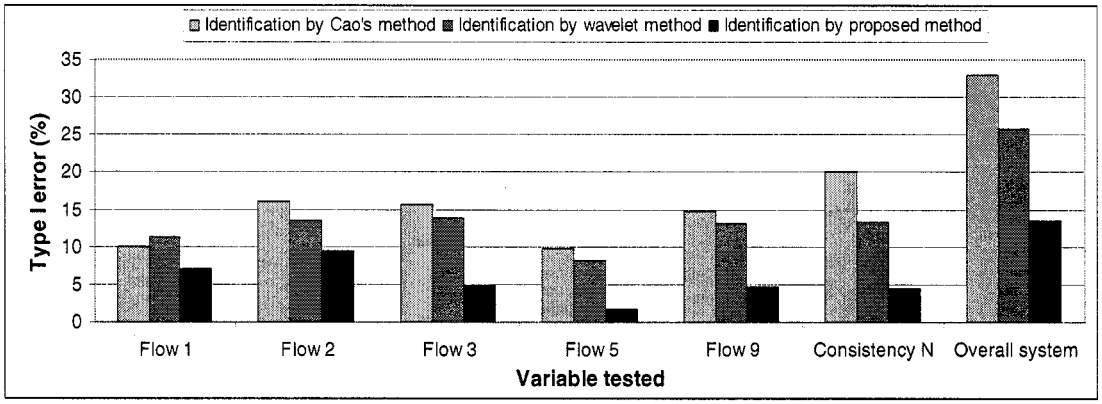


Figure 8. 10 : Type I error related to online steady state identification

In order to reconciled the data in steady state, one needs to identify previously the overall system steady state periods. Figure 11 shows the quantity of type II error related to the identification of true steady state periods. One can notice that the effect of the chosen steady state identification method does not have a high impact on the type II error related to the overall system status recognition. Since every variable under study must be under near steady state to record the overall system steadiness, type II errors related to the each component are diluted in the process of identifying overall system status.

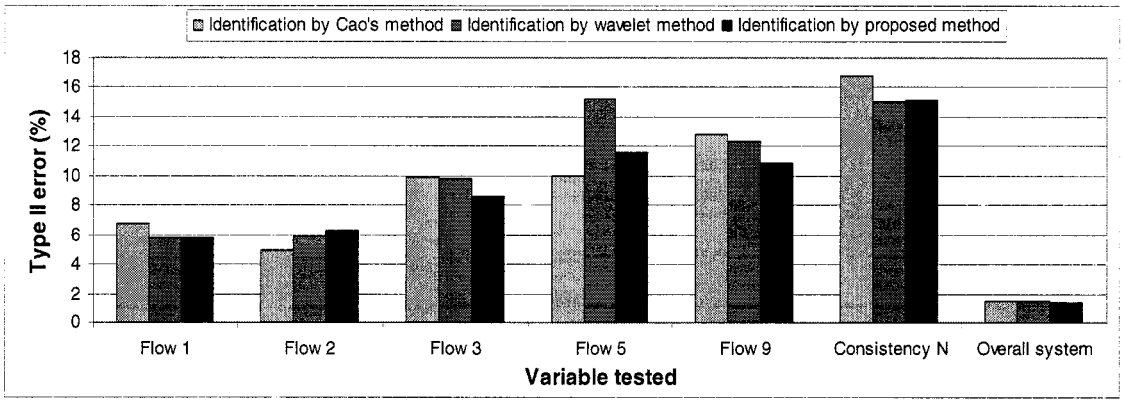


Figure 8. 11 : Type II error related to online steady state identification

8.2.5.3 Impact on online steady state data reconciliation

To analyze the impact of steady state detection on the on-line reconciliation results, the results of the steady state analysis by the 3 different methods were used to reconcile the data. Experimental data were reconciled by applying the exact same reconciliation method to each data set.

To reconciled on-line steady state data, the overall methodology proposed by Bellec *et al.* (2004) was used. First, to correct and filter abnormal measurements and high frequency features, data acquired by the data management system are analyzed based on wavelet features. Then, steady state detection is performed. That step insures to select adequate steady state data for steady state data reconciliation. Finally the data are reconciled by using Sigmafine software from OsiSoft Inc.

Figures 8.12 summarize the improvement of the average accuracy of the steady state data before and after the application of the different methodologies. Relative error is used as to represent accuracy in the data sets. Therefore low value on the figure indicates high accuracy. Equation 13 shows the average relative error has been calculated.

$$\%\mathrm{RE} = average\left(100\% * \frac{|\mathrm{Targeted.value} - \mathrm{measurement}|}{\mathrm{Targeted.value}}\right) \quad (13)$$

One can notice that the choice of steady state identification method as a direct impact on the quality of the steady state data reconciliation results. Generally, more adequate status identification leads to improvement in accuracy at the reconciliation step (see figure 8.1 for details). By analyzing Figure 8.12, one can notice that for some specific variables (cons N, level 8 and level 16) the relative error does not seem to be reduced by the utilization of the new method. Results on the graph for those variables are multiplied by

a factor of 100. Therefore, we are dealing with small errors in the range of 0.05%. These small errors are more difficult to correct by using data reconciliation techniques.

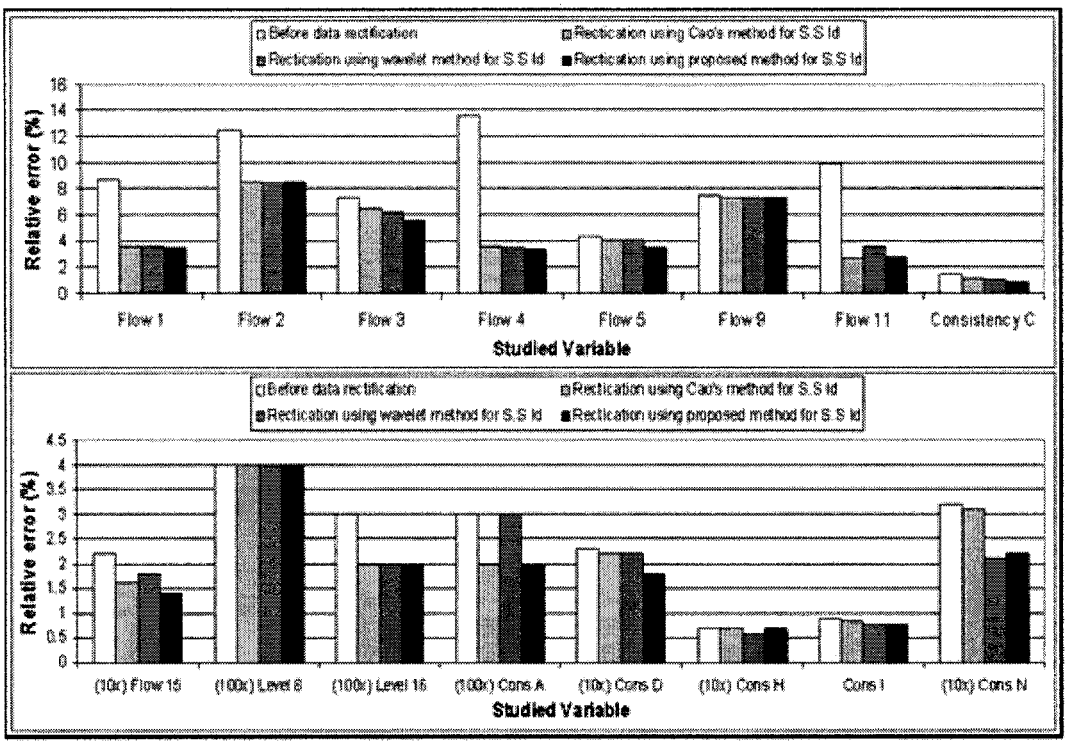


Figure 8. 12 : Accuracy improvement by the application of different online steady state detection techniques

Figure 8.13 summarize the improvement of the precision in the overall steady state data sets. Standard deviation to the average error is used to represent precision in the data sets. Therefore lower value on the figure indicates a higher precision. Equation 14 shows the standard deviation of relative error has been calculated. One can notice that the presence of gross errors (large amplitude errors on long period of time) will be reflected by a high standard deviation.

$$SD_{\%RE} = \sqrt{\frac{\sum_{i=1}^{n}\left[100\% * \left(average(RE) - RE_i\right)^2\right]}{n-1}} \quad (14)$$

By analyzing Figure 8.13, one can notice that the choice of the steady state identification method can have a huge impact on the precision of the reconciliation and therefore influence the correction of gross errors.

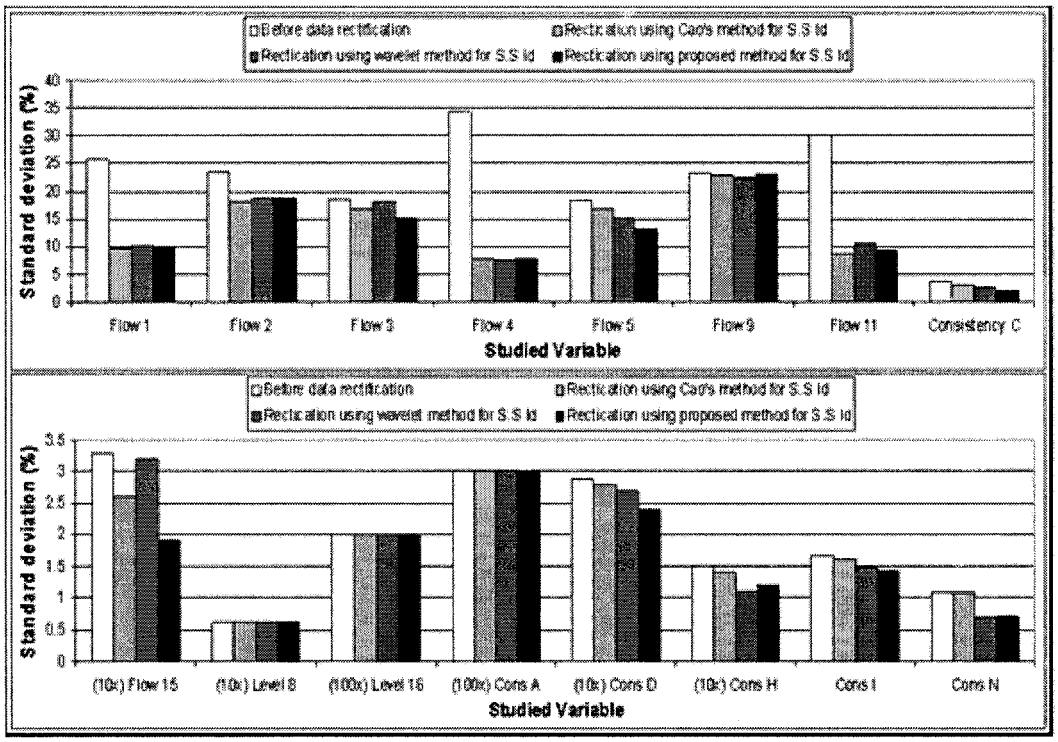


Figure 8. 13 : Precision improvement by the application of different online steady state detection techniques

8.2.6 Conclusions

A methodology to detect efficiently steady state status has been proposed and applied in this paper, which is suitable for on-line application. The method detects steady state periods using 3 signal processing steps, which allow reduction of Type I and Type II errors related to the identification of steady state periods. The method has been compared to other on-line steady state methods, using industrial data fro the stock approach system of a paper machine. Results show a higher accuracy improvement on

real time measurements when using the proposed methodology. Using this approach for steady state detection in a general methodology to reconcile data on-line could potentially lead to multiple benefits for on-line process operations such as the following: efficient on-line identification of out-of-calibration instruments, improved process control due to improvement of measurement accuracy; and improved operations planning and optimization.

Chapitre 9. Réconciliation des données

La réconciliation des données permet de rendre un ensemble de mesures, propres à un système, conforme à une loi quelconque. Dans ce cas-ci, la loi de la conservation de la matière est utilisée afin de réconcilier les mesures de débits et de composition en fibres des différents courants du procédé à l'étude.

L'étape de réconciliation est divisée en trois sous étapes (figure 9.1). Premièrement, les données mesurées sont classifiées. Cette étape permet d'identifier les mesures redondantes et les variables non mesurées qui peuvent être approximer à l'aide des informations connues. Par la suite, les erreurs brutes sont détectées et les mesures redondantes sont réconciliées. Dans le cas présent, ces deux dernières étapes sont effectuées de manière itérative.

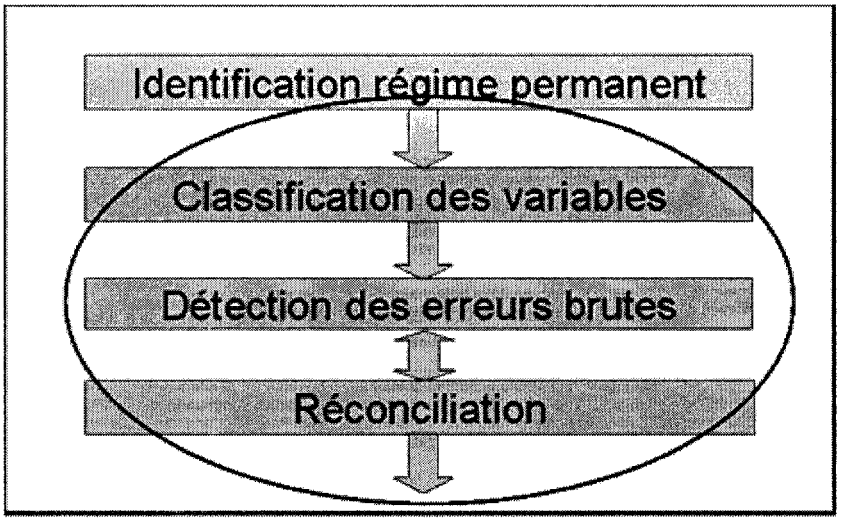


Figure 9. 1 : Étapes de la réconciliation des données

9.1 Classification des variables

La réconciliation des données est effectuée en faisant appel à une fonction objective qui minimise la somme des carrés des erreurs associées aux mesures effectuées. Tout cela en

tenant compte des contraintes relatives à la conservation massique dans le système étudié. Puisque la réconciliation des données est un problème d'optimisation, il est nécessaire que le problème soit surspécifié, donc que les mesures soit redondantes dans le système.

Une mesure est jugée redondante lorsqu'il est possible de calculer, d'au moins une manière, sa valeur à l'aide des autres données mesurées autour du système (Bagajewiczs 2000). Lorsque l'on fait référence à un système, on compare le nombre d'équations indépendantes relatives à un système et le nombre d'inconnu qui lui est associé. Si le nombre d'équation est plus grand que le nombre d'inconnu, alors le système est surspécifié.

Un premier inventaire des mesures effectuées en temps réels sur le procédé indique que les données mesurées n'y sont pas redondantes. Cependant, en utilisant les résultats des données échantillonnées en laboratoire autour du système, le système devient redondant.

Les données de laboratoire, permettant la redondance du système, ne sont pas échantillonnées à la même fréquence que les données prises en temps réels. Cette constatation pose un problème. En effet, puisque la valeur échantillonnée en laboratoire n'est pas prise à toutes les 10 secondes (comme celles en temps réel), le système n'est surspécifié qu'au moment de la prise de l'échantillon. Par contre, il est à noter que les échantillonnages effectués en laboratoires sont faits sur les mesures de consistance des différents courants et que celles-ci varient très peu dans le temps. Par conséquent, en supposant que la variable échantillonnée possède une valeur constante jusqu'à l'échantillonnage suivant, il est possible (et raisonnable) d'avoir constamment des données redondantes dans le système étudié.

9.2 Détection des erreurs brutes et réconciliation des données

Afin de détecter les erreurs brutes présentes dans le système et de les corriger via la réconciliation des données, une méthode itérative couplée au test sur les mesures est utilisée. Cette méthode est décrite dans la sous-section suivante. Les deux autres sous-sections sont dédiées à l'explication du test sur les mesures et à l'étape de résolution du problème de réconciliation à l'aide du logiciel Sigmafine.

9.2.1 Méthode itérative

La figure 9.2 représente la méthode itérative utilisée pour détecter et réconcilier les données en régime permanent. Premièrement les données sont réconciliées une première fois. Suite aux résultats de cette étape, les résultats de la réconciliation sont comparés aux valeurs mesurées à l'aide du test sur les mesures. Ainsi les mesures susceptibles d'être contaminées par une erreur brute sont identifiées. Celles possédant la plus grande correction est corrigée et la réconciliation est refaite une seconde fois. La seconde réconciliation permet de déterminer si la valeur corrigée est bel et bien celle affectée. Si non, une deuxième valeur susceptible d'être contaminée est corrigée. La boucle se poursuit ainsi jusqu'à ce que la réconciliation effectuée soit acceptable.

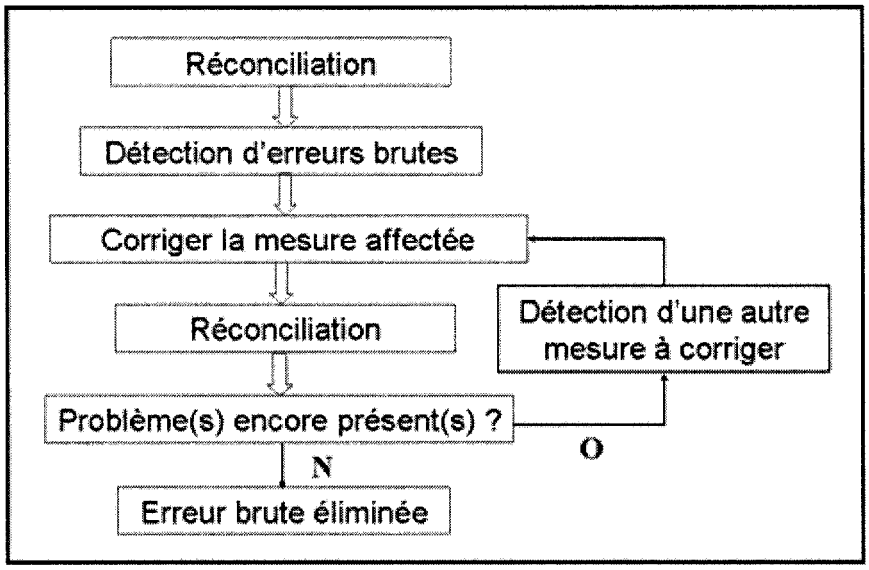


Figure 9. 2 : Méthode itérative pour détecter les erreurs brutes et les réconciliés

9.2.2 Détection des erreurs brutes

Afin de détecter les erreurs brutes présentes dans une série de données en régime permanent, le test sur les mesures est utilisé. Suite à la première réconciliation de la méthode itérative, la différence entre la donnée réconciliée et la donnée mesurée est calculée (équation 28) pour chacune des variables étudiées. Selon la théorie associée au test (voir section 3.3.2), η_i doit appartenir à une distribution normale. Afin de détecter les mesures susceptibles d'être contaminées par une erreur brute, il suffit d'identifier les valeurs de η_i qui ne font pas parti de la distribution normale. Il est à noter que pour effectuer cette identification, un test d'hypothèse avec un intervalle de confiance de 95% a été utilisé dans ce projet.

$$a = F - \tilde{F} \qquad (28)$$

$$\eta_i = \frac{a_i}{\sqrt{Q_{ii}}} \qquad (29)$$

9.2.3 Réconciliation à l'aide du logiciel Sigmafine

Les étapes de réconciliation des données sont effectuées à l'aide du logiciel Sigmafine de OsiSoft Inc. Ce logiciel permet une intégration directe avec les données emmagasinées en temps réel dans le système de gestion PI (OsiSoft Inc.).

9.2.3.1 Caractéristiques du logiciel

La technique de résolution du logiciel est basé sur le filtre de Kalman et ne requière aucune inversion matricielle. Il peu permettre de travailler sur de gros système sans problème de convergence. Ce qui le rend très pratique pour les applications industrielles.

Les fluctuations du niveau des réservoirs, présentes pendant les périodes de régimes permanents, ont été modélisées par un débit d'entrée/sortie des cuves (voir la section 3.3.3. pour plus de détails). Ce type de modélisation n'est pas tout à fait idéal lorsque l'on utilise le logiciel Sigmafine. En effet, puisque ce dernier ne permet pas d'avoir d'entrée de débit négatif, il est impossible de modéliser les fluctuations associées aux niveaux des réservoirs à l'aide d'un seul courant qui a une valeur négative lorsque le niveau augmente dans la cuve et une valeur négative lorsque celle-ci diminue.

9.2.3.2 Méthodologie utilisée

Pour palier au problème relié à l'utilisation impossible d'un débit négatif dans le logiciel Sigmafine, quatre modèles ont été créés. Puisque le système contient deux réservoirs dont la dynamique peut être modélisée par un courant entrant ou sortant de la cuve, ces quatre modèles inclus toutes les possibilités de combinaison possible reliée à la modélisation du système. Le choix du modèle à utiliser est donc basé sur le calcul de la variation du niveau des réservoirs.

La réconciliation a lieu en utilisant une fonction objective pondérée minimisant l'erreur entre les valeurs mesurées (équation 32) et celles nécessaire pour atteindre les contraintes relatives aux bilans massiques (équations 47 et 48). Lors de cette première étape, les poids attribués aux mesures dans la fonction objective sont inversement proportionnels à la variance des données associées à la variable.

$$\left.\begin{array}{l} Min\left[Z-\tilde{Z}\right]^T Q^{-1}\left[Z-\tilde{Z}\right] \\ s.t.: g(\tilde{Z})=0 \end{array}\right\} \quad (32)$$

où $g(\tilde{Z})$ inclus les contraintes reliées à l'opération des unités et aux instruments

$$\sum F_{in} = \sum F_{out} \quad (47)$$

$$\sum F_{in} Cons_{in} = \sum F_{out} Cons_{out} \quad (48)$$

9.3 Application de combinaisons de techniques à la méthodologie générale

Afin de choisir les techniques favorisants une application de la méthodologie la plus exacte et précise possible, les combinaisons logiques des différentes techniques comparées à chacune des étapes de la méthodologie générale (présentée au chapitre 6) ont été testées sur la série de données expérimentales créées.

9.3.1. Présentation des combinaisons étudiées

Dans la présente étude, deux techniques ont été étudiées pour la détection des anormalités en ligne : test de Hampel et détection par la méthode des ondelettes. 3 techniques de détections des régimes permanents ont également été comparées. Une possibilité de 6 combinaisons différentes de choix pour la détection des anormalités et l'identification des régimes permanents s'offrent donc naturellement (2 tests de détection des anormalités multipliés par 3 test d'identification des régimes permanents). Cependant, l'utilisation de la technique de détection des anormalités proposées par Cao

et Rhinehart (1995) n'est habituellement pas précédée d'un prétraitement des données. L'application de cette technique sera également étudiée. Il y aura donc, au total, 7 combinaisons de techniques étudiées dans cette étude.

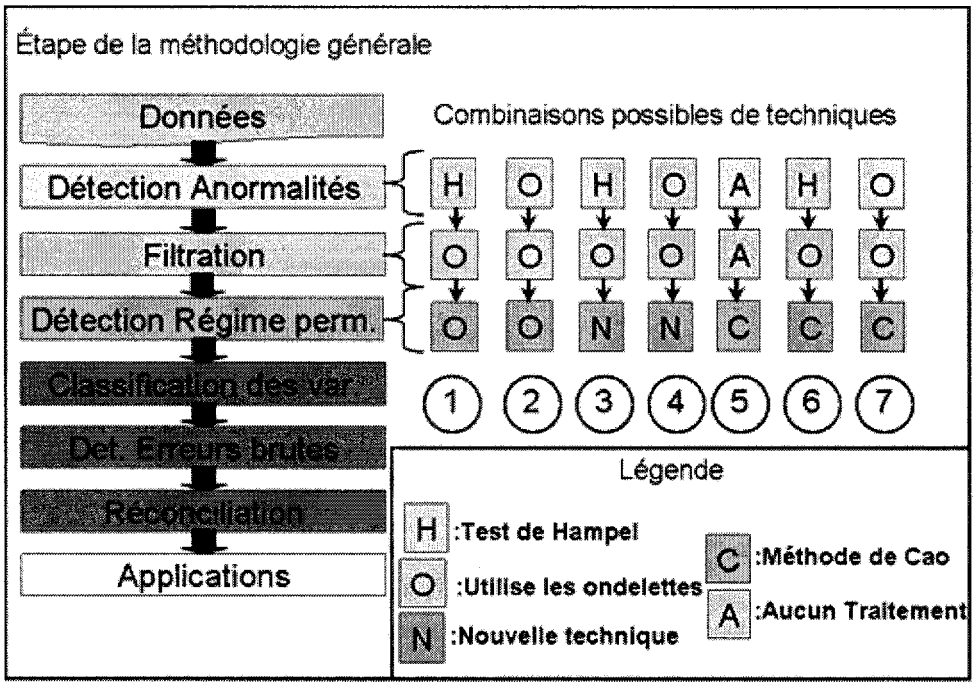


Figure 9. 3 : Représentation des sept combinaison de techniques étudiées

9.3.2 Résultats et discussion

Les 7 combinaisons de techniques pouvant être utilisés dans la méthodologie globale ont été testé à l'aide des données expérimentales crées. La précision et l'exactitude des résultats obtenus ici ont été calculées de la même façon que celles obtenues dans le chapitre 6. Il est à noter que tous les résultats intermédiaires permettant d'obtenir les résultats présentés dans cette section sont disponibles en annexe :

- Données expérimentales (Annexe II)
- Analyse se sensibilité
  - Associée aux paramètres permettant la détection des anormalités par la méthode des ondelettes (Annexe III)
  - Associée aux choix du niveau de coupure de l'étape de filtration (Annexe IV)
  - Associée à l'identification du régime permanent (Annexe V)
- Résultats intermédiaires des calculs d'exactitude et de précision (Annexe VI)
- Résultats finaux des calculs d'exactitude et de précision pour chacun des courants (Annexe VI)

9.3.2.1 Exactitude des différentes combinaisons de techniques candidates

La figure 9.4 montre un exemple des résultats reliés à l'exactitude des données réconciliées obtenus en utilisant les différentes combinaisons de techniques possibles dans la méthodologie proposée. On peut remarquer que la $4^{ème}$ combinaison, associée à l'exploitation des ondelettes pour détecter les anormalités et filtrer les données, suivi de l'utilisation de la technique d'identification des régimes permanents développée dans ce rapport, est favorisée par ces résultats.

Il est à noter que plus la moyenne des erreurs relatives est élevée, moins le résultat de la réconciliation effectuée est exacte. Les résultats relatifs aux autres variables mesurés sont disponibles en à l'annexe VI.

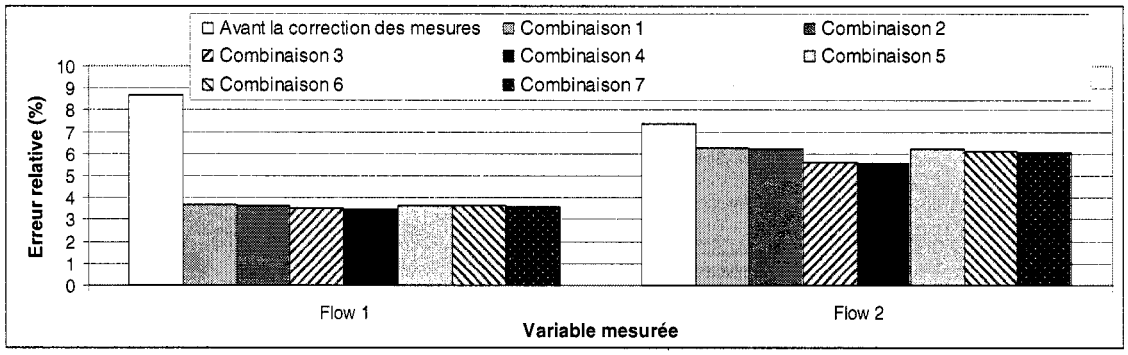


Figure 9. 4 : Exactitude associée à l'utilisation des différentes combinaison de techniques étudiées

9.3.2.2 Précision des différentes combinaisons de techniques candidates

La figure 9.5 montre un exemple des résultats reliés à la précision des données réconciliées obtenus en utilisant les différentes combinaisons de techniques possibles dans la méthodologie proposée. Encore une fois, la 4ème combinaison de techniques, associée à l'exploitation des ondelettes pour détecter les anormalités et filtrer les données, suivi de l'utilisation de la technique d'identification des régimes permanents développée dans ce rapport, est favorisée par ces résultats.

Il est à noter que plus l'écart-type des erreurs relatives est élevée, moins la réconciliation effectuée est précise. Les résultats relatifs aux autres variables mesurées sont disponibles à l'annexe VI.

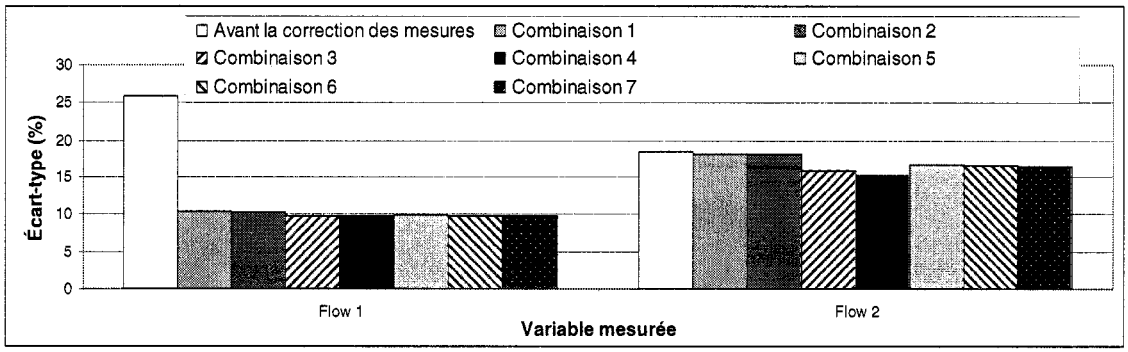


Figure 9. 5: Précision associée à l'utilisation des différentes combinaison de techniques étudiées

Chapitre 10. Interprétation, analyse et discussion générale

Tout au long de cette étude, une méthodologie générale permettant de traiter, en ligne, les données en régime permanent d'un procédé a été analysée (figure 6.1). Afin d'identifier les différentes techniques permettant d'optimiser l'amélioration de l'exactitude et de la précision dans les données mesurées, une étude systématique a été effectuée.

Suite à une revue de littérature exhaustive permettant d'identifier les différentes techniques pouvant être implantées en ligne, l'utilisation des différentes techniques permettant de détecter les mesures anormales a premièrement été comparée. Par la suite, l'application d'une nouvelle technique permettant d'identifier, en ligne, les régimes permanents dans une série de mesures a été confrontée aux techniques existantes dans la littérature. Ensuite, l'effet des différentes combinaisons possibles de techniques sur la précision et l'exactitude des résultats obtenus, en implantant la méthode proposée, a été comparé. Finalement, la combinaison optimale des techniques identifiées a été appliquée à une série de données en temps réels.

10.1 Discussion reliée à la détection des anormalités

Afin d'identifier la technique la plus appropriée pour détecter, en ligne, les mesures anormales, le test de Hampel et la méthode basée sur l'utilisation des caractéristiques propres aux ondelettes ont été appliqués sur 5 jours de données expérimentales. Afin de comparer les méthodes et d'identifier leurs forces et leurs faiblesses, les anormalités traitées ont été séparées en 4 catégories. Cette séparation est basée sur une caractérisation de l'amplitude et de la durée des anormalités étudiées (figure 7.2).

Les résultats présentés à la figure 7.3 démontrent que le test de Hampel est moins efficace que la méthode basée sur les ondelettes pour identifier les anormalités de longue durée. Cette constatation peut être expliquée par un point de brisure, associé à l'utilisation de la médiane dans le test de Hampel, d'une valeur de 50% (les détails sont présentés à la section 7.3.1). En ce qui concerne les autres catégories d'anormalités, la différence entre les résultats obtenus par les deux techniques est très faible.

En plus de l'efficacité des deux méthodes à détecter les anormalités, le taux d'erreurs de type II associées à l'utilisation de chacune des techniques a été comparé (figure 7.5). Dans les deux cas, le taux d'erreur est très faible.

Suite à ces constatations, il est possible d'affirmer que la méthode basée sur la théorie des ondelettes est plus performante que le test de Hampel pour détecter les anormalités présentes dans une série de mesures. En effet, la méthode basée sur les ondelettes détecte plus efficacement les mesures anormales, qui persistent sur une longue période, que le test de Hampel. Pour les autres types d'anormalités les deux techniques ont une efficacité et un taux d'erreur de type II qui sont semblables.

10.2 Discussion reliée à l'identification des régimes permanents

Afin de sélectionner la méthode permettant l'identification la plus efficace des différentes périodes en régime permanent présentes dans une série de mesures, une nouvelle technique élaborée dans la cadre de ces travaux (section 8.2.4) a été comparée aux techniques présentes dans la littérature (méthode de Cao et méthode basée sur l'utilisation des transformées d'ondelettes).

Premièrement, l'efficacité des différentes méthodes à identifier adéquatement les périodes de régimes permanents a été identifiée. Pour se faire, les taux d'erreurs de types I et II attribuables à la détection de chaque point mesuré en période de régime permanent

ont été calculés (figure 8.10 et 8.11). Les erreurs de type I sont systématiquement plus faibles lorsque l'on emploie la nouvelle méthode proposée. En ce qui concerne les erreurs de type II, celles-ci sont sensiblement du même ordre de grandeur dans les 3 cas. Plus spécifiquement, les erreurs de type II associées à l'identification du régime permanent du système en entier sont moins importantes dans le cas de l'utilisation de la nouvelle technique. Puisque c'est à partir de ce critère que les données sont sélectionnées afin d'être réconciliée, on peut affirmer que les erreurs de type II seront minimisées en utilisant la technique proposée. Ces constatations sur l'efficacité de la nouvelle technique à minimiser les erreurs, associées à l'identification des régimes permanents, peuvent être expliquées par la théorie associée à son élaboration. En effet, la nouvelle technique élaborée minimise, dans un premier temps, les erreurs de type I en surestimant la durée du régime permanent. Par la suite, en retournant dans le temps, les erreurs de type II sont éliminées. Cela permet de détecter le moment exact où un changement se manifeste dans la tendance du procédé.

Une fois l'étude sur l'efficacité de détection des techniques effectuée, l'influence du choix de la méthode sur l'exactitude et la précision des résultats obtenus a été étudiée (figure 8.11 et 8.12). Ces résultats démontrent que la nouvelle technique permet d'améliorer systématiquement l'exactitude des résultats de manière au moins aussi efficace que les autres techniques étudiées. Le choix de l'une ou l'autre des méthodes ne semble pas avoir un grand impact sur la précision des résultats obtenus. Cette dernière constatation peut facilement être expliquée en étudiant de plus près l'équation utilisée pour calculer la précision des résultats :

$$SD_{\%RE} = \sqrt{\frac{\sum_{i=1}^{n}\left[100\% * \left(average(RE) - RE_i\right)^2\right]}{n-1}} \quad (49)$$

Le résultat de cette équation est grandement influencé par les erreurs de grandes dimensions, donc par les erreurs brutes. Puisque la détection de ces erreurs n'est pas

nécessairement influencée par l'efficacité de l'identification du régime permanent, il est normal que les résultats reliés à la précision diffèrent peu d'une méthode à l'autre.

À l'intérieur de la revue de la littérature, il est spécifié que les erreurs brutes englobes plusieurs types d'erreurs (figure 2.4), qu'une détection efficace de ces inconsistances comporte quatre grande étape (Bagajewicz 2003) et qu'aucun test sur la détection des erreurs brutes ne permet de toutes les identifiées (Crowe 1996). En observant la figure 2.4, on peut comprendre que certains types d'erreurs brutes ne peuvent pas être identifiés en régime permanent. Dans le cadre de cette étude, la détection de ce type d'erreur est donc loin d'être parfaite. En effet, en plus de savoir qu'aucune technique existante ne permet l'identification de toutes ces erreurs, qu'une bonne détection est limitée par plusieurs étapes, on sait que le fait d'effectuer la réconciliation en régime permanent diminue les chances d'identifier ces erreurs.

Les résultats qui ont été discutés dans cette section démontrent que les erreurs de type I et de type II associées à l'identification des régimes permanents, propres à un système, sont réduites en utilisant la nouvelle méthode. De plus, cette méthode permet d'améliorer, de manière plus prononcée, l'exactitude des résultats obtenus suite à la réconciliation des données sans altérer la précision des mesures.

10.3 Discussion associée à la détermination de la combinaison optimale de techniques

Une fois les études sur la détection des anormalités et sur l'identification du régime permanent terminées, une étude comparative des différentes combinaisons de techniques applicables à chacune des étapes de la méthodologie proposée a été effectuée. L'effet provoqué par l'utilisation des 7 combinaisons logiques de méthodes a été étudié (figure 9.4 et 9.5). Ces résultats démontrent que la $4^{ème}$ combinaison de techniques permet

d'obtenir systématiquement les résultats les plus exacts. Cependant, on remarque que le choix des techniques utilisées ne semble pas influencer la précision des résultats.

La 4ème combinaison de techniques implique l'utilisation des méthodes basées sur la théorie des ondelettes aux étapes de détections des anormalités et de filtration. Par la suite les données en régime permanent sont identifiées à l'aide de la nouvelle méthode et finalement réconciliées à l'aide du logiciel Sigmafine de OsiSoft Inc. Le fait que l'amélioration optimale de l'exactitude des résultats soit obtenue à l'aide de cette combinaison peut être expliqué par les résultats présentés précédemment (détection des anormalité et identification des régimes permanents) et par les informations contenues dans la revue de la littérature. En effet, les résultats des sections précédentes démontrent que l'utilisation des ondelettes pour détecter les anormalités est plus efficace que le test de Hampel et que l'utilisation de la nouvelle technique pour identifier les périodes de régime permanent permet une plus grande amélioration de l'exactitude des résultats que l'application des deux autres méthodes. De plus, il a été démontré, dans la revue de la littérature, que l'utilisation d'ondelettes est plus efficace que celle des filtres passse-bas pour traiter les erreurs aléatoires.

Le fait qu'aucune combinaison de technique ne semble favoriser la diminution d'erreur de type II peut encore être expliqué par le calcul de la précision. Celui-ci est plus influencé par la présence d'erreurs brutes que par tout autre type d'erreurs (voir 10.3 pour l'explication complète).

10.4 Discussion reliée à l'application de la méthode optimale à des données en temps réels

Une fois la combinaison optimale de technique identifiée, elle a été appliquée, à l'aide de la méthodologie proposée, sur les données en temps réels d'une machine à papier. L'application a été effectuée sur 14 variables mesurées à toutes les 10 secondes et 4

variables dont les connaissances sont issues des données de laboratoire couplées à un jugement d'ingénierie. La méthodologie appliquée a permis d'améliorer l'exactitude et la précision des données mesurées ainsi que d'identifier et de corriger des erreurs de décalages et de défaillances totales (figure 6.11).

L'application de cette méthodologie à non seulement permis de corriger des données défaillantes, mais son application n'a pas engendré de défaillance supplémentaire. L'équation mathématique, issue de la littérature, permettant de définir de la robustesse est inversement proportionnel à :

$$Def = 100\% * \frac{n_{Def}}{n_{Mesure}} \quad (4)$$

Puisque le nombre de défaillance diminue avec l'application de la technique et que le nombre de mesures ne varie pas, on peut affirmer que la méthode permet d'augmenter la robustesse des mesures.

Chapitre 11. Conclusions

Tout au long de cette étude, une méthodologie générale permettant de traiter, en ligne, les données en régime permanent d'un procédé a été analysée. Cette étude avait pour but de démontrer la robustesse et la capacité d'améliorer l'exactitude et la précision des données traitées à l'aide de la méthodologie. Cet objectif général a été approché à l'aide de sous objectifs spécifiques tels que la comparaison de l'efficacité des techniques de détection des anormalités en ligne, la comparaison des performances d'une nouvelle méthode de détection du régime permanent en 3 étapes avec les méthodes déjà existantes pour traiter les problèmes en ligne et finalement la démonstration des capacités d'implantation de la méthode en milieu industriel.

La méthodologie étudiée permet de traiter tous les types d'erreurs de manière systématique. Pour se faire, une étape de détection et de correction des anormalités est premièrement effectuée. Par la suite, les données sont filtrées pour éliminer les erreurs aléatoires. Ensuite les différents régimes permanents sont identifiés et les données sélectionnées sont réconciliées. En plus de ces étapes, une identification du régime d'opération permet de sélectionner différents paramètres de réglage et de les utiliser aux différentes étapes décrites précédemment.

Pour chacune des étapes permettant de corriger les erreurs présentes dans les mesures, une étude systématique, basée sur une revue de la littérature et sur une comparaison entre l'efficacité des différentes techniques applicables, a été effectuée. Pour la détection des anormalités, le test de Hampel a été comparé à une méthode basée sur l'utilisation des caractéristiques propres aux ondelettes. Suite aux informations et aux exemples d'applications recueillis permettant de confirmer l'inefficacité des filtres passe-bas à résister aux phénomènes de distorsion et de délais, le choix de la méthode de filtration s'est arrêté sur l'utilisation des ondelettes. Par la suite, une nouvelle méthode permettant d'identifier les régimes permanents en 3 étapes a été comparée aux techniques utilisées pour détecter ces périodes en ligne. Une fois les études sur la détection des anormalités

et sur l'identification du régime permanent terminées, une étude comparative des différentes combinaisons de techniques applicables à chacune des étapes de la méthodologie proposée a été effectuée. Finalement la combinaison optimale de techniques, permettant l'amélioration la plus importante de l'exactitude et de la précision des données, a été appliquée, en ligne, à des données disponibles en temps réels.

L'efficacité de la détection des anormalités à l'aide du test de Hampel a été comparée à l'utilisation d'une méthode basée sur l'exploitation des caractéristiques propres aux ondelettes. Les résultats de cette étude démontrent que la méthode basée sur les ondelettes est plus efficace que le test de Hampel pour identifier les anormalités présentes sur de longues périodes. En ce qui concerne les autres anormalités, les efficacités des deux techniques sont très similaires. Par conséquent, on peut conclure que l'utilisation des ondelettes pour détecter les mesures anormales est plus efficace que le test de Hampel lorsque l'on étudie l'ensemble des anormalités pouvant être présentes dans une série de mesures.

Une nouvelle méthode d'identification du régime permanent en 3 étapes : détection du point de départ, filtration des données en régime permanent et détection du point final a été comparée à deux autres techniques permettant d'identifier les régimes permanents d'un système en ligne. Les résultats de cette analyse ont démontrés que les erreurs de type I et II associées à la détection des régimes permanents du système sont réduites en utilisant la méthode développée. De plus, l'exactitude des résultats obtenus après la réconciliation des données est grandement améliorée avec l'implantation de cette méthode sans négliger la précision des résultats obtenus. Par conséquent, on peut conclure que cette méthode permet d'identifier plus efficacement, en ligne, les régimes permanents que les techniques existantes.

Une comparaison systématique des combinaisons de techniques applicables à la nouvelle méthodologie a été effectuée. Les résultats de cette étude démontrent que la combinaison, basée sur la détection et la correction des anormalités ainsi que la filtration des données à l'aide des ondelettes, suivi d'une identification du régime permanent en 3 étapes et d'une réconciliation des données à l'aide du logiciel Sigmafine, permet la plus grande amélioration de l'exactitude des résultats tout en améliorant leurs précisions aussi efficacement que les autres combinaisons étudiées. Cette combinaison de techniques a donc été incluse à la méthodologie générale et a été implantée, en ligne, afin de traiter les données en temps réel d'une machine à papier. L'application de la technique a permis d'améliorer l'efficacité des données mesurées ainsi que d'identifier et de corriger des erreurs de décalage et de défaillances totales des instruments. Ces résultats ont ainsi permis de démontrer la robustesse de la méthode et sa possibilité d'implantation en industrie.

La méthodologie développée au cours de cette étude démontre, de manière robuste, l'amélioration systématique de l'exactitude et de la précision des résultats. Par contre, ces deux derniers points pourraient être améliorés davantage si les erreurs brutes présentes dans les mesures effectuées pouvaient être détectées plus efficacement en ligne. Ce défi est occasionné par un manque dans le corps de la connaissance scientifique et par le choix d'effectuer la réconciliation des données en régime permanent. En effet, selon Crowe(1996), aucune technique ne permet de détecter toutes les erreurs brutes et certains types de ces erreurs ne peuvent difficilement être détectées en régime permanent. Il serait donc intéressant dans les travaux futurs d'étudier de plus prêt l'influence de la réconciliation en dynamique sur le détection de ces types d'erreurs et de développer une technique de détection des erreurs brutes qui soit efficace pour les travaux en ligne.

En implantant cette méthodologie de traitement des données en ligne, les usines pourront développer des reproductions fidèles des régimes d'opérations quasi permanents de leur usine. Cette implantation peut mener à plusieurs bénéfices tel que :

- Une détection efficace des instruments hors de calibrations
- Une amélioration du contrôle de procédé
- Une connaissance plus détaillée des différents régimes d'opérations du procédé
- L'identification des tendances locales et globales du procédé étudié
- L'identification de fuite et d'accumulation dans le procédé.
- Une amélioration des opérations de planification et d'optimisation reliées au procédé étudié.

Références

BAGAJEWICZ, M. 1995. «*On the probability distribution and reconciliation of process plant data* », *Computers and Chemical Eng.* Vol 20 no6-7 pp 813-819

BAGAJEWICZ, M. 1999. «A brief review of recent developpements in data reconciliation and gross error detection/estimation », *Empromer 1999,* pp. 335-342.

BAGAJEWICZ, M. 2000. «Comparison of steady state and integral dynamic data reconciliation, », *Computers and Chemical Eng.* pp 2367-2383

BAGAJEWICZ, M. 2001 *Process plant instrumentation and upgrade*, Technomic Publishing company, Lancaster, Pensylvania, 333p

BAGAJEWICZ, M. 2003. «Data Reconciliation and instrument upgrade. Overview and Challenges », *FOCAPO Conference*, p103-115

BAKSHI, B.R.and STEPHANOPOULOS, G. 1995. «Representation of process trends III. Multiscal exraction of trends from process data». *Computers and chemicals eng.* Vol 18 No 4 pp267-302

BAZIN, C., HODOUIN, D., DUCHESNE, C., THIBAULT, J., TRUSIAK, A.R. 1998. «Reconciliation of mass and energy data measurements : application to a rotary dryer ». *Canadian Metallurgical Quaterly* Vol 37 No 3-4 pp333-342

BHAT, A. and SARAF, N. 2004. «Steady-State Identification, Gross Error Detection and data reconciliation for industrial process units». *Ind. Eng. Chem.* Vol 43 pp 4323-4336

BODINGTON, C.E., 1995. «Data Reconciliation», *Planning, scheduling and control integration in the process industries.* Chapitre 4.

CAO, S. and RHINEHART, R.R. 1995. «*An efficient method for on-line identification of steady state*», *Journal of process control* Vol 5 No6 pp363-374

CHEN, J. and ROMAGNOLI, J.A. 1997. «*A strategy for simultaneous dynamic data reconciliation and outlier detection*», *Computers and Chemical Eng.* Vol 22 No4-5 pp569-562

CROWE, M.C. 1996. «*Data Reconciliation- progress and challenges*», *Journal of process control* Vol 6 No 2-3 pp89-98

DABROS, M., PERRIER, M., FORBES, F., FAIRBANK, M., STUART, P.R., 2005. «Model-based direct search optimization of the broke recirculation system in a newsprint mill», accepted to *Journal of Cleaner Production.*

DAUBECHIES, I. 1990. «The wavelet transform time frequency localization and signal analysis». *IEE transactions on information theory* , vol 36, pp 961-1005

DAVIES, L. and GATHER, U. 1993. «The identification of multiple outliers», *Journal of the American statistical ass.*, vol 88, pp 782-792

FLEHMIG, F., WATZDORF, R.V., MARQUARDT, W. 1998. «Identification of trends in process measurements using the wavelet transform»., *Computers and Chemicals Engineering*, Vol.22, pp 491-496.

HAMPEL, F.R. 1971. «A general qualitative definition of robustness», *Annals of Mathematics Statistics* , vol 42, pp 1887-1896

HAMPEL, F.R. 1974. «The influence curve and its role in robust estimation», *Journal of the American statistical ass.*, vol 42, pp 382-393

HARISSON, R., LEROUX, R., STUART, P.R. 2003. «Multivariate analysis of refiner operating data from a thermo-mechanical pulp newsprint mill», *Pulp & Paper Canada*, 105:4, pp. 24-27.

HODOUIN, D. and MAKNI, S. 1996. «Real-time data reconciliation of mineral processing plant using bilinear material balance equations coupled to empirical dynamic models», *International Journal of mineral processing.*, vol 48, pp 245-264

JACOB, J. and PARIS, J. 2003. «Data sampling and reconciliation – application to pulp and paper mills. Part I: methodology and implementation», *Appita J.*, 56(1), 25-52.

JACOB, J. and PARIS, J. 2003. «Data sampling and reconciliation – application to pulp and paper mills. Part II: case studies», *Appita J.*, 56(2), 116-120.

JANSSEN, M., LAFLAMME-MAYER, M., ZEINOU, M.-H., STUART, P.R. 2004a. «Survey indicates mill needs to exploit IT systems with new business model», *Pulp and Paper*, 105:6, pp. 46-51.

JANSSEN, M., STUART, P.R., 2004b. «Business model framework for using real-time process data for design decision making». Proceedings, C-DEN Design Conference, Montreal, QC. Comm. 07B5.

JIANG, T., BINGZHEN, C., XIAORONG, H. 2000. «Industrial application of wavelet transform to the on-line prediction of side draw qualities of crude unit». *Computers and chemicals eng.* Vol 24 pp507-512

JIANG, T., BINGZHEN, C., XIAORONG, H., STUART, P. 2003a. «Application of steady state detection method on wavelet transform». *Computers and chemicals eng.* Vol 27 pp569-578

JIANG, T., CHEN, B., JASMIN, K. STUART, P.R. 2003b. «Strategy for improving data quality for a kraft pulp mill recausticizing plant», *FOCAPO Conference*, p467-470.

KALMAN, R.E. 1960. «New approach to linear filtering and prediction problems», *Basic Eng., ASME,* 82D, 35

MALLAT, S. and HWANG, W.L. 1992. «Singularity detection and processing with wavelets», *IEE transactions on information theory* , vol 38, pp 617-641

MALLAT, S. and ZHONG, W.L. 1992. «Caracterization of singnal from multiscale edges», *IEE transactions on pattern analysis and machine intellignece*, vol 14, pp 710-732

MITRA, S. and ACHARYA, T. 2003. *Data mining*, John Wiley and sons, 401p

NOUNOU, M. and BAKSHI, B.R. 1999. «Online multiscale filtering of random and gross errors without process models», *AICHE Journal* Vol 45 no5 pp1041-1058

NARASIMAHAN, S. MAH, R.S.H 1986. «A composite statistical test for detecting changes in steady state», *American Institute of chemical eng. Journal* , vol 32, pp 1409-1418

NARASIMAHAN, S., CHEN, S., MAH, R.S.H 1987. «Detecting changes of steady state using the mathematical theory of evidnece», *American Institute of chemical eng. Journal*, vol 33, pp 1930-1932

NARASIMAHAN and JORDACHE 2000. *Data Reconciliation and Gross Error Detection: an Intelligent use of Process Data*, Gulf Publishing Company, Houston, Texas

TURNER, O. and MCELRATH, H. 1996. *Engeneering statistics the industrial experience,* Wadsworth publishing company, 569p

ANNEXE I

Section I-A : Liste des régimes permanents identifiés à l'aide des données historiques

Tableau A.I.1 : Liste des régimes permanents associés aux données historiques

Status	Regimes	Flow 1 L/min	Flow 2 L/min	Flow 3 L/min	Flow 5 L/min	Flow 9 L/min	Level 17 %	Cons P %
Grade A	1	430.0	120.0	3425.0	4300.0	8700.0	100.00	3.87
	2	430.0	320.0	3675.0	4300.0	8500.0	100.00	3.80
	3	430.0	320.0	3075.0	4300.0	7525.0	100.00	3.77
	4	830.0	320.0	3075.0	4300.0	8500.0	100.00	3.85
	5	830.0	320.0	3075.0	4465.0	8700.0	100.00	3.80
	6	1260.0	320.0	3075.0	4465.0	7525.0	100.00	3.77
	7	1260.0	320.0	3075.0	4625.0	7525.0	100.00	3.85
	8	1260.0	320.0	3075.0	4100.0	8500.0	100.00	3.75
	9	625.0	230.0	3425.0	4465.0	9650.0	99.50	3.85
	10	625.0	230.0	3425.0	4300.0	11000.0	99.50	3.80
	11	2500.0	280.0	3075.0	4740.0	7525.0	99.00	3.75
	12	2500.0	280.0	3075.0	4625.0	7525.0	99.00	3.80
Grade B	13	715.0	0.0	2350.0	4465.0	9650.0	102.50	3.82
	14	715.0	0.0	2350.0	4100.0	9650.0	102.50	3.87
	15	1475.0	0.0	700.0	4465.0	9650.0	102.50	3.82
	16	2500.0	0.0	700.0	4300.0	11000.0	102.50	3.75
	17	430.0	0.0	2350.0	4300.0	11000.0	100.00	3.75
	18	430.0	0.0	2350.0	4300.0	12000.0	100.00	3.75
	19	430.0	0.0	3675.0	4465.0	9650.0	100.00	3.77
	20	430.0	0.0	3675.0	4100.0	9650.0	100.00	3.82
	21	715.0	0.0	3675.0	4465.0	9970.0	100.00	3.77
	22	715.0	0.0	2350.0	4100.0	11000.0	99.70	3.75
	23	1475.0	0.0	2350.0	4100.0	12000.0	99.70	3.75
	24	1475.0	0.0	2350.0	4465.0	9650.0	99.70	3.77
	25	14000.0	0.0	0.0	4740.0	9970.0	99.70	3.75
Maintenance	26	0.0	0.0	0.0	0.0	0.0	100.00	3.75
Break	27	0.0	0.0	0.0	800.0	0.0	100.00	3.75

Section I-B : Tableau identifiant les bases de comparaison des différents régime

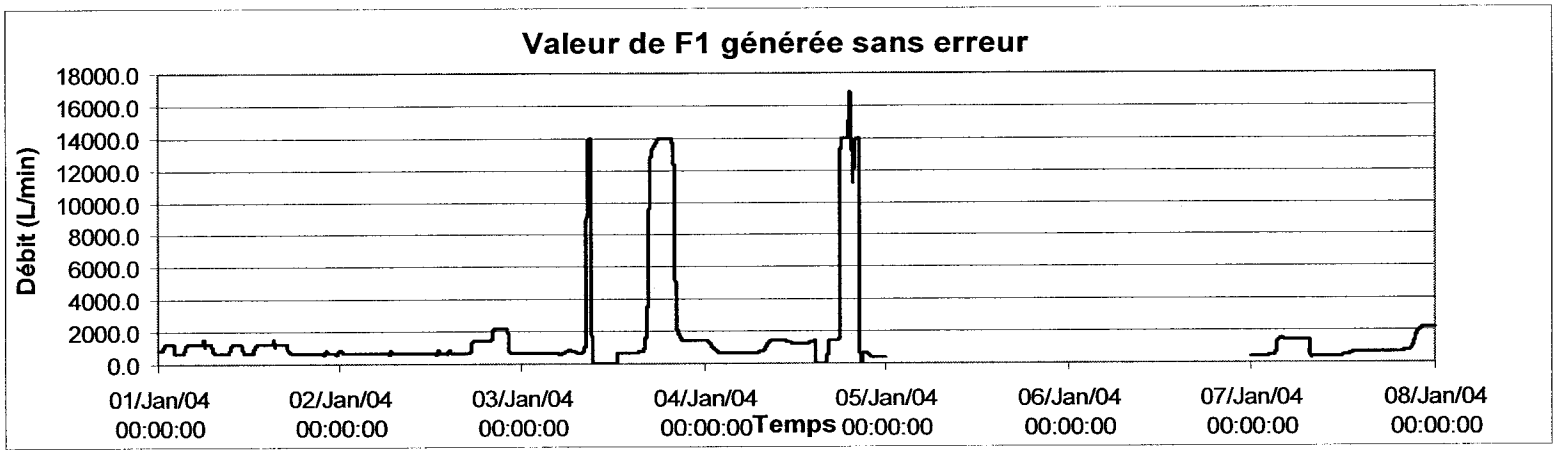
Ce tableau identifie, pour chacun des couples de régimes, la variable mesurée sur laquelle une analyse statistique permet de prouver que les régimes sont différents.

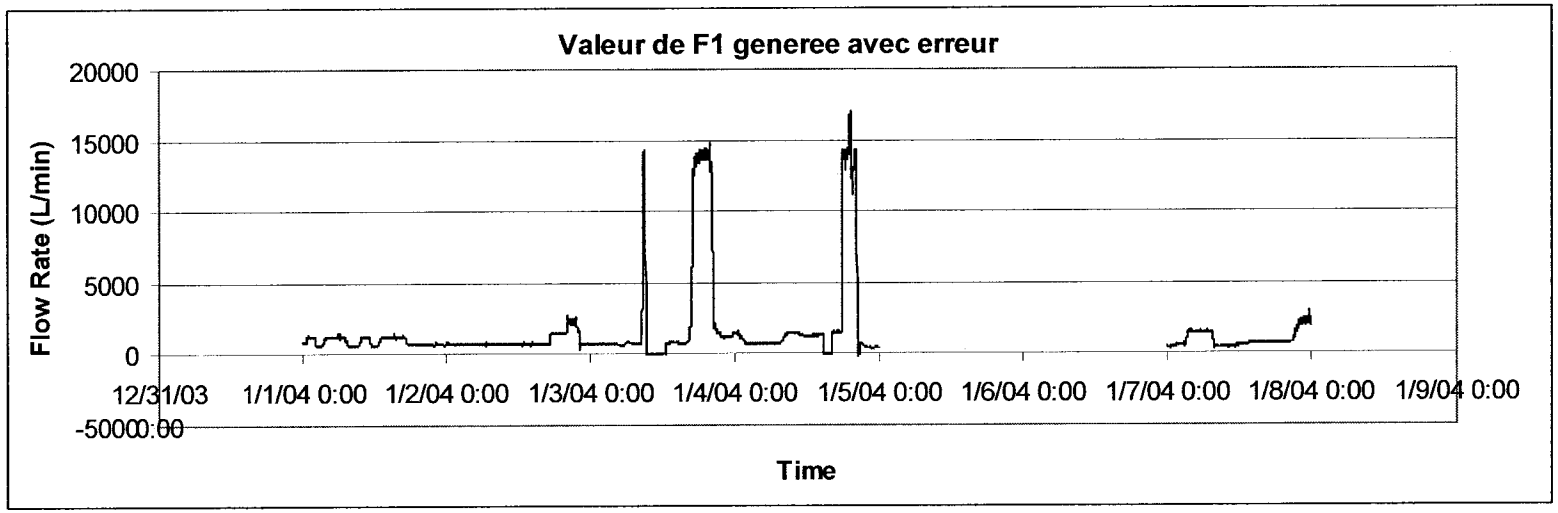
Tableau A.I.2 : Variable comparée pour l'analyse de différenciation des régimes d'opérations

	Régime 1	Régime 2	Régime 3	Régime 4	Régime 5	Régime 6	Régime 7	Régime 8	Régime 9	Régime 10	Régime 11	Régime 12	Régime 13	Régime 14	Régime 15	Régime 16	Régime 17	Régime 18	Régime 19	Régime 20	Régime 21	Régime 22	Régime 23	Régime 24	Régime 25	Régime 26	Régime 27
Régime 1																											
Régime 2	F2																										
Régime 3	F2	F3																									
Régime 4	F1	F1	F1																								
Régime 5	F1	F1	F1	Cons P																							
Régime 6	F1	F1	F1	F1	F1																						
Régime 7	F1	F1	F1	F1	F1	Cons P																					
Régime 8	F1	F1	F1	F1	F1	F9	F9																				
Régime 9	F1	F1	F1	F1	F1	F1	F1	F1																			
Régime 10	F1	F1	F1	F1	F1	F1	F1	F1	F9																		
Régime 11	F1	F1	F1	F1	F1	F1	F1	F1	F1	F1																	
Régime 12	F1	F1	F1	F1	F1	F1	F1	F1	F1	F1	Cons P																
Régime 13	F1	F1	F1	F2	F2	F1	F1	F1	F2	F2	F1	F2															
Régime 14	F1	F1	F1	F2	F2	F1	F1	F1	F2	F2	F1	F2	F5														
Régime 15	F1	F1	F1	F1	F1	F1	F1	F1	F1	F1	F1	F2	F1	F1													
Régime 16	F1	F1	F1	F1	F1	F1	F1	F1	F1	F1	F2	F2	F1	F1	F1												
Régime 17	F2	F2	F2	F1	F1	F1	F1	F1	F1	F1	F1	F2	F1	F1	F1	F1											
Régime 18	F2	F2	F2	F1	F1	F1	F1	F1	F1	F1	F1	F2	F1	F1	F1	F1	F9										
Régime 19	F2	F2	F2	F1	F1	F1	F1	F1	F1	F1	F1	F2	F1	F1	F1	F1	F3	F3									
Régime 20	F2	F2	F2	F1	F1	F1	F1	F1	F1	F1	F1	F2	F1	F1	F1	F1	F3	F3	F5								
Régime 21	F1	F1	F1	F2	F2	F1	F1	F1	F2	F2	F1	F2	F3	F3	F1	F1	F1	F1	F1	F1							
Régime 22	F1	F1	F1	F2	F2	F1	F1	F1	F2	F2	F1	F2	F9	F9	F1	F1	F1	F1	F1	F1	F3						
Régime 23	F1	F1	F1	F1	F1	F1	F1	F1	F1	F1	F1	F2	F1	F1	F3	F1	F1	F1	F1	F1	F1	F1					
Régime 24	F1	F1	F1	F1	F1	F1	F1	F1	F1	F1	F1	F2	F1	F1	F3	F1	F1	F1	F1	F1	F1	F1	F5				
Régime 25	F1	F1	F1	F1	F1	F1	F1	F1	F1	F1	F1	F2	F1	F1	F1	F1	F1	F1	F1	F1	F1	F1	F1	F1			
Régime 26	F1	F1	F1	F1	F1	F1	F1	F1	F1	F1	F1	F1	F1	F1	F1	F1	F1	F1	F1	F1	F1	F1	F1	F1	F1		
Régime 27	F1	F1	F1	F1	F1	F1	F1	F1	F1	F1	F1	F1	F1	F1	F1	F1	F1	F1	F1	F1	F1	F1	F1	F1	F1	F5	

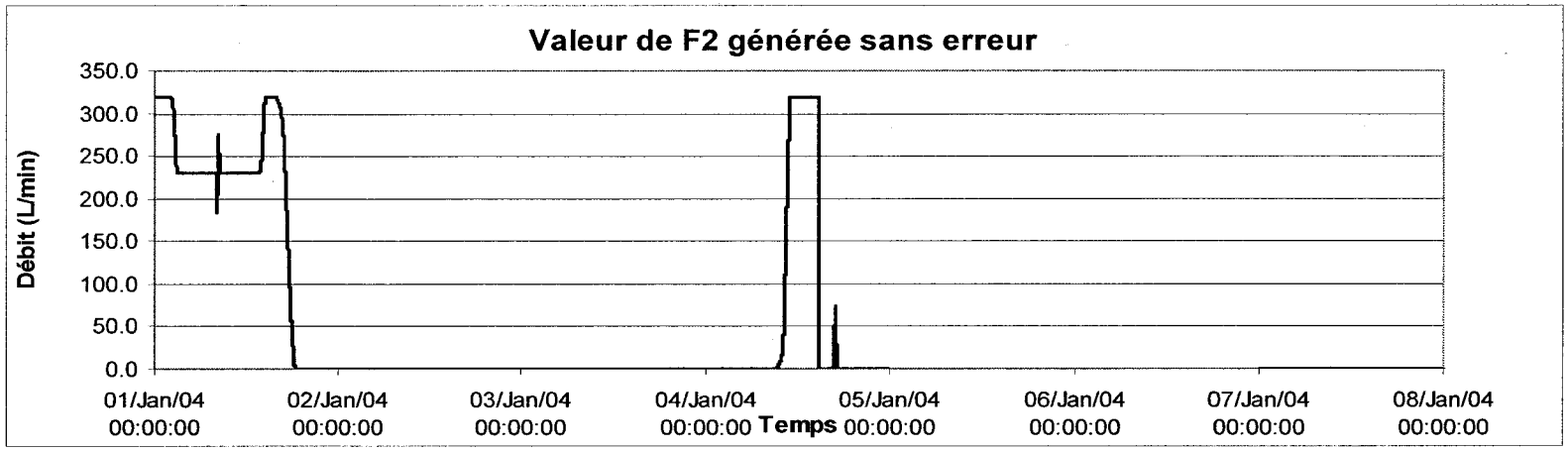
ANNEXE II

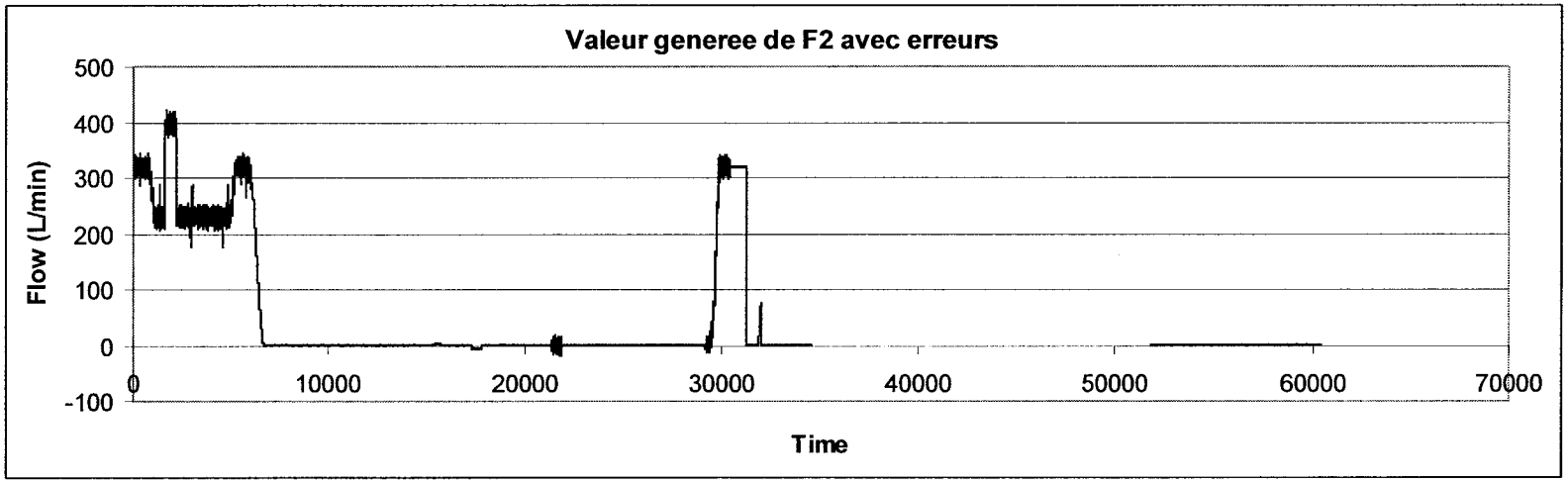
Données crées pour les mesures de débit pour le courant 1





Données crées pour les mesures de débit pour le courant 2





ANNEXE III

Section III-A : Résultats de l'analyse de sensibilité sur λ_1

Tableau A-III.1 : Évolution des erreurs de type I en fonction de λ_1

	Erreurs de type I		
	Lambda = 0.9	Lambda = 0.95	Lambda = 1
	(% Erreur)	(% Erreur)	(% Erreur)
F1	3.45	4.36	9.96
F2	3.57	4.08	9.09
F3	3.42	4.21	7.46
F4	3.88	4.12	9.17
F5	3.90	4.54	10.73
F6	3.33	4.08	8.20
F7	4.03	4.27	5.83
L8	3.25	4.16	5.15
F9	3.12	4.30	7.51
F10	3.72	4.55	6.26
F11	3.05	4.17	7.81
F12	3.76	4.54	7.36
F13	3.37	4.38	8.84
F14	3.24	4.52	10.54
F15	3.62	4.47	7.12
L16	4.35	4.44	6.50
CA	3.56	4.33	8.08
CB	3.00	4.20	5.09
CC	3.74	4.51	4.70
CD	3.67	4.19	4.39
CE	3.98	4.22	7.27
CF	3.57	4.46	8.07
CG	3.70	4.43	10.22
CH	3.94	4.42	9.71
CI	3.25	4.36	4.63
CJ	3.15	4.26	7.63
CK	4.35	4.51	6.28
CL	3.54	4.26	10.34
CM	3.50	4.17	4.30
CN	3.52	4.18	7.47
CP	3.92	4.46	9.16
Moyenne	3.60	4.33	7.58

Tableau A-III.2 : Évolution des erreurs de type II en fonction de λ_1

	Erreurs de type I		
	Lambda = 0.9	**Lambda = 0.95**	**Lambda = 1**
	(% Erreur)	(% Erreur)	(% Erreur)
F1	0.88	0.19	0.11
F2	0.65	0.24	0.23
F3	0.38	0.14	0.11
F4	0.33	0.09	0.08
F5	0.52	0.15	0.12
F6	1.01	0.21	0.20
F7	0.37	0.10	0.06
L8	0.27	0.06	0.03
F9	0.35	0.09	0.06
F10	0.47	0.18	0.09
F11	0.55	0.21	0.20
F12	0.39	0.18	0.10
F13	0.11	0.03	0.02
F14	0.47	0.11	0.09
F15	0.45	0.14	0.10
L16	0.70	0.17	0.12
CA	0.51	0.14	0.09
CB	0.88	0.21	0.20
CC	0.59	0.13	0.07
CD	0.19	0.09	0.05
CE	0.33	0.09	0.06
CF	0.04	0.01	0.01
CG	0.06	0.03	0.03
CH	0.05	0.02	0.01
CI	0.52	0.18	0.17
CJ	0.87	0.22	0.19
CK	0.74	0.17	0.15
CL	0.60	0.13	0.09
CM	0.45	0.18	0.16
CN	0.84	0.23	0.11
CP	0.22	0.10	0.08
Moyenne	**0.48**	**0.14**	**0.10**

Section III-B : Résultats de l'utilisation du test de Hampel

Légende :

GC : Anormalités de grande amplitude et de courte durée
FC : Anormalités de faible amplitude et de courte durée
GL : Anormalités de grande amplitude et de longue durée
FL : Anormalités de faible amplitude et de longue durée

Tableau A-III.3 : Résultats sur les erreurs de type I associés à l'application du test de Hampel selon les différentes classes d'anormalités

	Erreurs de type I			
	GC	FC	GL	FL
	(% Erreur)	(% Erreur)	(% Erreur)	(% Erreur)
F1	0.39	15.51	4.83	21.66
F2	0.65	16.24	3.88	18.62
F3	1.38	15.66	4.58	18.96
F4	1.24	12.75	5.26	20.14
F5	1.82	16.55	5.85	19.76
F6	1.73	18.54	4.73	19.96
F7	1.76	19.19	4.06	17.46
L8	0.94	18.44	6.06	26.91
F9	0.04	15.62	4.08	24.11
F10	1.28	14.84	3.81	20.13
F11	2.19	14.65	3.58	16.90
F12	0.46	13.86	5.88	21.98
F13	1.61	14.03	4.05	21.84
F14	1.39	12.60	3.43	20.79
F15	1.24	12.74	4.58	24.12
L16	1.66	19.50	6.19	26.70
CA	1.34	13.28	5.33	22.10
CB	1.49	18.21	5.22	22.76
CC	1.75	18.53	5.40	23.65
CD	0.19	15.81	3.71	17.14
CE	0.15	12.34	3.86	21.10
CF	1.73	13.47	4.01	20.61
CG	1.87	13.39	5.44	20.95
CH	1.58	13.75	5.44	20.70
CI	1.44	13.70	5.30	20.83
CJ	0.81	15.32	4.35	20.50
CK	0.37	16.44	6.27	23.60
CL	0.29	17.63	4.47	25.10
CM	0.31	16.69	6.29	23.70
CN	1.19	17.61	5.82	23.79
CP	0.78	12.38	5.70	22.53
Moyenne	1.13	15.46	4.89	21.58

Tableau A-III.4 : Résultats sur les erreurs de type II associés à l'application du test de Hampel selon les différentes classes d'anormalités

	Erreurs de type II			
	GC	FC	GL	FL
	(% Erreur)	(% Erreur)	(% Erreur)	(% Erreur)
F1	0.26	0.29	0.27	0.27
F2	0.13	0.15	0.15	0.13
F3	0.06	0.09	0.08	0.09
F4	0.06	0.07	0.06	0.10
F5	0.27	0.29	0.28	0.28
F6	0.28	0.29	0.29	0.29
F7	0.02	0.02	0.04	0.05
L8	0.34	0.35	0.34	0.39
F9	0.04	0.07	0.06	0.06
F10	0.25	0.26	0.27	0.28
F11	0.27	0.28	0.27	0.28
F12	0.21	0.24	0.23	0.22
F13	0.05	0.05	0.06	0.07
F14	0.18	0.22	0.18	0.24
F15	0.26	0.27	0.28	0.27
L16	0.36	0.37	0.37	0.40
CA	0.21	0.22	0.23	0.22
CB	0.23	0.26	0.25	0.24
CC	0.21	0.23	0.22	0.25
CD	0.07	0.09	0.10	0.09
CE	0.02	0.02	0.03	0.03
CF	0.23	0.26	0.25	0.25
CG	0.23	0.27	0.24	0.26
CH	0.23	0.26	0.25	0.27
CI	0.02	0.03	0.02	0.05
CJ	0.29	0.29	0.30	0.31
CK	0.28	0.28	0.29	0.31
CL	0.30	0.30	0.30	0.32
CM	0.20	0.23	0.26	0.29
CN	0.26	0.27	0.28	0.27
CP	0.01	0.02	0.03	0.05
Moyenne	0.19	0.20	0.20	0.21

ANNEXE IV

Section IV-A : Sélections finales du niveau de coupure pour filtrer les données de débits à l'aide des ondelettes

Variable	Niveaux de coupures sélectionnées
F1	5
F2	6
F3	6
F4	6
F5	5
F6	5
F7	4
L8	4
F9	6
F10	5
F11	6
F12	6
F13	5
F14	4
F15	5
L16	4

Section IV-B : Sélections finales du niveau de coupure pour filtrer les données de consistance à l'aide des ondelettes

Variable	Niveaux de coupures sélectionnées
CA	5
CB	7
CC	7
CD	7
CE	6
CF	7
CG	7
CH	7
CI	6
CJ	6
CK	5
CL	5
CM	5
CN	6
CP	6

ANNEXE V

Section V-A : Sélections finales de fenêtre de temps pour l'analyse par ondelette

Variables	Fenêtre de temps analysé (minutes)
F1	5
F2	10
F3	10
F5	10
F9	10
CP	15

Section V-B : Sélection optimale de λ_2 pour l'identification à l'aide des ondelettes

Variables	Lambda 2 optimal
F1	1.2
F2	1.3
F3	1.3
F5	1.3
F9	1.3
CP	1.4

Section V-C : Sélection optimale du point de coupure sur l'index de la méthode des ondelettes

Variables	Point de coupure optimal
F1	0.8
F2	0.85
F3	0.85
F5	0.8
F9	0.85
CP	0.9

Pour le système en entier : la valeur sélectionnée est de : 0.7

Section V-D : Sélection optimale du Rcrit lors de l'utilisation de la méthode de Cao et Rhinehart

Variables	Rcrit
F1	1.5
F2	1.1
F3	1
F5	1.2
F9	1.2
CP	1.3

Section V-E : Sélections optimales des valeurs de rho lors de l'utilisation de la méthode de Cao et Rhinehart

Variables	Rho1
F1	0.08
F2	0.1
F3	0.15
F5	0.15
F9	0.12
CP	0.15

Variables	Rho2
F1	0.2
F2	0.3
F3	0.3
F5	0.3
F9	0.4
CP	0.2

Variables	Rho3
F1	0.2
F2	0.3
F3	0.3
F5	0.3
F9	0.4
CP	0.2

Section V-F : Sélection optimale de la fenêtre de temps utilisé lors de l'application de la nouvelle méthode

Variables	Fenêtre de temps analysé (minutes)
F1	5
F2	10
F3	10
F5	10
F9	10
CP	15

Section V-G : Sélection optimale de λ_2 pour l'identification du point de départ du régime permanent

Variables	Lambda 2 optimal
F1	1.2
F2	1.3
F3	1.3
F5	1.3
F9	1.3
CP	1.4

Section V-H : Sélections optimales des valeurs de rho lors de l'utilisation de l'étape de filtration de la méthode proposée

Variables	Rho1
F1	0.6
F2	0.76
F3	0.82
F5	0.85
F9	0.8
CP	0.86

Variables	Rho2
F1	0.05
F2	0.1
F3	0.06
F5	0.04
F9	0.1
CP	0.04

Variables	Rho3
F1	0.005
F2	0.005
F3	0.005
F5	0.005
F9	0.005
CP	0.005

Section V-H : Sélections optimales de la valeur de alpha lors de l'étape d'identification du point final par la méthode proposée

Variables	alpha 3
F1	0.8
F2	0.75
F3	0.75
F5	0.75
F9	0.75
CP	0.7

Section V-I : Taux d'erreurs de type I associée à l'application de chacune des méthode pour chacun des régime permanent créé

Sur F1 (Résultats en % d'erreurs)

Régime	New Method	Wavelet	CAO
1	45.36	46.99	66.67
2	3.70	25.74	3.15
3	0.00	24.62	9.68
4	7.96	31.85	6.11
5	1.65	11.02	25.34
6	2.22	30.77	0.55
7	1.67	33.61	1.67
8	2.78	28.89	1.67
9	7.13	26.30	2.87
10	0.31	28.53	6.98
11	4.09	31.42	6.58
12	8.44	27.11	4.00
13	14.02	31.53	22.63
14	12.47	32.12	20.66
15	8.64	38.54	7.75
16	1.48	31.96	1.92
17	1.19	32.74	0.48
18	5.30	33.64	9.49
19	0.76	29.52	1.02
20	8.83	24.28	3.09
21	4.95	27.60	5.22
22	3.09	33.53	4.17
23	25.16	46.65	21.63
24	0.00	27.33	0.00
25	4.77	29.51	9.39
26	3.41	27.31	18.47
27	9.35	33.28	13.24
28	1.68	23.85	0.00
29	10.62	37.04	0.49
30	23.53	41.18	15.29
31	0.52	34.90	6.77
32	0.00	30.93	0.00
33	0.00	21.15	10.24
34	0.00	30.55	0.00
35	0.83	25.70	2.10
36	16.37	34.69	27.79
37	2.99	28.03	3.54

Sur F2 (Résultats en % d'erreurs)

Régime	New Method	Wavelet	CAO
1	7.84	11.43	29.05
2	6.24	17.21	10.45
3	0.70	13.75	3.84
4	0.00	3.41	0.18
5	0.00	12.36	0.00
6	6.53	13.61	20.30
7	7.90	15.73	20.69
8	9.74	14.76	30.88
9	3.10	10.04	3.94
10	5.08	16.05	2.03
11	0.00	10.33	0.00
12	4.83	11.47	15.63

Sur F3 (Résultats en % d'erreurs)

Régime	New Method	Wavelet	CAO
1	9.68	10.75	66.31
2	6.84	16.78	14.79
3	1.46	11.30	3.50
4	0.00	0.00	11.83
5	0.00	20.83	0.00
6	5.02	15.73	18.27
7	7.65	15.07	27.97
8	0.00	10.54	0.51
9	6.37	14.03	23.18
10	0.00	15.70	0.50
11	8.78	18.58	18.46
12	0.00	15.92	0.00
13	0.44	14.24	1.31
14	7.66	14.02	25.08
15	0.17	8.20	0.00
16	8.56	22.92	12.43
17	0.00	12.76	0.00
18	2.23	12.15	1.88
19	0.00	9.82	0.39
20	8.98	11.89	37.34
21	7.67	18.32	12.25
22	0.49	11.66	2.88
23	4.07	12.26	15.23

Sur F5 (Résultats en % d'erreurs)

Régime	New Method	Wavelet	CAO
1	3.32	7.68	33.96
2	0.00	10.07	0.00
3	0.00	7.61	0.00
4	1.82	8.00	8.55
5	0.00	3.20	2.46
6	0.42	7.98	5.12
7	0.76	7.49	9.48
8	0.46	4.41	9.38
9	0.00	7.30	2.21
10	4.82	8.97	20.07
11	0.00	3.84	2.92
12	0.00	8.98	1.23
13	2.15	9.45	10.62
14	0.00	6.96	1.47
15	4.95	10.67	19.91
16	1.36	7.69	0.00
17	0.27	7.65	1.64
18	2.40	10.13	17.74
19	1.10	7.34	12.75
20	0.00	6.00	11.39
21	4.57	11.58	15.49
22	0.85	7.35	13.16
23	4.20	10.64	5.32
24	0.00	5.80	3.56
25	5.12	12.08	6.46
26	1.85	7.40	13.00
27	3.32	7.83	17.54
28	2.65	9.86	11.15
29	0.00	8.21	1.57
30	0.00	5.01	0.49
31	0.00	5.32	19.68

Sur F9 (Résultats en % d'erreurs)

Régime	New Meth	Wavelet	CAO
1	10.47	12.20	59.27
2	0.00	17.49	0.66
3	0.00	7.59	0.00
4	0.00	13.05	0.00
5	4.77	11.93	6.86
6	0.00	12.08	0.00
7	2.52	16.88	7.38
8	12.08	14.34	28.22
9	7.60	13.11	25.67
10	3.95	13.97	8.85
11	6.08	14.72	13.00
12	8.74	17.13	20.76
13	12.36	17.70	32.36
14	3.52	15.08	3.52
15	0.56	12.73	2.51
16	0.57	6.42	3.28
17	2.71	14.45	10.55
18	1.54	5.05	15.40
19	7.11	13.21	31.86
20	6.74	13.13	5.25
21	3.22	17.00	3.95
22	24.18	24.91	28.21
23	2.43	5.11	9.73
24	0.00	7.35	0.85
25	4.27	11.10	22.87
26	0.00	10.96	4.21
27	5.69	13.23	14.76

Sur CP (Résultats en % d'erreurs)

Régime	New Method	Wavelet	CAO
1	2.02	6.53	53.95
2	7.33	66.67	66.67
3	0.00	9.79	2.93
4	0.00	8.55	0.00
5	5.74	7.92	24.77
6	4.43	10.49	21.00
7	2.27	6.41	30.03
8	6.21	12.56	26.18
9	2.11	6.32	26.40
10	4.55	8.55	32.93
11	16.30	18.52	47.41
12	3.97	6.10	27.12
13	4.57	5.82	28.35
14	8.70	12.36	33.93
15	2.95	5.86	27.83
16	0.00	5.00	8.53
17	9.15	12.33	38.22
18	1.47	4.78	22.49

ANNEXE VI

Section VI-A : Valeur entrées dans la réconciliation pour chaque régime identifiée et correction finale après les itérations

Pour F1

Pour Cao	
Valeur moyenne	Correction
230.9	0
313.4	1.8
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0

Pour ondelettes	
Valeur moyenne	Correction
628.3	116.3
624.3	1.2
1258.8	-138
715.5	-1.2
716.7	-1.8
716.8	0.1
715.9	-1.3
713.8	3.3
713.8	0.6
716.4	-1
714.6	0.1
714.8	-0.1
716.4	-0.1
716.5	-1.2
1473.4	0.4
1477.2	-1.7
2253.1	-0.1
713.4	0.3
715.8	0
713.8	-2.3
714.1	0.1
43.1	-42.3
1	-0.9
1.1	-1.1
0.2	-0.1
13990.7	6
13982.3	7.5
1196.6	277.2
1174.2	238.4
1179.2	219.6
715.7	0.2
713.8	0.9
714.5	-16.9
715.8	0
715	-0.2
1.2	-1.2
123.6	-135.1
427.6	-2
424.4	-0.8
588.1	-152.3
598	-26.4
1474.3	1.6
1476.4	-2.6
427.1	-1.7
563	-57
714.4	0.8
714.4	0.2
715.7	-0.1
714.5	0.1
716.3	-1
2248.6	-0.9

Pour nouvelle méthode	
Valeur moyenne	Correction
826.5	1.5
623.9	0
1257.5	0
628.4	-2.5
625.6	3
1260.3	0.2
1257.3	-9.2
714.8	0.2
715.4	0.2
715.9	-0.5
716.5	0.4
713.4	1.2
713.8	0.3
714.4	-0.1
716	-0.2
715.8	0
1474.7	1.2
1475.9	0
2252.3	-0.3
715.7	0
714.5	0.2
716	-1.1
715.4	-0.7
714.9	-0.1
13800.3	58.8
0.5	0.3
1.1	-1
0	0
717.2	3
13530.5	8
13994.7	-0.4
13979.8	7.5
1575.2	2.2
1251.9	40.5
1178.7	270.7
715.8	0
715	-0.2
714.4	-0.9
0.3	-0.3
1476.6	-0.5
13835.4	129.8
13976.7	11.4
435.9	0.3
424.8	-0.1
422.9	-1.9
1542	-5.6
1476.1	-3
1472	1.4
426.8	-0.3
424.8	0.4
424.7	0.1
713.8	-0.2
714.9	1
714.9	-0.4
713	0.2
717	-0.1
714.4	-0.4
713.9	0
715.4	0.1

Pour F2

Pour Cao	
Valeur moyenne	Correction
230.9	0
313.4	1.8
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0

Pour ondelettes	
Valeur moyenne	Correction
231.3	0
229.2	0
317.4	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
1	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0

Pour nouvelle méthode	
Valeur moyenne	Correction
320.1	0
236	-0.1
231.1	0
231.3	0
228.9	1.2
228.8	0
316.3	1
0.3	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
1.2	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0.3	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0

Pour F3

Pour Cao	
Valeur moyenne	Correction
3424.6	0
3079	4.4
3684.6	1
3677.8	-8
3692	-0.8
3661.1	-2.5
2351.3	0
3653.4	-1.3
3678.3	-0.6
3678.1	0.3
0.9	9.6
50.7	10.7
3667.7	6.4
3694	-10
2354.4	-1.2

Pour ondelettes	
Valeur moyenne	Correction
3341	-2.2
3422	0
3077.5	0
3683.7	-1.7
3674.2	-1
3687.8	-2.8
3656	15.3
3695.2	-12.5
3695.2	-17.9
2397.5	-38.9
2354.4	0
3661.3	4.4
3638.1	31.3
3683.7	0
3662.1	3.8
3673.8	0.1
3667.8	0
2334.8	12.2
2349.9	0
2349.3	-0.1
3681.4	0
183	-164.8
0.5	-0.4
0.3	-0.3
0	0
3658.7	13
3680	0
3681.1	0.6
3677.1	95.3
3633.2	244.5
3662.4	10.1
3673.2	-7.8
3687.7	-10.2
3645.7	13.6
3676.5	2
0.9	-0.6
0	0
3664.9	9.9
3684.8	-5.6
2352.6	0
2347.9	0
2352.6	0.5
2353.6	-1.1
3687.9	10.3
1747.8	1978.1
2353	0
2353.5	0
2345.5	0
3677.1	0
3666.2	7
3683.5	-2.2

Pour nouvelle méthode	
Valeur moyenne	Correction
3074.2	0.1
3426.1	0
3423.6	0.1
3339	-2.3
3423.9	2
3389.3	-0.4
3078.6	-0.7
3672.7	2.6
3681	-1.1
3677.6	-0.9
3689.7	-0.7
3659.9	0.5
3653.4	4.5
2354.4	0
2362.1	-0.7
3659.7	2.5
3659.7	9
3679.1	-0.4
3670.1	0
3675.3	0
3668.8	0.7
2370.5	-0.4
2348	0
2352.2	0
3670.1	0
1.2	0
0.9	0
0.2	0
3626.4	0
3675.9	0
3658.1	3.5
3681.2	0
3677.4	0
3679.3	-0.6
3634.4	37.5
3662.4	0
3670.2	2
3689.2	-0.3
0.6	-0.6
47.8	9.2
80.8	0.1
3.1	-0.1
3664.5	10
3679	-2.9
2355.5	-0.9
2349.4	0
2354.4	-0.7
2349.9	0.7
3682.4	0
3669.2	1.5
3683	-1.8
3652.3	22.5
2353.1	0
2353.6	0.5
2350.2	0
2347	0
2350.7	-1.8
3673.8	0.4
3674.2	0

Pour F4

Valeur moyenne	Correction
800.1	0
799.5	0.6
906.2	0
906.1	0
799.9	0
801.8	0
800	0
800.5	0
799.3	0
800.4	0
800.2	0
800.4	0.2
800.4	0
799.8	1.7
800.4	0

Valeur moyenne	Correction
800.2	0
799.9	0
798.8	0
905.9	-84.8
800.4	0
799.9	0
799.8	0
800.3	0
800.3	-0.1
825.1	0
825.9	-20.7
800.2	0
799.8	0
798.2	0
906.4	-85.2
825.6	-20.4
799.4	0
799.6	0.9
800	0
801.9	-1.5
801	-0.8
800.2	0
799.1	0
799.7	0
0.3	-0.3
799.7	0
799.9	0
801.9	-1.5
802.3	-1.7
799.7	0.2
799.7	0
800.2	0
799.9	0
784.5	14.2
752.3	38.2
800.3	0
799.4	0
800.2	0
800.2	-0.1
837.3	-29.9
800.5	0
800.7	0
800.3	-0.1
799.6	0
800.3	0
800.1	0
800.8	-0.6
799.7	0
801.1	-0.9
799.9	0
799.4	0.5

Valeur moyenne	Correction
801.7	-0.2
799.9	0
800	0
800.2	0
799.7	0.1
799.6	0
799.4	0.4
901.6	-89.1
906.4	-73
800.1	0
799.9	0
800	0
799.6	0.1
825.8	0
799.5	0
800.9	0
906.6	-86.1
800.6	0
799.6	0
799.9	0
799	0
799.9	0
800.1	0
800.1	0
800	0
799.3	0
799.9	0
42.3	-35.8
800.3	0
799.4	0
799.6	0
800	0
800.5	0
802	0.1
799.7	0.2
799.8	0
800	0
799.8	0
799.8	0
800.4	-0.1
800.7	0
800	0
800.2	0
800.1	0
799.9	-0.1
800.1	0
800.4	-0.1
800.1	0
799.5	0
800.6	0
800.9	0
799.8	0
800.1	0
800.2	-0.4
798.7	0
799.8	0
799.4	-0.1
799.9	0
801.2	0

Pour F5

Pour Cao	
Valeur moyenne	Correction
4594	1.7
4102.7	-3
4230.2	35.2
4420	6.9
4423.7	-0.9
4257.7	1.7
4262.1	4.1
4258.1	-0.7
4259.6	0.1
4272.7	-4.3
0.2	-0.2
2295.1	3.2
4258.4	0.6
4245	-5
4260.5	3.8

Pour ondelettes	
Valeur moyenne	Correction
4625.2	1.3
4625.7	0.9
4109.9	-8
4424.2	-0.1
4406.1	15.6
4424	3
4427	-5.9
4423.5	11.2
4423.5	-0.3
4424.5	-6.8
4422.7	0.1
4286.5	-21.4
4265.1	-1.3
4257.1	1.3
4255.1	0.6
4133	2.9
4259.6	2.6
4237.9	16.3
4066.4	0.1
4143	-0.1
4425.3	-0.6
4426.5	-1.2
0.2	-0.2
0	0.2
0	0
4259.8	-0.4
4256.9	-0.6
4261.4	-0.7
4247.8	-0.2
4256.8	-0.1
4267.8	0.2
4424.4	0
4427.2	0
4426.5	-0.1
4409.6	0
58.2	-52.4
94.1	-84.4
4260.6	-0.4
4259.3	-6
3998.2	14.6
4256.6	7.5
4258.2	0.4
4258.5	-1.8
3998.1	3.4
4938.9	-542.6
4254.3	-0.3
4239.3	26
4258.3	-1.2
4273.6	2
4424	-0.2
4260.4	-7.6

Pour nouvelle méthode	
Valeur moyenne	Correction
4461.4	1.3
4103.8	0.9
4588.4	-0.6
4625.2	-0.1
4625.3	2.6
4619.9	3
4103.7	-5.9
4335.8	11.2
4430	-0.3
4402.1	-6.8
4423.7	0.1
4428.4	-1.1
4428.7	-1.3
4425.3	1.3
4241.6	0.6
4267.5	2.9
4257.6	2.6
4163.2	16.3
4259.3	0.1
4254.3	-0.1
4262.8	-0.6
4258.9	-1.2
4065.2	-10.2
4256.2	1
4424	0
0.4	-0.4
0	-0.6
0.6	-0.7
4258.1	-0.2
4259.2	-0.1
4260.6	0.2
4257.1	0
4269.5	0
4257.8	-0.1
4256.4	0
4268.2	0.2
4415	2.6
4427.3	-0.4
42.4	-32.6
2257.5	14.6
3573.8	23.9
5138.5	0.4
4259.5	-1.8
4247.8	3.4
3999.9	2.4
4259.4	-0.3
4259.2	47
4264.5	-1.2
3999	2
4007.1	-0.2
4030.9	-40.5
4257	-0.4
4257.4	2.8
4246.2	3.6
4260.1	0.6
4257.2	1.3
4271.8	-14.6
4262.5	-0.4
4425.7	-0.7

Pour F9

Pour Cao	
Valeur moyenne	Correction
10628.1	3.6
8506.3	0.8
9516.2	4
9515	412.7
9969.7	-0.3
9475.7	0
9957.3	2.7
9978.4	-0.1
9892.7	-5.8
9470	-0.1
0	0
9430.5	279.7
9477.3	0.7
9831.6	-8.5
11816.3	-18.6

Pour ondelettes	
Valeur moyenne	Correction
9672.4	-18.1
9683.4	-2.2
8502.5	-1.9
9522.5	404.8
9976.2	-1.2
9967.5	-0.2
9964.5	-2.9
9957.3	1
9957.3	0.2
9960.8	2.2
9956.8	7.7
9479.2	-0.1
9475.6	0.1
9442.8	4.1
8655.8	672.5
8653.4	657.5
9408.6	64.2
9955	4.6
9978.6	-0.2
9966.9	-0.2
9975	3.7
345.4	-327
0.7	-0.7
2.7	-2.7
5.8	-5.8
9830.1	-2.2
9817	0.6
9481.7	0
9481.4	1.2
9470.7	-0.1
9475.4	0.8
9474.4	-0.3
9473.9	-0.2
9475.9	5.6
9465.3	1.7
0	0
204.6	-200.4
9503.7	-5.3
9498.4	-0.3
9478.8	-166.2
9478.9	3.7
9478.9	-14.8
9478.2	-4.2
9478.9	5.2
9472.6	-97.8
9476.4	0.1
9479.4	0
9478.2	2.2
9478.4	-0.4
9478.9	0
9478.4	-4.5

Pour nouvelle méthode	
Valeur moyenne	Correction
8700.7	-0.5
10528.9	174.2
10628.9	6.3
9672.7	-18.1
9672.3	-7.8
8523.7	-1.3
8502.7	-0.6
9509.8	32.9
9967.9	0.5
9975.5	-2.7
9967.6	-0.2
9982.9	-5.6
9974	-3.7
9950.1	3.3
9471.9	1.2
9476.7	0.1
8680.6	766.8
8650.7	521.7
9439.1	29.1
9901.4	-21.7
9969.5	-1.1
9975.2	-5.1
9972.6	-2.1
9956.1	14.7
9935.6	38.9
0	0
2.2	-0.4
0.7	-0.6
9978.7	-2.4
9895.5	-7.3
9830.5	-4.1
9817.2	0.6
9469.9	0
9482.1	0
9470.5	1.5
9475.1	0.1
9473.6	0.7
9474.2	-0.2
2.3	-2.3
9437.7	105
9479.7	-3
9435.1	37.8
9487.6	-5.3
9717.6	-3.6
9466.9	17.9
11893	-78.3
11825.1	-24.2
11823.3	-27.6
11849.7	-27.6
11691.1	106.9
11950.7	-126.4
9467.2	7.4
9477.8	0.1
9477.2	-2.6
9467.5	4.8
9472.2	2.2
9473.3	1.5
9475.7	-0.1
9980.7	-9.1

Pour CP

Pour Cao	
Valeur moyenne	Correction
3.82	0.02
3.83	0
3.83	-0.008
3.82	0
3.81	-0.01
3.77	0
3.77	0
3.82	0
3.83	0
3.85	0
3.83	-0.01
3.78	0
3.79	0
3.79	0
3.73	-0.01

Pour ondelettes	
Valeur moyenne	Correction
3.85	0
3.84	0
3.85	0
4.49	-0.57
3.84	0
3.83	0
3.82	0
3.82	-0.01
3.8	0
3.81	0
3.8	0
3.79	0
3.77	0
3.77	0
3.77	0
3.76	0
3.77	0
3.77	0
3.77	0
3.76	0
4.03	0
3.96	0
3.77	0
3.82	0
3.81	0
3.81	0
3.85	0
3.86	0
3.87	0
3.87	0
3.88	0
3.83	0
3.78	0
3.79	0
3.79	0
3.78	0
3.78	0
3.87	0
3.75	0
3.72	0
3.72	0
3.81	0
3.79	0
3.81	0
3.85	0
3.82	0
3.8	0
3.78	0
3.78	0
3.76	0
3.75	0
3.74	0

Pour nouvelle méthode	
Valeur moyenne	Correction
3.74	0
3.85	0
3.84	0
3.85	0
4.49	-0.57
3.84	0
3.83	0
3.82	0
3.82	-0.01
3.8	0
3.81	0
3.8	0
3.79	0
3.77	0
3.77	0
3.77	0
3.76	0
3.77	0
3.77	0
3.77	0
3.76	0
4.03	0
3.96	0
3.77	0
3.82	0
3.81	0
3.81	0
3.81	0
3.82	0
3.83	0
3.83	0
3.83	0
3.85	0
3.85	0
3.86	0
3.87	0
3.87	0
3.88	0
3.83	0
3.78	0
3.79	0
3.79	0
3.78	0
3.78	0
3.87	0
3.75	0
3.72	0
3.72	0
3.81	0
3.79	0
3.81	0
3.85	0
3.82	0
3.8	0
3.78	0
3.78	0
3.76	0
3.75	0
3.74	0

Section VI-B : Résultats quotidiens de la réconciliation des données en terme de précision et d'exactitude pour l'application de la méthode de Cao

Pour F1 :

	n S.S	Average exactitude après	Average exactitude avant
Day 1	12	0.18	0.18
Day 2	7	0.21	0.22
Day 3	5	15.14	25.45
Day 4	8	5.66	22.70
Day 5	5	2.01	1.70
Overall	37	3.64	8.68

	n S.S	Precision après	Précision avant
Day 1	12	0.23	0.24
Day 2	7	0.46	0.49
Day 3	5	23.38	45.17
Day 4	8	9.04	40.61
Day 5	5	3.37	3.56
Overall	37	9.82	25.83

Pour F2 :

	n S.S	Average exactitude après	Average exactitude avant
Day 1	5	5.03	5.27
Day 2	1	5.61	5.20
Day 3	1	1.87	37.45
Day 4	4	17.08	17.25
Day 5	1	0.32	12.18
Overall	12	8.44	12.52

	n S.S	Precision après	Précision avant
Day 1	5	9.47	9.45
Day 2	1	NA	NA
Day 3	1	NA	NA
Day 4	4	31.68	32.04
Day 5	1	NA	NA
Overall	12	18.62	23.50

Pour F3 :

	n S.S	Average exactitude après	Average exactitude avant
Day 1	6	0.09	0.12
Day 2	3	0.10	0.11
Day 3	5	13.72	13.74
Day 4	5	13.62	17.61
Day 5	4	2.95	2.96
Overall	23	6.49	7.37

	n S.S	Precision après	Précision avant
Day 1	6	0.16	0.19
Day 2	3	0.12	0.12
Day 3	5	33.81	33.81
Day 4	5	18.36	24.01
Day 5	4	5.76	5.77
Overall	23	16.75	18.44

Pour F4 :

	n S.S	Average exactitude après	Average exactitude avant
Day 1	1	2.76	2.77
Day 2	1	1.24	1.24
Day 3	3	6.93	30.20
Day 4	1	0.53	0.54
Day 5	1	0.13	0.13
Overall	7	3.63	13.61

	n S.S	Precision après	Précision avant
Day 1	1	NA	NA
Day 2	1	NA	NA
Day 3	3	11.89	52.03
Day 4	1	NA	NA
Day 5	1	NA	NA
Overall	7	7.60	34.34

Pour F5 :

	n S.S	Average exactitude après	Average exactitude avant
Day 1	7	0.04	0.06
Day 2	6	0.01	0.01
Day 3	5	17.83	4.25
Day 4	6	5.40	17.93
Day 5	7	0.83	0.84
Overall	31	4.12	4.36

	n S.S	Precision après	Précision avant
Day 1	7	0.05	0.08
Day 2	6	0.01	0.02
Day 3	5	39.47	8.83
Day 4	6	9.78	40.22
Day 5	7	2.12	2.14
Overall	31	16.64	18.22

Pour F9 :

	n S.S	Average exactitude après	Average exactitude avant
Day 1	7	0.25	0.46
Day 2	5	1.35	1.35
Day 3	5	7.31	7.21
Day 4	6	25.58	26.00
Day 5	4	0.11	0.04
Overall	27	7.37	7.49

	n S.S	Precision après	Précision avant
Day 1	7	0.41	0.81
Day 2	5	2.96	2.97
Day 3	5	15.37	15.15
Day 4	6	49.48	50.24
Day 5	4	0.17	0.06
Overall	27	22.87	23.18

Pour CP :

	n S.S	Average exactitude après	Average exactitude avant
Day 1	5	0.06	0.08
Day 2	3	0.04	0.03
Day 3	3	1.55	1.55
Day 4	4	0.07	0.10
Day 5	3	0.06	0.08
Overall	18	0.31	0.32

	n S.S	Precision après	Précision avant
Day 1	5	0.09	0.14
Day 2	3	0.02	0.02
Day 3	3	2.66	2.66
Day 4	4	0.10	0.12
Day 5	3	0.06	0.09
Overall	18	1.09	1.09

Section VI-C : Résultats quotidiens de la réconciliation des données en terme de précision et d'exactitude pour l'application de la méthode basée sur les ondelettes

Pour F1 :

	n S.S	Average exactitude après	Average exactitude avant
Day 1	12	1.16	0.18
Day 2	7	0.21	0.22
Day 3	5	14.54	25.45
Day 4	8	5.62	22.70
Day 5	5	0.38	1.70
Overall	37	3.65	8.68

	n S.S	Precision après	Précision avant
Day 1	12	2.98	0.24
Day 2	7	0.46	0.49
Day 3	5	25.18	45.17
Day 4	8	8.94	40.61
Day 5	5	0.66	3.56
Overall	37	10.26	25.83

Pour F2 :

	n S.S	Average exactitude après	Average exactitude avant
Day 1	5	4.98	5.27
Day 2	1	5.32	5.20
Day 3	1	1.32	37.45
Day 4	4	17.20	17.25
Day 5	1	0.51	12.18
Overall	12	8.40	12.52

	n S.S	Precision après	Précision avant
Day 1	5	9.49	9.45
Day 2	1	NA	NA
Day 3	1	NA	NA
Day 4	4	31.87	32.04
Day 5	1	NA	NA
Overall	12	18.75	23.50

Pour F3 :

	n S.S	Average exactitude après	Average exactitude avant
Day 1	6	0.20	0.12
Day 2	3	0.09	0.11
Day 3	5	12.71	13.74
Day 4	5	13.76	17.61
Day 5	4	2.17	2.96
Overall	23	6.19	7.37

	n S.S	Precision après	Précision avant
Day 1	6	0.41	0.19
Day 2	3	0.11	0.12
Day 3	5	31.20	33.81
Day 4	5	28.29	24.01
Day 5	4	4.21	5.77
Overall	23	18.12	18.44

Pour F4 :

	n S.S	Average exactitude après	Average exactitude avant
Day 1	1	2.59	2.77
Day 2	1	1.03	1.24
Day 3	3	6.82	30.20
Day 4	1	0.45	0.54
Day 5	1	0.07	0.13
Overall	7	3.51	13.61

	n S.S	Precision après	Précision avant
Day 1	1	NA	NA
Day 2	1	NA	NA
Day 3	3	11.71	52.03
Day 4	1	NA	NA
Day 5	1	NA	NA
Overall	7	7.50	34.34

Pour F5 :

	n S.S	Average exactitude après	Average exactitude avant
Day 1	7	0.03	0.06
Day 2	6	0.01	0.01
Day 3	5	9.33	4.25
Day 4	6	12.77	17.93
Day 5	7	0.65	0.84
Overall	31	4.13	4.36

	n S.S	Precision après	Précision avant
Day 1	7	0.04	0.08
Day 2	6	0.02	0.02
Day 3	5	20.46	8.83
Day 4	6	27.60	40.22
Day 5	7	1.63	2.14
Overall	31	15.11	18.22

Pour F9 :

	n S.S	Average exactitude après	Average exactitude avant
Day 1	7	0.35	0.46
Day 2	5	1.08	1.35
Day 3	5	6.00	7.21
Day 4	6	26.13	26.00
Day 5	4	1.05	0.04
Overall	27	7.37	7.49

	n S.S	Precision après	Précision avant
Day 1	7	0.67	0.81
Day 2	5	2.35	2.97
Day 3	5	12.47	15.15
Day 4	6	49.51	50.24
Day 5	4	1.59	0.06
Overall	27	22.47	23.18

Pour CP :

	n S.S	Average exactitude après	Average exactitude avant
Day 1	5	0.06	0.08
Day 2	3	0.04	0.03
Day 3	3	1.55	1.55
Day 4	4	0.07	0.10
Day 5	3	0.06	0.08
Overall	18	0.31	0.32

	n S.S	Precision après	Précision avant
Day 1	5	0.09	0.14
Day 2	3	0.02	0.02
Day 3	3	2.66	2.66
Day 4	4	0.10	0.12
Day 5	3	0.06	0.09
Overall	18	1.09	1.09

Section VI-D : Résultats quotidiens de la réconciliation des données en terme de précision et d'exactitude pour l'application de la méthode développée

Pour F1 :

	n S.S	Average exactitude après	Average exactitude avant
Day 1	12	0.23	0.18
Day 2	7	0.22	0.22
Day 3	5	14.86	25.45
Day 4	8	5.26	22.70
Day 5	5	1.64	1.70
Overall	37	3.48	8.68

	n S.S	Precision après	Précision avant
Day 1	12	0.28	0.24
Day 2	7	0.47	0.49
Day 3	5	23.77	45.17
Day 4	8	8.23	40.61
Day 5	5	3.47	3.56
Overall	37	9.69	25.83

Pour F2 :

	n S.S	Average exactitude après	Average exactitude avant
Day 1	5	5.00	5.27
Day 2	1	7.16	5.20
Day 3	1	1.32	37.45
Day 4	4	17.05	17.25
Day 5	1	0.40	12.18
Overall	12	8.41	12.52

	n S.S	Precision après	Précision avant
Day 1	5	9.48	9.45
Day 2	1	NA	NA
Day 3	1	NA	NA
Day 4	4	32.06	32.04
Day 5	1	NA	NA
Overall	12	18.80	23.50

Pour F3 :

	n S.S	Average exactitude après	Average exactitude avant
Day 1	6	0.19	0.12
Day 2	3	0.09	0.11
Day 3	5	12.69	13.74
Day 4	5	10.27	17.61
Day 5	4	2.88	2.96
Overall	23	5.55	7.37

	n S.S	Precision après	Précision avant
Day 1	6	0.39	0.19
Day 2	3	0.11	0.12
Day 3	5	31.19	33.81
Day 4	5	16.54	24.01
Day 5	4	5.70	5.77
Overall	23	15.22	18.44

Pour F4 :

	n S.S	Average exactitude après	Average exactitude avant
Day 1	1	1.44	2.77
Day 2	1	1.05	1.24
Day 3	3	6.92	30.20
Day 4	1	0.53	0.54
Day 5	1	0.13	0.13
Overall	7	3.42	13.61

	n S.S	Precision après	Précision avant
Day 1	1	NA	NA
Day 2	1	NA	NA
Day 3	3	11.87	52.03
Day 4	1	NA	NA
Day 5	1	NA	NA
Overall	7	7.64	34.34

Pour F5 :

	n S.S	Average exactitude après	Average exactitude avant
Day 1	7	0.14	0.06
Day 2	6	0.02	0.01
Day 3	5	5.64	4.25
Day 4	6	12.41	17.93
Day 5	7	0.85	0.84
Overall	31	3.54	4.36

	n S.S	Precision après	Précision avant
Day 1	7	0.31	0.08
Day 2	6	0.03	0.02
Day 3	5	12.20	8.83
Day 4	6	26.71	40.22
Day 5	7	2.11	2.14
Overall	31	13.15	18.22

Pour F9 :

	n S.S	Average exactitude après	Average exactitude avant
Day 1	7	0.39	0.46
Day 2	5	1.08	1.35
Day 3	5	7.17	7.21
Day 4	6	25.74	26.00
Day 5	4	0.09	0.04
Overall	27	7.36	7.49

	n S.S	Precision après	Précision avant
Day 1	7	0.74	0.81
Day 2	5	2.37	2.97
Day 3	5	15.09	15.15
Day 4	6	49.74	50.24
Day 5	4	0.16	0.06
Overall	27	22.94	23.18

Pour CP :

	n S.S	Average exactitude après	Average exactitude avant
Day 1	5	0.06	0.08
Day 2	3	0.04	0.03
Day 3	3	1.55	1.55
Day 4	4	0.07	0.10
Day 5	3	0.06	0.08
Overall	18	0.31	0.32

	n S.S	Precision après	Précision avant
Day 1	5	0.09	0.14
Day 2	3	0.02	0.02
Day 3	3	2.66	2.66
Day 4	4	0.10	0.12
Day 5	3	0.06	0.09
Overall	18	1.09	1.09

ANNEXE VII

Section VII-A : Publication pour Pulp and Paper Canada

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

Steve Bellec[1], Taiwen Jiang[1], Bruce Kerr[2], Martin Diamond[3], Paul Stuart[1]

[1] NSERC Design Chair in Process Integration, Department of Chemical Engineering, École Polytechnique - Montréal
Contact: paul.stuart@polymtl.ca
[2] Kruger Inc., Montreal, Quebec
[3] Kruger Inc., Trois-Rivières, Quebec

ABSTRACT

In order to increase the value of real-time data from the process data management systems implemented at many pulp and paper mills in recent years, on-line process data treatment techniques have been evaluated in this study. The goal of applying data processing and reconciliation techniques is to correct for various sources of errors including random noise, abnormalities and inconsistencies – and thus have better quality data for decision-making. The application of these techniques on-line can potentially assist mills to quickly identify instrument maintenance requirements, to troubleshoot operating problems, and to improve decision-making related to process activities. This paper presents a robust and practical methodology for on-line data processing and steady state reconciliation that improves the accuracy and precision of measurements when the process is operating at near steady state conditions.

INTRODUCTION

A variety of decisions such as planning and scheduling are generally made based on steady state data sets. To better be able to take such decisions in real time, pulp and paper mills capture real time process data in their information management systems. Janssen *et al* (2004a) showed that mill personnel recognize that the implementation of data management systems improve the accessibility to real time data but until now however, the data have generally been used for impromptu process analysis. Only in a few cases have data been used for process-wide analysis or optimization, e.g. for plant-wide multivariate analysis of processes (Harrison *et al*, 2003), for plant-wide process optimization (Dabros *et al*, 2005), or for plant-wide business modeling (Janssen *et al*, 2004b). Part of the reason for this may be attributable to data quality. Measurements are inherently corrupted with various sources of error including instrument miscalibration or malfunction, power supply fluctuation, wiring and process noise. Unless they are carefully treated, using process data to make process decisions in real time can lead to misinterpretations caused by data quality.

Figure 1 shows a set of flowrate measurements containing the 3 different types of error present in data. Random errors are mostly concentrated in high frequencies and can be associated with power supply fluctuation and signal conversion noise. An Abnormality is a supernatural random error. Ths high frequency and amplitude measurement can be caused by electric voltage surges and wiring problems. Gross errors are associated to bias and drifting in measurements, complete failure and precision degradation of instruments (Narasimhan and Jordache 2000), and are related to instrument wear, corrosion, miscalibration or malfunction and inconsistencies in a set of measurements around a system, e.g. process leaks or accumulation.

Data processing and reconciliation techniques have been applied in the pulp and paper industry for various **off-line** applications. For example Jacob and Paris (2003a and 2003b) 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. This paper proposes an on-line methodology for improving steady state data quality in **real time**.

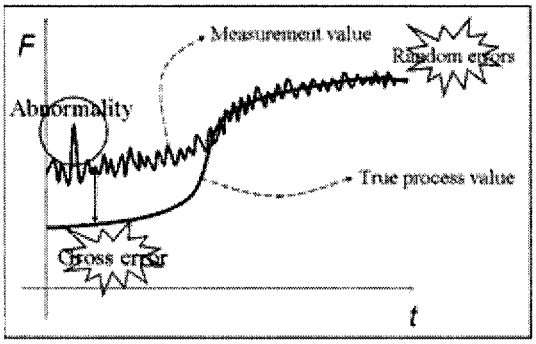


Figure 1: Types of errors in measurements

METHODOLOGY

Figure 2 illustrates the systematic methodology used in this study based on Jiang *et al* (2003) to reduce the various types of errors included in a set of measurements. First, abnormal measurements were detected and corrected, followed by random error correction. To correct gross errors, steady state data reconciliation was performed. For this, near steady state data sets must be carefully identified, which was done using a novel technique based on wavelet transform and statistical theories.

Critically for on-line application, the methodology included a *system status identification*. This step has two main purposes, as follows:

- to select data processing parameters from a library which were developed based on historical data
- to identify data tuning parameters to be used in each of the different steps of the data treatment methodology, based on the measurement sources and their link to the system operation regimes.

Selection of data processing parameters and identification of tuning parameters are based on the association of unique combinations of measurements values and offline properties investigation for optimal results in the data treatment procedure.

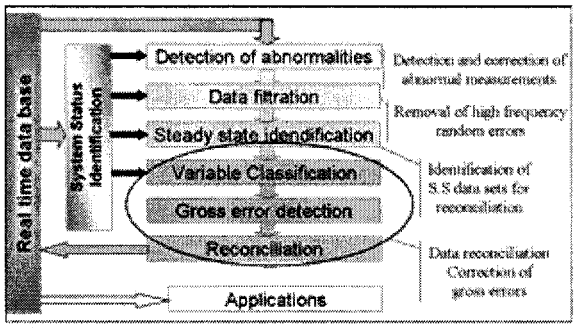


Figure 2: General methodology applied in this study to improve data quality

For the first two steps in the methodology shown in Figure 2, data treatment using wavelet transform was compared to a) utilization of statistical theory based on the median and its absolute deviation to detect abnormal features, and b) low pass filtering to correct random errors. Due to their ability to detect abnormalities over short or long time periods and their power to avoid distortion in the corrected signal, wavelet transforms were preferred over the alternatives (Bellec *et al*, 2004).

Data reconciliation corrects the gross errors in redundant measurements around a system by comparing their reliability using mass and energy balance relationships. Due to the importance of redundancy in data reconciliation, the first task to accomplish in data reconciliation is to classify the variables under study (Bagajewicz 2001). If the measurements are redundant, then gross error detection and reconciliation can be performed. To do so, errors between measurements and real process values are minimized by considering physical constraints and using a least square equation. Figure 3 illustrates this in a hypothetical case by treating measured values (100, 61, 38) to arrive at reconciled values (100, 60.3, 39.7).

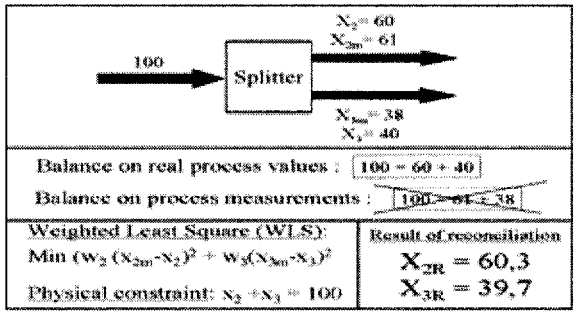


Figure 3: Example of reconciliation problem around a splitter

CASE STUDY

The applicability and performance of the proposed methodology for on-line processing and steady state data reconciliation was tested using real time data from the stock approach system at an integrated newsprint paper mill (Figure 4). The mill has a process data management system (PI system supplied by OsiSoft Inc.). In order to assess and evaluate the methodology, the system was investigated in 3 steps: 1) process data analysis and creation of a hypothetical "measured" data set from a "true" data set, 2) systematic analysis of the data processing and reconciliation methodology using the hypothetical data set, and finally 3) application of the methodology on process real time measurements.

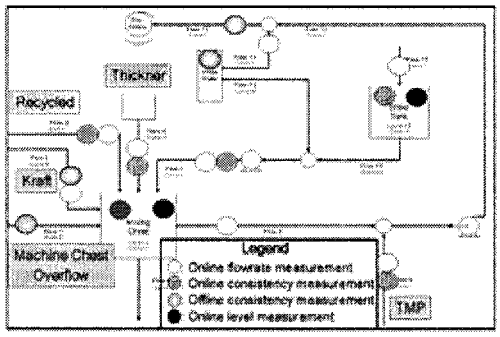


Figure 4: Paper machine stock approach system studied

Before starting the detailed study around the system, a redundancy analysis was performed. By using engineering judgment to couple on-line measurements with other data, it was possible to improve the measurement redundancy and therefore the possibility for reconciling data around the system.

Process Data Analysis and Creation of Hypothetical "True" and "Measured" Data Sets

To be able to characterize the accuracy and precision of the methodology, treated data must be compared with the true value of the process. Since on-line measurements are always corrupted with errors, the true value is never exactly known. Therefore, in order to investigate the accuracy of this methodology, hypothetical data were created.

In order to create hypothetical data that are representative of actual process operations, a 3 step study of the system was performed as follows.

Signal processing and near steady state operation regime identification: 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 coupled to identify overall system steady state regimes. Twenty-seven different operating regimes were grouped into four system status conditions. The system status conditions were: production of paper grade A, production of paper grade B, maintenance stops, and during the occurrence of a paper machine break.

Steady state measurement correction through mass balances: Once near steady state operating regimes had been identified, measurements were reconciled to fit mass balances for flow and fiber content in the different process streams. After correction, the probability of occurrence, length and delays, relative to each status and regime of operation were compiled based on historical data. Figure 5 summarizes the twenty-seven operation regimes grouped in the four system status conditions of the system. An example demonstrating the differences between two regimes of operation within the same status is also shown.

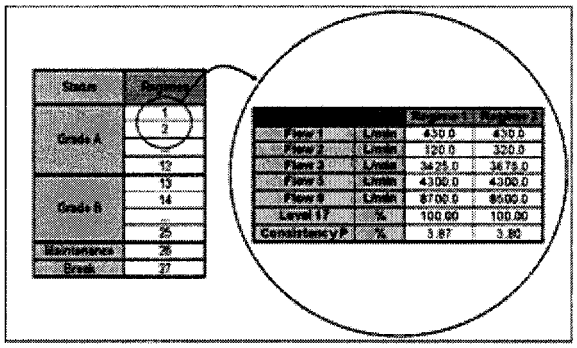


Figure 5: System status and regimes of operation identified

Creation of hypothetical data with known errors: Based on the system analysis and statistics associated with each signal, hypothetical data where created. In order to generate the corrupt hypothetical set of

measurements, errors were then applied to the true value. Abnormalities and random errors where added to the signals according to the standard deviation and accounting for the probability of abnormal measurement presence in the historical data.

To demonstrate how well the hypothetical data represented the real process operation, operation regimes of the system where analyzed over a one week period and reproduced by the hypothetical data. Figure 6 shows the result of that investigation for a particular flow rate measurement in time. The general characteristics of the two signals is the same, but one can notice that the hypothetical data can contain more or less noise in particular regions, and that part of the signal has been corrected for overall mass balances in the system.

Investigation of Data Processing and Reconciliation Methodology Using the Hypothetical Data Sets

The various methodologies for on-line steady state data reconciliation were applied using the hypothetical data set. Figure 7 shows an example of the results obtained, for a particular flow rate measurement. At each step of the methodology, the raw measurements in the process being corrected are compared to the targeted value ("true" hypothetical data). In the first two steps, abnormal measurements and random errors are corrected using wavelet transform. Steady state identification is performed and steady state data sets are selected. Finally, the selected data are reconciled through mass balances on the overall system. One can notice the improvement in the accuracy of the corrected steady state measurement compare to the raw data.

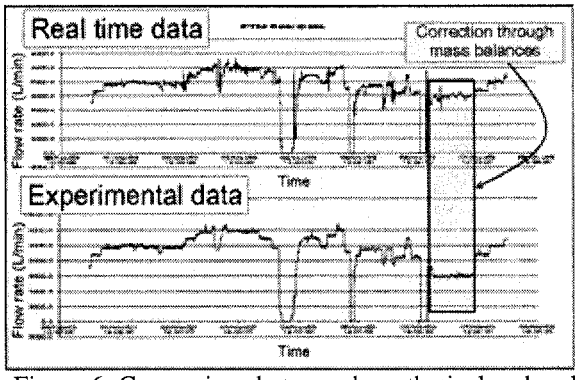


Figure 6: Comparison between hypothetical and real time data for the same operation regimes

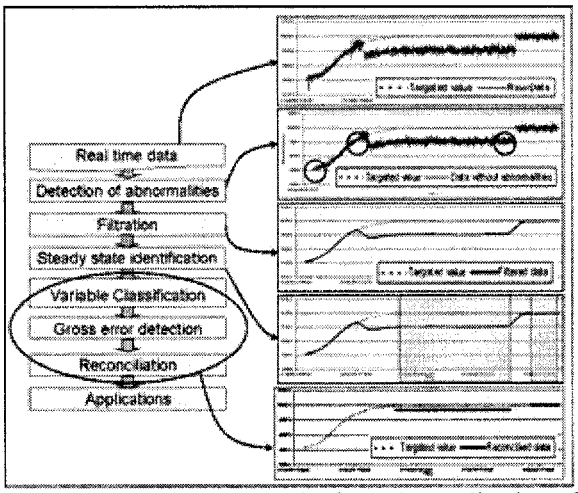


Figure 7: Step by step results from the application of on-line processing and reconciliation methodology

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. More specifically, Figure 8 shows the relative error improvement for the different measured variables, while Figure 9 reflects the effect of the methodology on the standard deviation of measurements.

Accuracy is defined as the average of relative error (%RE) from the targeted value (Equation 1) while precision is shown as the standard deviation ($SD_{\%RE}$) of such errors (Equation 2). Therefore, lower values indicate improvements in accuracy and precision. Improvements in the accuracy underline the overall improvement of steady state data due to the methodology, while the improvement of precision

underlines the efficiency of the method for correcting gross errors.

$$\%\mathrm{RE} = 100\% * \frac{|\mathrm{Targeted.value} - \mathrm{measurement}|}{\mathrm{Targeted.value}} \quad (1)$$

$$SD_{\%RE} = \sqrt{\frac{\sum_{i=1}^{n}\left[100\% * \left(average(RE) - RE_i\right)^2\right]}{n-1}} \quad (2)$$

Through the application of the proposed methodology, it can be concluded that overall accuracy and precision has been improved. In general, the methodology has a larger effect on flowrate than on measurements of consistency. This can be explained by the fact that flowrate measurements are associated with two levels of mass balance constraints (overall and fiber mass balances), and that consistency measurements are only implicated only in the fiber mass balances.

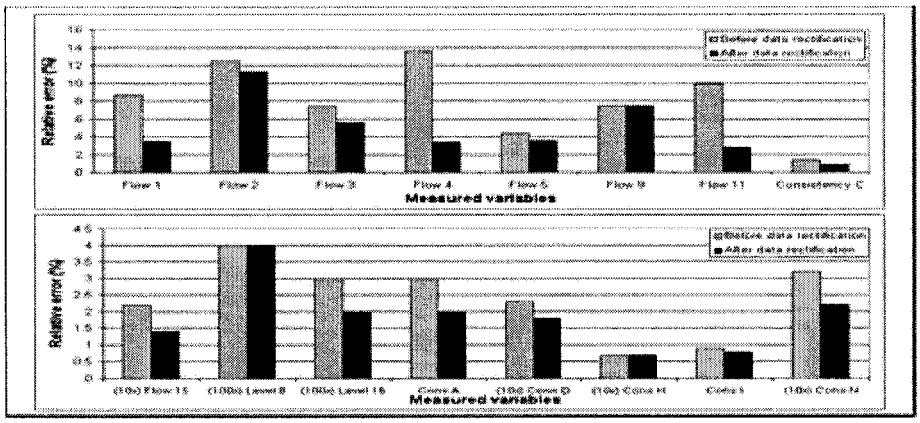


Figure 8: Summary of improvement in data **accuracy** from the application of the data treatment methodology

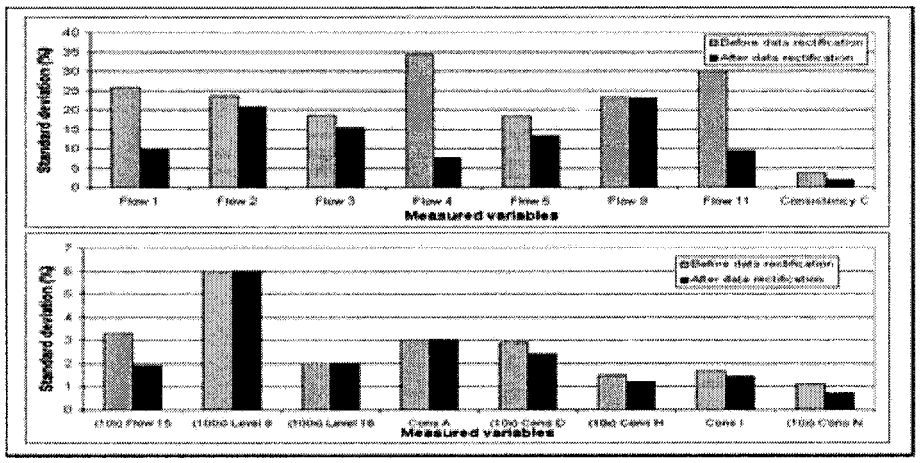


Figure 9: Summary of improvement in data **precision** from the application of the data treatment methodology

Evaluation of Methodology Robustness Using Real Time Process Measurements

After proving that by applying the proposed methodology, overall accuracy and precision of hypothetical process data can be improved, real time data were treated for longer time periods. The method under investigation was found to be robust for treating real time process data using 10 flow and 8 consistency signals over a time period of 3 days. In order to ensure that process time lags were taken into account, only steady state periods over a certain time threshold were considered.

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

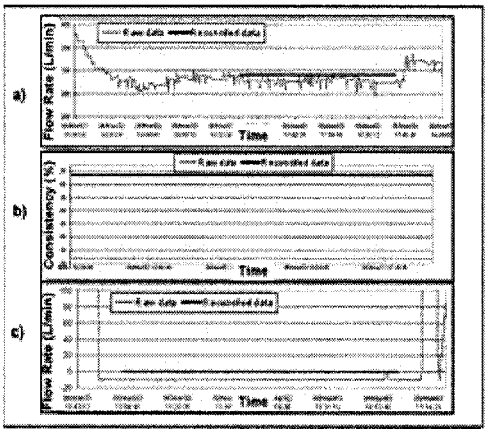


Figure 10: Examples of real time measurements improvement:
a) Accuracy improvement
b) Total failure correction c) Offset correction

One should be aware that, as the system under study becomes larger, the amount of operating regimes quickly increases, and information needed for online data reconciliation will augment considerably. Therefore, process operating regimes must be carefully and logically combined in order to minimize the different operating regimes.

IMPLICATION OF RESULTS

A novel methodology for data processing and reconciliation has been illustrated using real-time data from an integrated newsprint mill. 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 were identified, inconsistencies were removed using data reconciliation. By applying this data treatment methodology, data accuracy and reliability were improved. From this activity, confidence in data quality for well-defined process operating regimes can be achieved (called "direct benefits" in Figure 11), and this alone can likely justify the effort and expense of implementing such a system. In this section, the implications and **further benefits** of implementing such an on-line data treatment system are outlined.

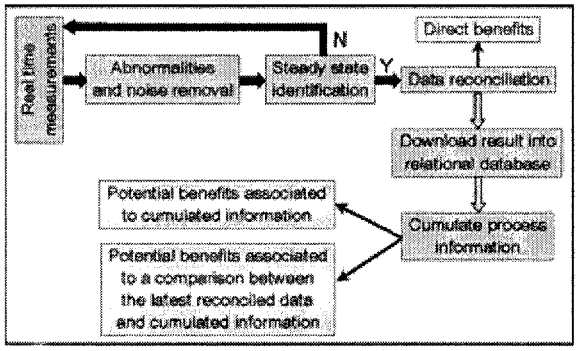


Figure 11: Types of Benefits Associated with the Application of the Proposed Methodology

By compiling on-line data reconciliation results for near steady state operating regimes, mills can develop relational databases of accurate process data. This data might be expressed as the probability of occurrence of different operating parameters. Figure 12 shows the example of the probability of occurrence of a flow rate measured at a process point, during a specified operating regime.

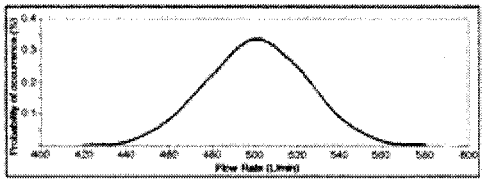


Figure 12: Example of Cumulated Process Information from Successive Online Data Reconciliation Results

This could potentially lead to multiple benefits related to process operation analysis and troubleshooting. Depending on the preliminary steps necessary to diagnose potential benefits, those can be classified in two major categories (Figure 11):

- Benefits obtained as a consequence of changes in the data reconciliation outcome compared to many previous on-line analyses, and
- 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 addressed:

- By comparing gross error detection and correction results for the current data set with previous data sets, one can quickly identify out-of-calibration instruments and process leaks.
- One can perform real-time yield accounting (performance analysis) for near steady-state operating conditions.

In the second case, information, issued from the compiled data reconciliation results (illustrated in Figure 12) can be used to localize, assess and analyze process operations. The better-resolved data at different process operating regimes improves understanding of process during near steady state operating conditions. This definition of different process operations can be used in combination with process simulation models to examine "what-if" scenarios.

The compiled data from different operating regimes issued from the reconciliation methodology can also enable more precise conditions for optimization techniques to enhance process planning (i.e. scheduling and plant-wide optimization). 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.

CONCLUSIONS

A robust methodology to improve steady state data accuracy and precision was assessed in this study. Improvements in data precision and accuracy were demonstrated by generating hypothetical "true" and "measured" data sets based on actual process data from an integrated newsprint mill.

The application of this methodology on-line could yield significant benefits including rapid instrument failure/process leak detection, improved process yield calculations, and better 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.

REFERENCES

Bagajewicz, M. (2001) "Process Plant Instrumentation and Upgrade", Technomic Publishing company, Lancaster, Pennsylvania

Bellec, S., Jiang, T., Stuart, P.R. (submitted for publication December 2004), "Steady State Identification for On-line Data Reconciliation Based on Wavelet Transform and Filtering", *Comp Chem Eng*

Dabros, M., Perrier, M., Forbes, F., Fairbank, M., Stuart, P.R. (2005). "Model-Based Direct Search Optimization of the Broke Recirculation System in a Newsprint Mill", accepted to *Journal of Cleaner Production.*

Jacob, J. and Paris, J. (2003a), "Data Sampling and Reconciliation – Application to Pulp and Paper Mills. Part I: Methodology and Implementation", *Appita J.*, 56(1), 25-52.

Jacob, J. and Paris, J. (2003b), "Data Sampling and Reconciliation – Application to Pulp and Paper Mills. Part II: Case Studies", *Appita J.*, 56(2), 116-120

Janssen, M., Laflamme-Mayer, M., Zeinou, M.-H., Stuart, P.R. (2004a), "Survey Indicates Mill Needs to Exploit IT Systems with New Business Model", *Pulp and Paper*, 105:6, pp. 46-51.

Janssen, M., Stuart, P.R. (2004b). "Business Model Framework for Using Real-Time Process Data for Design Decision Making". Proceedings, C-DEN Design Conference, Montreal, QC. Comm. 07B5.

Jiang, T., Chen, B., Jasim, K. Stuart, P.R. (2003), "Strategy for Improving Data Quality for a Kraft Pulp Mill Recausticizing Plant", FOCAPO, p 467-470

Harisson, R., Leroux, R., Stuart, P.R. (2003) "Multivariate Analysis of Refiner Operating Data From a Thermo-Mechanical Pulp Newsprint Mill", *Pulp & Paper Canada*, 105:4, pp. 24-27.

Narasimhan and Jordache (2000), "Data Reconciliation and Gross Error Detection: an Intelligent use of Process Data", Gulf Publishing Company, Houston, Texas

Section VII-B : Publication pour Computers and Chemical Engineering

Steady state identification for on-line data reconciliation based on wavelet transform and filtering

Steve Bellec, Taiwen Jiang, Paul Stuart[1]

NSERC Design Engineering Chair in Process Integration, Department of Chemical Engineering, École Polytechnique de Montréal
PO Box 6079, Succ. Centre-Ville, Montreal Que., Canada H3C 3A7

Abstract

In this paper a process steady state detection technique suitable for on-line application is developed, based on wavelet transform (WT) and filtering. First, measurement abnormalities and high frequency components are removed by using a wavelet based multi-scale processing technique. Then steady state is detected in 3 steps, as follows. 1) The start point of the steady state period is detected by using the first order WT and its first derivative. 2) Steady state duration is approximate by using an f-test to compare the ratio of the filtered square deviation from the mean and the filtered squared difference of successive data. 3) Finally, the end point of the steady state period is identified by using first order WT features. Compared to other methods tested, this method is efficient at identifying steady state periods, improved steady state data reconciliation results, and is suitable for on-line implementation. A case study based on data from the stock approach system of a paper machine is presented to illustrate the method.

Keywords: Steady state identification; data reconciliation; filtering; wavelet transform; on-line application; data quality.

[1] Telephone +1 514 340 4711 extension 4384; fax +1 514 340 5150
Email address: paul.stuart@polymtl.ca

Nomenclature

A	Distance between measured value and steady state average
B	Distance between successive measurements
D1	Filtered distance between measured value and steady state average
D2	Filtered distance between successive measurements
$\alpha_1, \alpha_2, \alpha_3$	Threshold used for steady state starting and ending point identification
$\beta_1, \beta_2, \beta_3$	Filtering parameters
$\sum_{i \epsilon I_j} C_{j,i} \varphi_{j,i}$	Smoothed signal
$\sum_{k \epsilon K_L} d_{j,k} \psi_{j,k}$	Detailed signal
σ	Standard deviation
WT	Wavelet transform
$\frac{dWT}{dt}$	First derivative of wavelet transform
$\bar{x}$	Filtered average

1. Introduction

In recent years, the emergence of plant-wide data management systems has improved accessibility to real time data and consequently, the possibility of enhancing value extracted from data by its analysis, e.g. for plant-wide multivariate analysis of processes (Harrison *et al*, 2003), for plant-wide process optimization (Dabros *et al*, 2005), or for plant-wide business modeling (Janssen *et al*, 2004). Unfortunately, measurements are inherently corrupted with various sources of errors, e.g. instrument miscalibration or malfunction, power supply fluctuation, wiring and process noise. Therefore, real time data must be processed and reconciled before being used in these applications. There are still additional benefits possible from *on-line* data processing and reconciliation. In the present study, methods for on-line determination of near steady state process conditions are addressed.

False detection of the process steady state can lead to misinterpretation of true process features, especially if the incorrect steady state data are subsequently reconciled. Under-estimating the true process steady state periods can result in only partial correction of gross errors (Figure 1a), while over-estimating steady state periods can lead to false input to data reconciliation (Figure 1b).

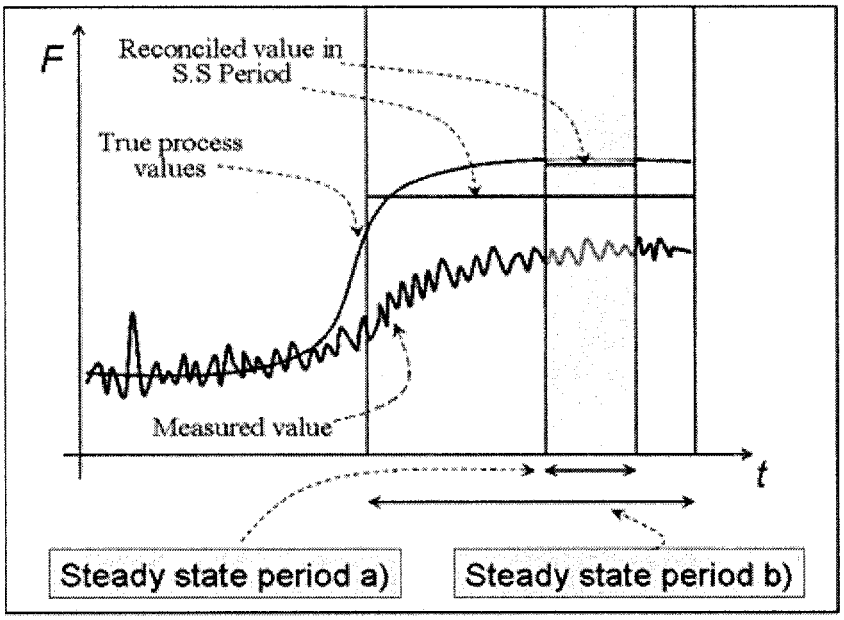


Figure 1: Inaccurate Estimation of Steady State Periods
b) Under-Estimated Period b) Over-Estimated Period

A variety of techniques for on-line process status identification have been proposed in the literature. Bakshi and Stephanopoulos (1993) developed a geometric approach for the description of process trend. Cao and Rhinehart (1995) proposed a steady state identification technique based on low pass filtering technique and statistical theory. Flehmig *et al* (1998) used wavelet transform features to approximate process measurements by a polynomial of limited degree and identify trend in process. Nounou and Bakshi (1999) used wavelet features to identify and remove random and gross errors. More recently Jiang *et al* (2003a) used 1st and 2nd wavelet transform features to correct process measurements and extract process trends. In the so-called direct approach, linear regression of the measured value is calculated over a data window, and a t-test is performed on the regression slope. This approach is executed over a specified time period, which is not ideal when dealing with real time measurements.

Cao and Rhinehart (1995) proposed a method based on the comparison of data variances calculated in different ways. In this method, a weighted moving average is

used to filter the sample mean. Then, the filtered mean square deviation from the new mean is compared with the filtered squared difference of successive data. This method uses a low pass filter to estimate the mean value. On the one hand, the computational requirements and storage are significantly reduced. On the other hand, low pass filters are less sensitive to abnormalities, and cause time delays in transient periods. Furthermore, using a weighted average to filter the calculated variances creates a delay in the characterization of process measurement frequency. These delays cause problems of detection in periods where the frequency varies in real time measurements. Figure 2 shows the effect of a sudden change in signal amplitude on steady state identification when using a low pass filter to estimate the signal variance. One can notice that the presence of frequency variation causes a problem of identification at the moment of such variation in the measurement signal.

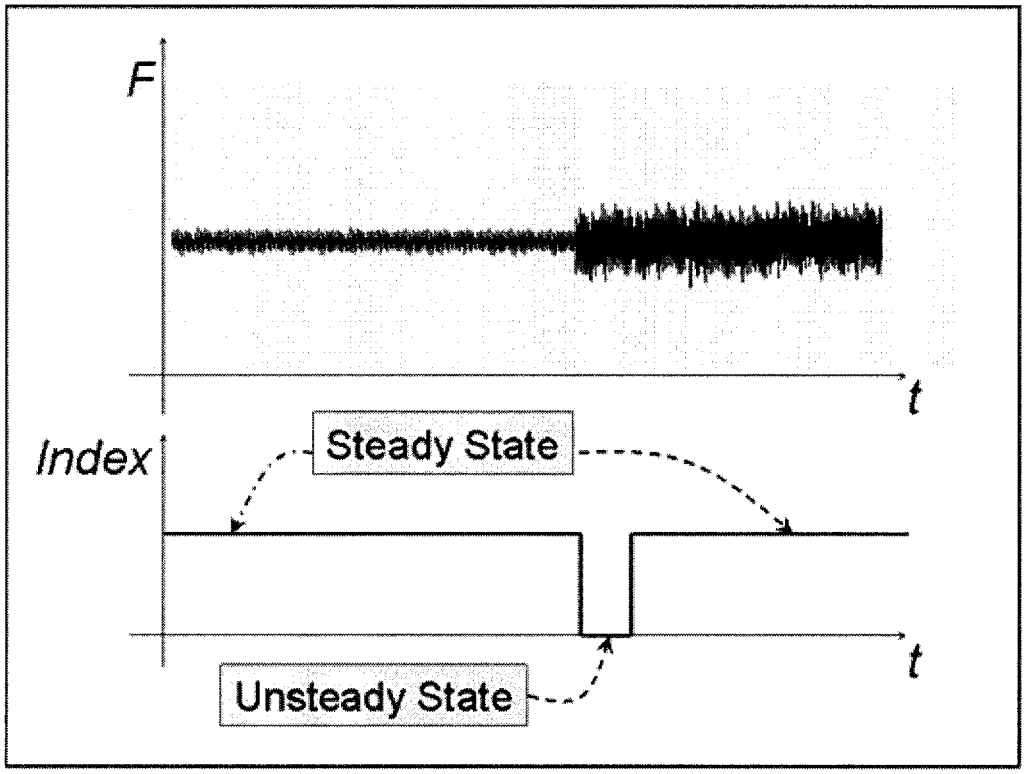


Figure 2: Steady State Detection Problems Related to Method of Filtering Variances by a Low Pass Filter

Jiang *et al* (2003a) proposed a wavelet transform (WT) based method for the detection of near steady state periods. The wavelet based multi-scale data processing technique is used to eliminate random noise and abnormalities. Then, the process status is analyzed according to the modulus of the first and second order wavelet transforms. The method has the ability to analyze the high frequency components and abnormalities accurately. When applying the multi-scale method, the accurate choice of scale is critical. When selecting a scale that is too low, the WT will be corrupted by high frequency noise (Figure 3a), i.e. process status identification is corrupted by temporal features. If a too high is selected, then process measurements are excessively smoothed which creates distortion (Figure 3b). This creates a deviation from the true process trend and leads to an incorrect reflection of process status. Jiang *et al* (2003a) proposes to select the optimal scale by taking in consideration the response time constants and sampling intervals. This criterion is adequate for off-line purposes, but not practical for on-line treatment of real time data because on-line measurements can be corrupted with different high frequency features over time. Therefore, the scale choice must be known *a priori* for on-line wavelet based treatment of real time data. Furthermore, this method uses the second order WT of the signal to distinguish zero-crossing points from steady state periods. The second WT is directly proportional to the second derivative of the smoothed signal at the sample cutting scale. It is adequate to represent process trend but requires high computational storage and requirements.

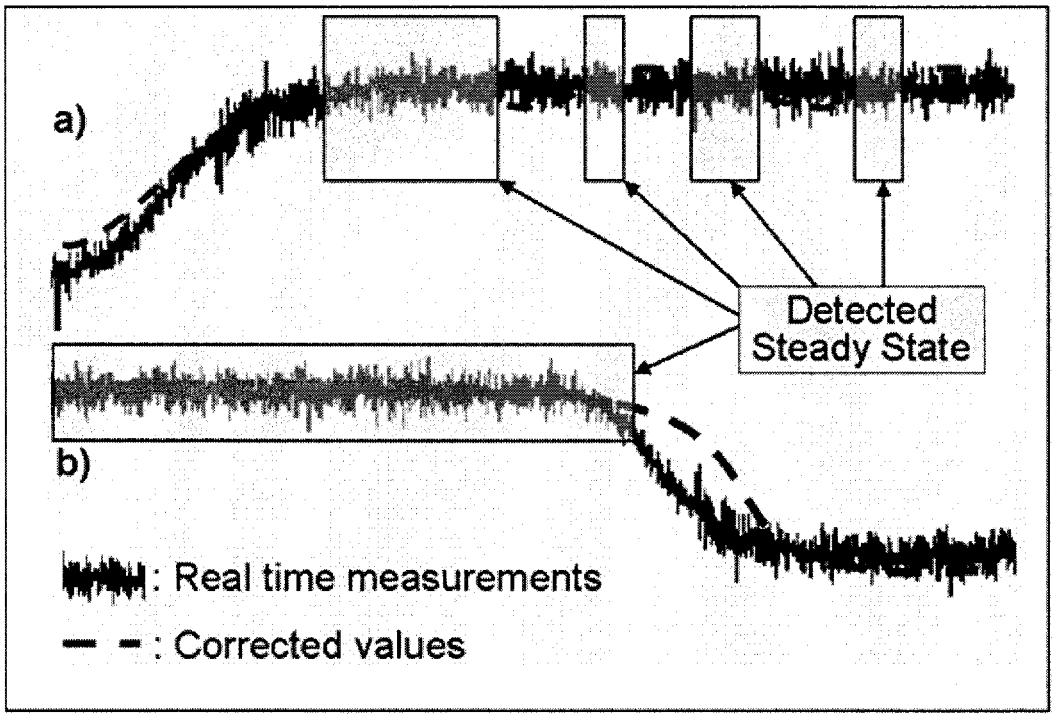


Figure 3: Scaling Choice for On-Line Data Treatment
c) Impact of Under-Scaling on Steady State Identification
d) Impact of Over-Scaling on Steady State Identification

In this paper, a steady state detection technique for on-line estimation of process status is proposed. First, by using multi-scale wavelet data processing (coupled with historical process data analysis in order to select appropriate wavelet cutting scale), random noise and abnormalities are eliminated. Then, the process status is evaluated by a 3-step method based on wavelet transform and statistical theory. The identification of the steady state period start point is identified using wavelet transform and its first derivative. Then the steady state duration is approximated by coupling an f-test with filtration. Finally the end point of the period is identified by using wavelet transform features.

2. Elimination of random noise and abnormalities

High frequency process measurement features corrupt the process trend, and have a direct impact on process status identification. Jiang *et al* (2000) proposed a multi-scale wavelet method for processing measurements which is effective for removing noise and detecting abnormalities in real time data sets. The approach for steady state detection presented in this study builds on this multi-scale wavelet methodology.

2.1 Representing process measurement features by multi-scale wavelet approach

The objective of the multi-scale wavelet technique is to represent a real time measurement as a limit successive approximation (Figure 4). First, the process measurement signal ($\Sigma C_{0,i}\varphi_{0,i}$) is decomposed into smoothed ($\Sigma C_{1,i}\varphi_{1,i}$) and detailed signals ($\Sigma D_{1,k}\Psi_{1,k}$). $\Sigma C_{1,i}\varphi_{0,i}$ represents the low frequency feature of the process measurements, while $\Sigma D_{1,k}\Psi_{1,k}$ represents the higher one. In subsequent scales(j), the smoothed signal of the previous scale ($\Sigma C_{j-1,i}\varphi_{j-1,i}$) is also decomposed in a smoother signal ($\Sigma C_{j,i}\varphi_{j,i}$) and a lower frequency detailed signal ($\Sigma D_{j,k}\Psi_{j,k}$). Multi-scale wavelet cuts the real time measurements signal into different frequency components represented at each scale. Therefore it is critical to identify the proper scale(j) that allows noise elimination in the detailed signals, and does not remove the true trend (Mallat and Hwang 1992).

$$f(t) = f_0 = \sum_{i \in I_0} C_{0,i}\varphi_{0,i} = \sum_{i \in I_1} C_{1,i}\varphi_{1,i} + \sum_{k \in K_1} d_{1,k}\psi_{1,k}$$
$$= \sum_{i \in I_2} C_{2,i}\varphi_{2,i} + \left(\sum_{k \in K_1} d_{1,k}\psi_{1,k} + \sum_{k \in K_2} d_{2,k}\psi_{2,k}\right)$$
$$= \cdots$$
$$= \sum_{i \in I_j} C_{j,i}\varphi_{j,i} + \left(\sum_{k \in K_j} d_{j,k}\psi_{j,k} + \sum_{L=1}^{j}\sum_{k \in K_L} d_{L,k}\psi_{L,k}\right)$$

Equation 1 [from Jiang *et al* (2000)]

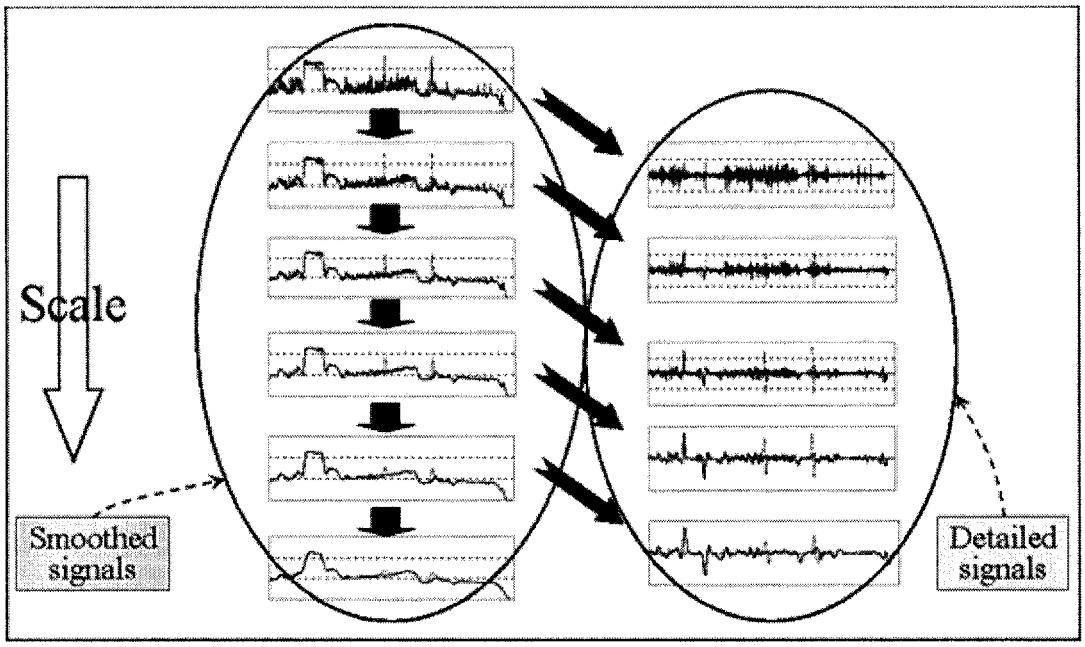


Figure 4: Multi-Scale Wavelet Representation of Real-Time Measurements

2.2 Identification of abnormal measurements

Abnormalities are defined as supernatural peeks of short duration. In other words, they are errors represented as large changes at high frequency. Such changes in real time measurements can be detected by using the first order WT, which is proportional to the first derivative of the smoothed signal (Equations 2 and 3). Since the extrema of the first derivative indicates fast changes in the function under study, one can detect fast changes in a set of measurements using the first order WT (Jiang *et al* 2003a).

$$\text{for first order WT: } \psi_j(t) = \frac{d\varphi_j(t)}{dt} \quad (2)$$

$$WT_j f(t) = f * \varphi_j(t) = f * (2^j \frac{d\varphi_j}{dt})(t) = 2^j \frac{d}{dt}(f * \varphi_j)(t) \quad (3)$$

2.3 Denoising based on multi-scale wavelet features and historical process data

Real time measurements contain noise at higher frequency than the true process trend. Therefore multi-scale wavelet decomposition of process measurements is effective for denoising if the correct scale is employed. Optimal scale choice for signal denoising is based on the response time, which cannot be predicted in real time and thus is not useful for real time data application. The technique proposed in this paper bases the scale choice on historical data, and assumes that the scale choice is time invariant.

As shown previously, choosing a WT scale that is too high creates a distortion of process measurements, and leads to an inaccurate reflection of process trend. On the other hand, choosing a scale that is too low will leave the smoothed signal with some random noise. This second possibility does not affect the true process trend. At the chosen scale, the process trend is still available but corrupted with higher frequency measurements. Therefore, in a subsequent step the corrupted smoothed signal can be refiltered to isolate the process trend from higher frequency perturbations. This opportunity is not possible in the case where scale is over-estimated.

One should choose a scale that does not affect the process trend. By studying historical process measurements, one can investigate the optimal cutting criterion for different process operation. The scale employed for on-line implementation can be selected in such a way that high frequency features are mostly deleted, and the true process trend is not affected by signal distortion.

3. Steady state detection based on WT features, filtering and statistical theory

3.1 Three-step methodology to perform steady state detection

Once abnormal measurements and high frequency features are removed, real time data steady state detection can be performed on-line. To do so precisely, a 3-step methodology based on wavelet transform and statistical theory is proposed, as follows:

4. The start point of the steady state period is detected by using WT characteristics and its first derivative,

5. High frequency features of the measured signal, which were not eliminated in the first step, are removed by filtering,
6. Finally, steady state period end point is detected through WT features analysis.

3.1.1. Step 1: Detecting steady state starting point with WT and the first derivative

The first order WT is proportional to the first derivative of the smoothed signal. Hence the WT measures variation in the smoothed signal and can be used to represent process variation (Jiang *et al* 2003a). WT extrema indicate fast changes in the data set, while near zero values indicate the presence of a slow change or a zero crossing point of the WT.

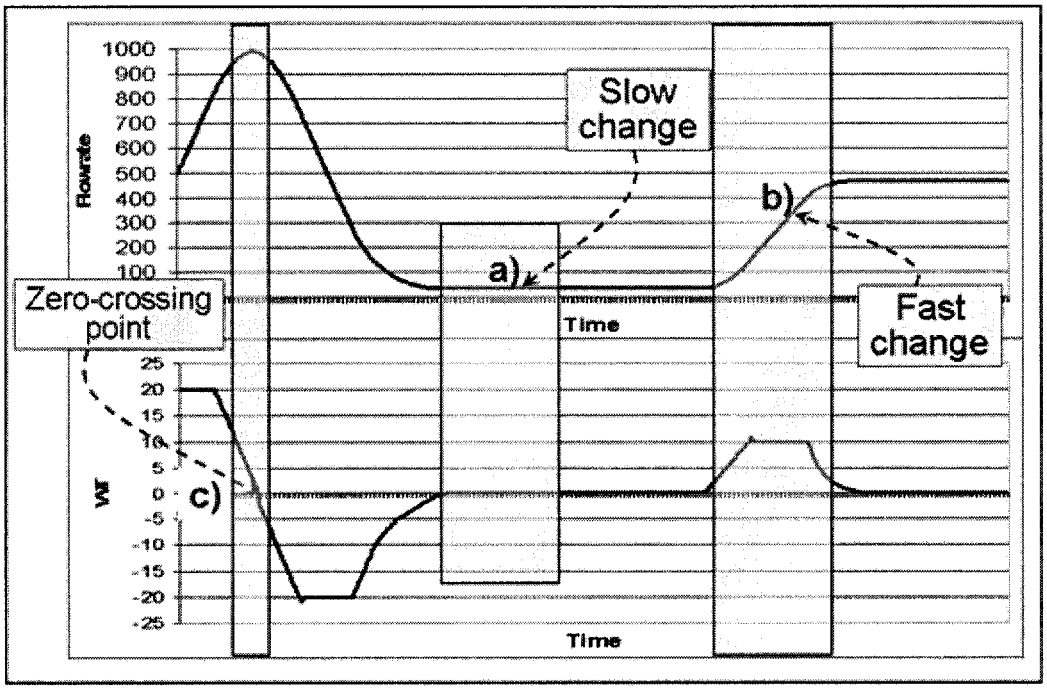


Figure 5: Using Wavelet Transform Features to Represent Process Trend

d) Slow change in process variation: 1st WT an its slope are near 0 values
e) Fast change in process variation: 1st WT value is not near 0
f) Zero-crossing point: 1st wavelet value is near zero but not the slope value

As shown in Figure 5, a zero crossing point corresponds to a peak in the process trend. Therefore the first derivative slope at that point is not equal to zero, as opposed to the case of steady state data. Based on the above information, the steady state starting point can be detected when the following equations are verified for the first time following a transient period:

$$\begin{cases} |WT(f)| < \alpha_1 \\ \left|\dfrac{dWT(f)}{dt}\right| < \alpha_2 \end{cases} \quad (4)$$

Figure 6 shows the methodology used to detect steady state starting point by taking advantage of wavelet transform features. Starting point is identified when steady is detected when the first order WT and its first derivative are simultaneously at near zero values. One can also notice that zero-crossing points are discarded by using this approach.

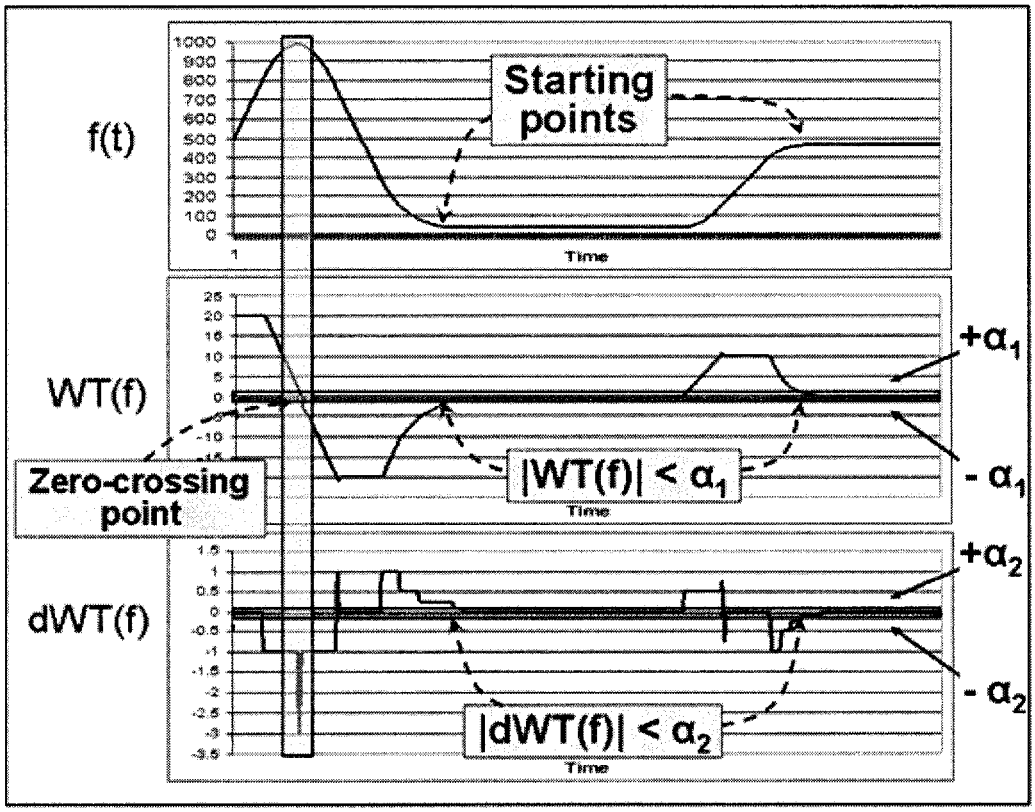


Figure 6: Detection of Steady State Starting Point Using WT and First Derivative

3.1.2. Step 2: Filtering steady state data

As discuss previously, over a long time period, the WT scale choice for high frequency features removal must be done carefully. Since the on-line scaling choice was done in order to insure that distortion is not incorporated in the signal, at certain points in time, some high frequency component will not be removed from process measurements. The oscillation created by this phenomenon can lead to false detection by the WT method. Figure 7 shows that when high frequency residuals exceed the minimal threshold acceptance, steady state identification is partial. Therefore, using the method described in the previous section is not suitable for on-line application over a long period when selecting the scale as proposed.

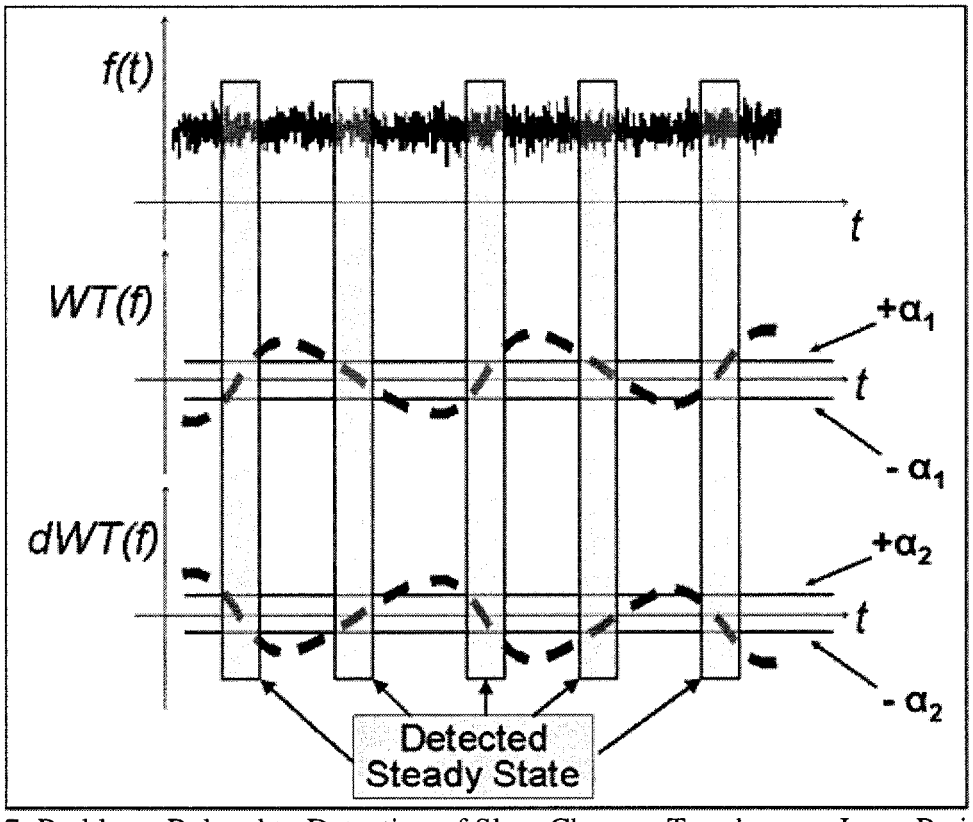


Figure 7: Problems Related to Detection of Slow Changes Trend over a Long Period

To overcome this problem, the steady state period can be approximated through a statistical test. An f-test on the ratio of the filtered squared deviation from the mean and the filtered squared difference of successive data is proposed. Before performing the test, an average value is approximated at each point in time by using a conventional moving average filter (Equation 5). Since this can create time delays and distortion in the corrected signal, such filters are not suitable for transient period correction, but are useful over steady state analysis.

$$\bar{x}_i = \beta_1 x_i + (1-\beta_1)\bar{x}_{i-1} \quad (5)$$

Figure 8 shows the basic concept behind the f-test proposed by Cao and Rhinehart (1995). "*A*" represents the deviation from the mean, while "*B*" represents the difference between two successive data. To avoid corruption by high frequency features that remains in the signal, the two estimators are filtered (Equations 6 and 7) and their ratio is calculated and compared to the threshold value of 1 (Equation 8). At this step of the analysis, real time measurements have already been treated through a wavelet processing method. Therefore, all features with frequencies higher than a certain threshold have been eliminated, and utilization of a low pass filter on the estimators does not affect the steady state method.

$$D_{1,i} = \beta_2 A_i + (1-\beta_2)D_{1,i-1} \quad \text{where } A_i = x_i - \bar{x}_{i-1} \quad (6)$$

$$D_{2,i} = \beta_3 B_i + (1-\beta_3)D_{21,i-1} \quad \text{where } B_i = x_i - x_{i-1} \quad (7)$$

$$\text{if } R_i = \frac{D_{1,i}}{D_{2,i}} < 1, \qquad \text{then the process is near steady state} \quad (8)$$

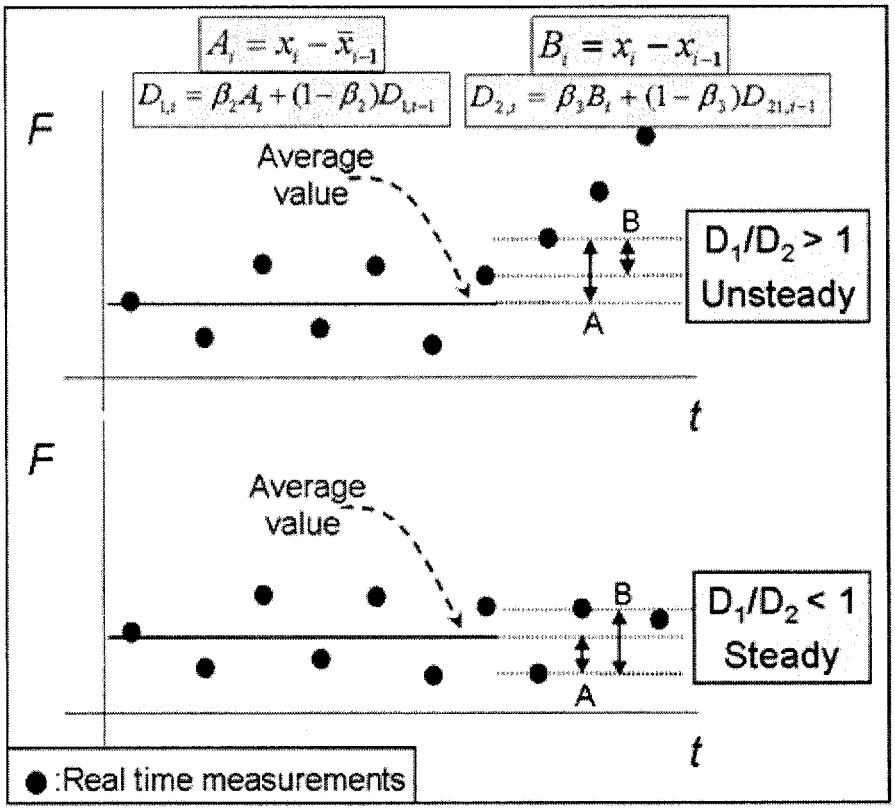


Figure 8: Representation of Deviation from Mean and Difference in Successive Data Points

Figure 9 shows, that as transient period evolves, distance between the measured value and filtered average (A_i) is increasing considerably. A low value of β_2 is selected in order to over-estimate the steady state period. In this way, D_1 and R variation will be minimized, even in the beginning of the transient period. In the third step of this methodology, the selected the over-estimation of the steady state period is corrected.

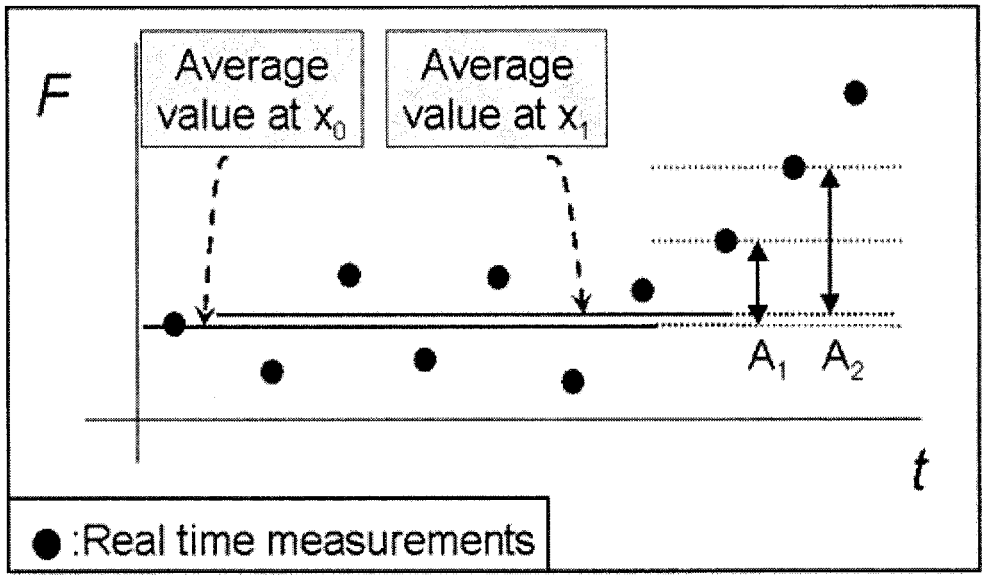


Figure 9 Evolution of Deviation from Average Value (A) During Transient Periods

3.1.3 Step 3: Detecting steady state period end point with WT

The final step of the methodology is to identify the end point of the steady state period. To do so, the last portion of the period identified in the previous step is analyzed. The WT of first order is use to detect the last moment before process status becomes transient. Since the WT value is proportional to the rate of change in the measurements one can detect, by analyzing the data backwards in time, the ending point of the steady state period by selecting the first value of WT under a certain threshold:

$$|WT(f)| < \alpha_3 \quad (9)$$

Since the selected set of data is already at steady state, no analysis related to the zero-crossing is needed here.

3.2 Selecting threshold values and filtering parameters for steady state detection

In order to take advantage of the proposed methodology, an accurate selection of appropriate threshold values and filtering parameters are critical. Since the selection of these parameters is done based on the degree of fluctuation of process measurements,

threshold values and filtering parameters can be selected *a priori* based on historic measurements.

3.2.1 Selection of threshold values (α_1, α_2 and α_3)

α_1 and α_2 are used to identify the steady starting point while α_3 is used in the detection of the period's ending point. More specifically, α_1 and α_2 are respectively used to ensure that WT modulus and its first derivative are at near zero values while α_3 is helpful to characterize process features in the WT modulus at the end of the steady state period.

The first two threshold values are used to detect steady state starting point. These are selected according to the degree of fluctuation in first order WT and its first derivative. To choose α_1 effectively, select adequate successive measurements under steady state, perform the first order WT and compute the standard deviation of the WT modulus (σ_{WT}) Jiang *et al* (2003a). The selection of α_2 is made to differentiate zero-crossing points from true steady state values. When comparing the true steady state WT modulus to the zero crossing point, one can notice (Figure 5) that the steady state values are associated with a slow change in the WT modulus. Therefore, one can perform the first derivative on the WT and compute its standard deviation ($\sigma_{\frac{dWT}{dt}}$). Then the threshold values are selected as follows:

$$\alpha_1 = \sigma_{WT} \quad (10)$$

$$\alpha_2 = \sigma_{\frac{dWT}{dt}} \quad (11)$$

Selection of α_3 is based on the WT features to identify significant changes in the process trend. To choose α_3 effectively, multiple sets of different historic steady state process measurements are selected, and the first order WT and its standard deviation

are computed. Starting with a λ value of 1, calculate the following threshold value and judge on the efficiency of the method on the selected data sets:

$$\alpha_3 = \lambda \sigma_{WT} \quad (12)$$

If the ending point of the steady state period is detected to late, then decrease the λ value (around 0.1) and repeat. If it is detected to early, increase λ value. That process is repeated until a suitable solution for all steady state set is achieved.

3.2.2. Selection of filtering parameters (β_1, β_2 and β_3)

Filtering parameters have a significant impact on the performance of the second step of the steady state detection. To avoid, as much as possible, undesired results due to unexpected on-line events, the investigation of the optimal filtering parameter must be performed simultaneously on different historic steady state periods. Optimal filtering parameters are obtained through the following steps (inspired by Bhat and Saraf 2004):

7. Select a value of β_3 (around 0.01)
8. Select a low value of β_2 (around 0.005)
9. Select a high value of β_1 (around 0.25)
10. Increment β_1 by .05 while the filtered average is affected by the high frequency component
11. Increment β_2 by .005 while every steady state period is over-estimate
12. Increment β_3 by 0.005 and go back to step 5 while all steady state period is detected.

4. Case Study : Application of Methodology to Paper Machine Process

The efficiency of the method described above was investigated on real time plant data from a papermachine stock approach system in an integrated pulp and paper mill. The plant is equipped with a process data management system (PI system from OsiSoft inc.) that samples tags approximately every 10 seconds. The papermachine under study has multiple process operation regimes associated with different characteristics in the process measurements and high frequency features of the signal.

4.1 Performance indices

In order to investigate the efficiency of the proposed methodology, it as been compared to two other methods used to detect steady state properly. The comparison was done by calculating, for each method, the type I (steady state is not detected) and type II (false detection) errors related to steady state detection.

The accuracy and precision of the steady state detection technique for on-line steady state data reconciliation was also evaluated. Accuracy is calculated for each steady state period, as the relative error between the true process value and the reconciled measurement. Precision is defined as the standard deviation of such errors. Therefore, low values for those two indices indicate high accuracy and precision.

Hypothetical real time true process signals were systematically corrupted with various types of errors. Since the performance indices used to investigate the efficiency of the proposed methodology are based on the true process data value, it was impossible to use directly process data to asses this task. Therefore, based on a systematic study of the process operations, experimental data were created to investigate the methodology performance.

4.2 Efficiency of the on-line steady state detection method

The proposed methodology for on-line steady state detection has been compared to the method based on first and second order WT and Cao's method. Optimal parameters were used for the WT based and Cao's method. WT based parameters were selected according to Jiang *et al*'s description (2003a), while optimum parameters for Cao's method were selected based on the method proposed by Bhat and Saraf (2004).

Figure 10 shows the quantity of type I error related to the identification of true steady state periods. One can notice that, for the system under study, the application of the proposed method reduces considerably the amount of type I error. That can be explained by the presence of a large amount of different high frequency features in the measurements:

3. Filtering the raw measurement variance by a low pass filter can lead to misidentification of process trend when high frequency features varies suddenly (figure 2).
4. Utilization of one the wavelet features in a conventional way, using a single cutting scale per variable, can lead to under-scaling or over-scaling problems when dealing with signal associated to multiple frequency features (figure 3).

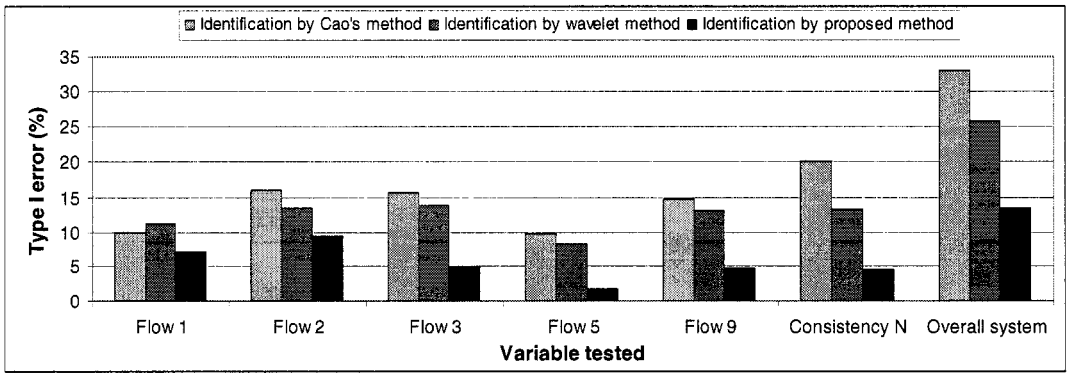


Figure 10: Type I Error Related to On-Line Steady State Detection

In order to reconciled the data in steady state, one needs to identify previously the overall system steady state periods. Figure 11 shows the quantity of type II error related to the identification of true steady state periods. One can notice that the effect of the chosen steady state identification method does not have a high impact on the type II error related to the overall system status recognition. Since every variable under study must be under near steady state to record the overall system steadiness, type II errors related to the each component are diluted in the process of identifying overall system status.

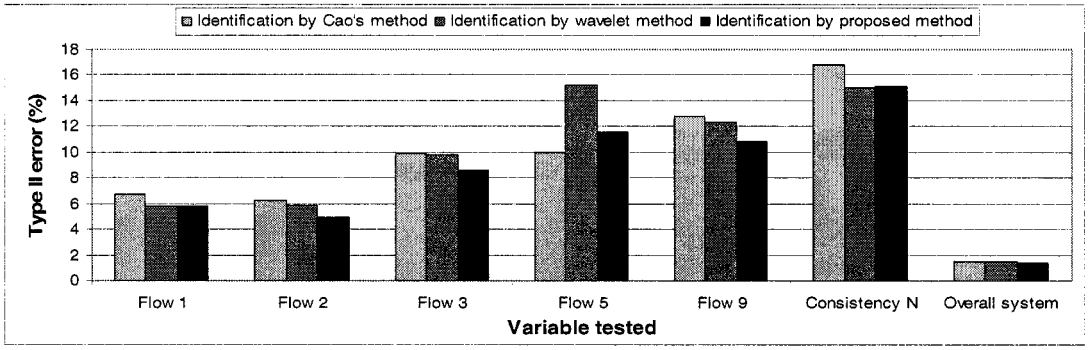


Figure 11: Type II Error Related to On-Line Steady State Detection

4.3 Impact on on-line steady state data reconciliation

To analyze the impact of steady state detection on the on-line reconciliation results, the results of the steady state analysis by the 3 different methods were used to reconcile the data. Experimental data were reconciled by applying the exact same reconciliation method to each data set.

4.3.1 Steady state data reconciliation overall methodology

To reconciled on-line steady state data, the overall methodology proposed by Bellec *et al.* (2004) was used. First, to correct and filter abnormal measurements and high frequency features, data acquired by the data management system are analyzed based on wavelet features. Then, steady state detection is performed. That step insures to

select adequate steady state data for steady state data reconciliation. Finally the data are reconciled by using Sigmafine software from OsiSoft Inc.

4.3.2 Steady state data reconciliation results

Figures 12 summarize the improvement of the average accuracy of the steady state data before and after the application of the different methodologies. Relative error is used as to represent accuracy in the data sets. Therefore low value on the figure indicates high accuracy. Equation 13 shows the average relative error has been calculated.

$$\%\mathrm{RE} = average\left(100\% * \frac{|\mathrm{Targeted.value} - \mathrm{measurement}|}{\mathrm{Targeted.value}}\right) \quad (13)$$

One can notice that the choice of steady state identification method as a direct impact on the quality of the steady state data reconciliation results. Generally, more adequate status identification leads to improvement in accuracy at the reconciliation step (see figure 1 for details). By analyzing Figure 12, one can notice that for some specific variables (cons N, level 8 and level 16) the relative error does not seem to be reduced by the utilization of the new method. Results on the graph for those variables are multiplied by a factor of 100. Therefore, we are dealing with small errors in the range of 0.05%. These small errors are more difficult to correct by using data reconciliation techniques.

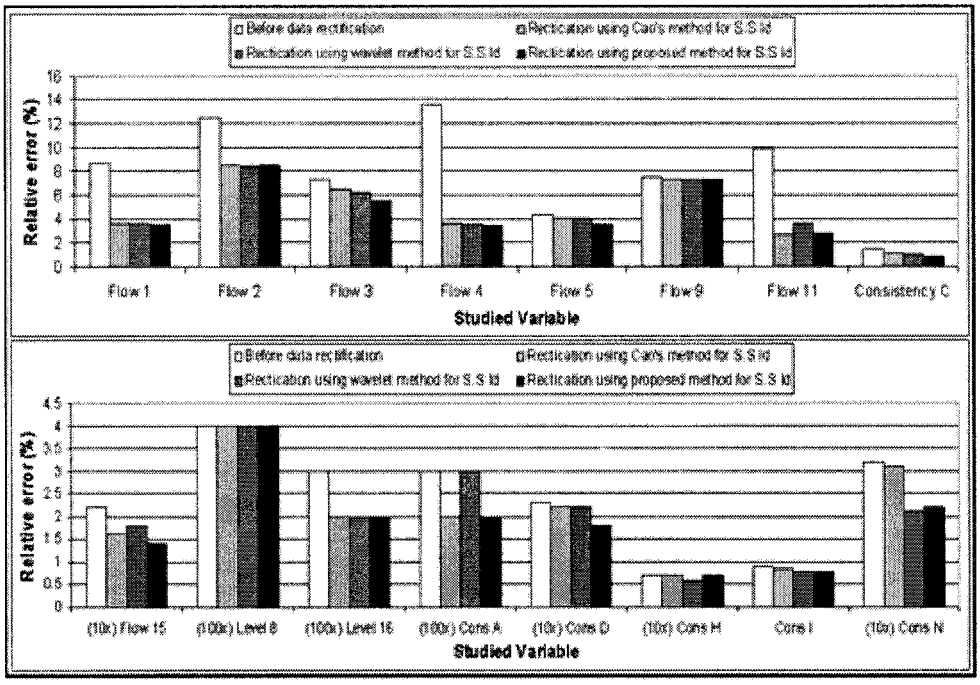


Figure 12: Accuracy Improvement by the Application
of Different On-Line Steady State Detection Techniques

Figure 13 summarize the improvement of the precision in the overall steady state data sets. Standard deviation to the average error is used to represent precision in the data sets. Therefore lower value on the figure indicates a higher precision. Equation 14 shows the standard deviation of relative error has been calculated. One can notice that the presence of gross errors (large amplitude errors on long period of time) will be reflected by a high standard deviation.

$$SD_{\%RE} = \sqrt{\frac{\sum_{i=1}^{n}\left[100\% * \left(average(RE) - RE_i\right)^2\right]}{n-1}} \quad (14)$$

By analyzing Figure 13, one can notice that the choice of the steady state identification method can have a huge impact on the precision of the reconciliation and therefore influence the correction of gross errors.

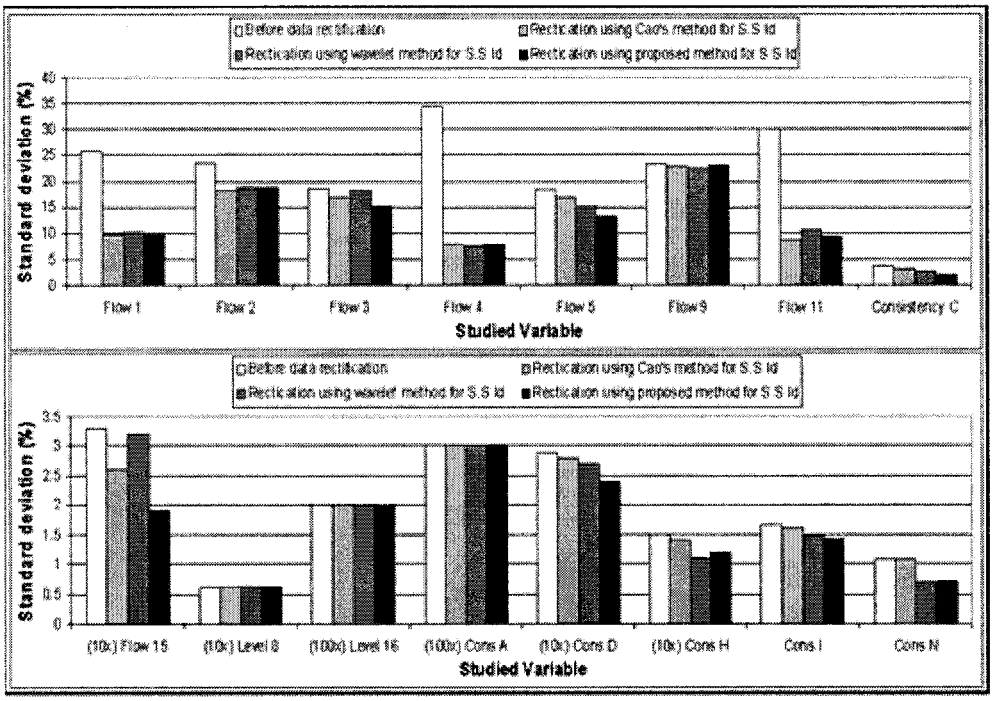


Figure 13: Precision Improvement by the Application
of Different On-Line Steady State Detection Techniques

5. Conclusions

A methodology to detect efficiently steady state status has been proposed and applied in this paper, which is suitable for on-line application. The method detects steady state periods using 3 signal processing steps, which allow reduction of Type I and Type II errors related to the identification of steady state periods. The method has been compared to other on-line steady state methods, using industrial data fro the stock approach system of a paper machine. Results show a higher accuracy improvement on real time measurements when using the proposed methodology. Using this approach for steady state detection in a general methodology to reconcile data on-line could potentially lead to multiple benefits for on-line process operations such as the following: efficient on-line identification of out-of-calibration instruments, improved process control due to improvement of measurement accuracy; and improved operations planning and optimization.

Acknowledgements

This work was completed with the financial assistance of the Natural Sciences and Engineering Research Council (NSERC) of Canada.

References

BAKSHI, B.R., STEPHANOPOULOS, G. (1993). Representation of Process Trends III. Multiscal Exraction of Trends from Process Data. *Computers and Chemicals Engineering*, Vol.18, pp 267-302.

BELLEC, S., JIANG, T., KERR, B., DIAMOND, M., STUART, P.R. (Submitted for publication December 2004). On-line Processing and Reconciliation of Pulp and Paper Mill Data, *Pulp & Paper Canada.*

BHAT A., SARAF N.,(2004). Steady-State Identification, Gross Error Detection and Data Reconciliation for Industrial Process Units. *Industrial and Engineering Chemistry*, Vol 43, pp 4323-4336.

CAO, S., RHINEHART, R.R.(1995). An Efficient Method for On-line Identification of Steady State. *Journal of Process Control*, Vol.5, No.6, pp 363-374.

DABROS, M., PERRIER, M., FORBES, F., FAIRBANK, M., STUART, P.R. (2005). Model-Based Direct Search Optimization of the Broke Recirculation System in a Newsprint Mill, accepted to *Journal of Cleaner Production.*

FLEHMIG F., WATZDORF, R.V., MARQUARDT, W. (1998). Identification of Trends in Process Measurements using the Wavelet Transform, *Computers and Chemicals Engineering*, Vol.22, pp 491-496.

HARISSON, R., LEROUX, R., STUART, P.R. (2003). Multivariate Analysis of Refiner Operating Data From a Thermo-Mechanical Pulp Newsprint Mill, *Pulp & Paper Canada*, 105:4, pp 24-27.

JANSSEN, M., STUART, P.R. (2004). Business Model Framework for Using Real-Time Process Data for Design Decision Making". Proceedings - *C-DEN Design Conference*, Montreal, QC. Paper 07B5.

JIANG, T., CHEN, B., He, X. (2000). Industrial Application of Wavelet Transform to the On-line Prediction of Side Draw Qualities of Crude Unit. *Computers and Chemical Engineering*, Vol.24, pp 507-512.

JIANG, T., CHEN, B., HE, X., STUART, P.R. (2003a). Application of a Steady State Detection Method Based on Wavelet Transform, *Computers and Chemical Engineering*, Vol. 27, pp 569-578.

JIANG, T., CHEN, B., JASMIN, K., STUART, P.R. (2003b), Strategy for Improving Data Quality for a Kraft Pulp Mill Recausticizing Plant, Proceedings - *FOCAPO Conference*, pp 467-470.

MALLAT, S., HWANG, W.L. (1992). Singularity Detection and Processing with Wavelets. *IEE Transactions on Information Theory*, Vol.38, pp 617-641.

NOUNOU, M., BAKSHI, B.R. (1999). On-line Multiscale Filtering of Random and Gross Errors without Process Models. *AIChE Journal*, 45:5, pp.1041-1058.