Validation et réconciliation de données. Approche conventionnelle, difficultés et développements

Benoît Marx, Gilles Mourot, Didier Maquin, José Ragot
Centre de Recherche en Automatique de Nancy. INPL/UHP/CNRS UMR 7039.
Institut National Polytechnique de Lorraine.
2, Avenue de la forêt de Haye. F - 54 516 Vandoeuvre les Nancy Cedex.
{benoit.marx, gilles.mourot, didier.maquin, josé.ragot}@ensem.inpl-nancy.fr

19 décembre 2006

Résumé

Résumé. On assiste, depuis une bonne dizaine d'années, à un développement considérable de méthodes concourant à l'efficacité du fonctionnement des systèmes technologiques. Ces méthodes s'appliquent lors de la conception d'un système mais aussi dans la phase d'exploitation de ce système. Dans ce cas, les méthodes visent d'une part à mieux connaître à chaque instant le mode de fonctionnement du système en question, et d'autre part, en cas de dysfonctionnement avéré ou de dérive par rapport au fonctionnement souhaité, à réagir sur les paramètres de contrôle du système. Cet exposé est focalisé sur le problème de la validation des mesures, les mesures validées servant précisément à mieux connaître l'état du système. La validation de donnée procède par comparaison des estimés d'une même grandeur et nécessite donc de disposer de redondances qui sont d'ordre matériel ou analytique. Dans cet exposé, trois approches sont considérées correspondant à trois niveaux de redondance : traitement de données par analyse en composantes principales, réconciliation de données à partir de modèles statiques, estimation d'état à partir de modèles dynamiques.

Mots-clefs : réconciliation de données, diagnostic, détection de défauts, estimation d'état, méthode avec modèle, analyse en composantes principales, robustesse.

1 Introduction

Pour fonctionner correctement, les systèmes de contrôle-commande et de supervision des installations industrielles ont besoin de recevoir, en permanence, des informations représentatives de leur état. L'élaboration de commandes complexes est en effet inefficace, si les informations prises en compte par les algorithmes qui les génèrent sont erronées et/ou incohérentes. La performance et la fiabilité de l'ensemble des moyens de commande et de contrôle sont liées à la qualité des systèmes de mesures. Toute défaillance de l'instrumentation conduit à la génération d'informations erronées. La validation de données qui permet de s'assurer de la cohérence des informations acquises constitue donc une étape essentielle qui doit précéder toute tentative de conduite rationnelle. L'étude de cette cohérence est délicate car les données sont le plus souvent hétérogènes (grandeurs physiques de natures différentes) et incomplètes (en milieu industriel, il est impossible, pour des raisons de coût ou des raisons technologiques, de disposer d'information sur chaque variable du système). De plus, les installations considérées sont souvent d'assez grande dimension et ont un comportement non linéaire.

La validation de données s'appuie sur la connaissance plus ou moins précise d'un modèle de comportement du système à surveiller ou d'une partie de celui-ci et a pour objet principal :

- d'élaborer, à partir de variables mesurées, des estimés cohérents avec ce modèle et des indicateurs globaux de bon fonctionnement ;
- d'enrichir la base de données en fournissant des estimés de grandeurs inaccessibles à la mesure ;
- de détecter et localiser, le plus précocement possible, l'apparition de défauts de mesure (cette étape peut éventuellement être suivie d'une caractérisation des défauts, c'est-à-dire de l'estimation de leur amplitude).

L'une des techniques permettant de s'assurer de la crédibilité d'une mesure consiste à créer une redondance d'informations ; celle-ci peut être obtenue en utilisant, par exemple, les relations structurellement exactes de bilan matière ou de bilan énergie. Ce type de redondance est qualifié d'analytique ou fonctionnelle, contrairement à la redondance matérielle obtenue en multipliant les capteurs mesurant une même grandeur.

La mise en évidence et l'extraction de redondance constitue le point de départ des méthodes de diagnostic. Quelques nuances peuvent être apportées selon que ces redondances sont explicitées ou non. Ainsi, lorsque on cherche à vérifier si un bilan de production est respecté (égalité de la somme des flux entrants et de la somme des flux sortants dans l'installation considérée), on a besoin de façon explicite des équations de conservation de la matière (ou de l'énergie) pour procéder à ce test ; le modèle du procédé est donc explicitement connu et pour cela nécessite une phase préalable de modélisation. Par contre, certaines techniques de traitement de données sont capables de détecter des données aberrantes et même des modifications de comportement d'un procédé sans modèle préalable explicite de ce procédé. C'est le cas, par exemple, des analyses en composantes principales (*ACP*) qui sont capables de mettre en évidence des hétérogénéités dans des flux de données. En réalité, l'examen attentif des techniques d'*ACP* montre l'existence d'un modèle sous-jacent prenant en compte l'existence de relations linéaires ou quasi-linéaires entre les différentes variables.

L'exposé est articulé autour de trois points. La section 2 traite de l'*ACP* et rappelle brièvement sa formulation classique et son application à la détection de données aberrantes. Quelques compléments donnent un aperçu sur des extensions de cette technique en abordant notamment les problèmes de robustesse. La section 3 est consacrée à la validation de données par réconciliation, cette technique étant présentée dans sa version de base puis étendue à des situations plus complexes. Enfin, la section 4 traite de l'estimation de l'état de fonctionnement des systèmes en évoquant des problèmes délicats d'excitations inconnues, de commutation de modes de fonctionnement, de systèmes décrits par plusieurs modèles.

On notera que les trois sections utilisent une connaissance de plus en plus précise du système que l'on cherche à surveiller : pour l'*ACP* le modèle du système n'est pas connu a priori, pour la réconciliation de données le modèle est sous forme statique, enfin les observateurs d'état font usage de modèles dynamiques.

2 Analyse en composantes principales et cohérence de données

Bien que généralement classée parmi les méthodes sans modèle, l'analyse en composantes principales élabore implicitement un modèle du système à partir de données expérimentales prélevées sur le système. L'*ACP* peut donc être considérée comme une méthode de diagnostic basée sur le concept de redondance analytique à part entière. Cette méthode permet :

– la mise en évidence de toutes les relations linéaires entre les variables du système sans les formuler explicitement. C'est un point important pour des systèmes de grande dimension dont les composants peuvent être fortement liés (degré de redondance élevé) ;
– la prise en compte de critères propres au diagnostic (détectabilité des défauts) lors de la synthèse du modèle [13]. Eventuellement, l'isolation des défauts peut être effectuée à partir d'*ACP* utilisant une partie adéquate des données.

2.1 Principe de base

Soit $\mathbf{x}(k) = [x_1(k)\ x_2(k)\ \dots\ x_m(k)]^T$ le vecteur contenant les m variables observées du système (mesures ou commandes) à l'instant k. Considérons la matrice de données $X = [\mathbf{x}(1)\ \mathbf{x}(2)\ \dots\ \mathbf{x}(N)]^T \in \Re^{N\times m}$ comprenant N observations $\mathbf{x}(k)$ recueillies sur ce processus en fonctionnement normal.

L'*ACP* détermine une transformation optimale (vis-à-vis d'un critère de variance) de la matrice de données X :

$$T = XP \quad \text{et} \quad X = TP^T \tag{1}$$

avec $T = [t_1\ t_2\ \dots\ t_m] \in \Re^{N\times m}$, où les t_i sont les composantes principales et la matrice $P = [p_1\ p_2\ \dots\ p_m] \in \Re^{m\times m}$, où les vecteurs orthogonaux p_i sont les vecteurs propres correspondant aux valeurs propres λ_i de la décomposition en valeurs et vecteurs propres de la matrice de covariance (ou de corrélation) Σ de X :

$$\Sigma = P\Lambda P^T \qquad \text{avec} \qquad PP^T = P^TP = I_m \tag{2}$$

avec Λ une matrice diagonale où les termes diagonaux sont ordonnés dans l'ordre décroissant. Les valeurs propres les plus petites par rapport aux autres indiquent l'existence de relations linéaires ou quasi-linéaires entre les différentes composantes de $\mathbf{X}$.

Pour une valeur de l'entier ℓ donnée, les matrices des vecteurs propres et des composantes principales sont partitionnées sous la forme :

$$P = \left[\ \hat{P}_\ell \ \middle|\ \tilde{P}_{m-\ell}\ \right], \quad T = \left[\ \hat{T}_\ell \ \middle|\ \tilde{T}_{m-\ell}\ \right] \tag{3}$$

A partir de l'équation (1), on peut alors expliciter la partie $\hat{X}$ des données expliquées uniquement par les ℓ premiers vecteurs propres et la partie résiduelle E expliquée par les composantes restantes :

$$\hat{X} = X\hat{P}_\ell\hat{P}_\ell^T = X\hat{C}_\ell \tag{4}$$

$$E = X - \hat{X} = X(I - \hat{C}_\ell) \tag{5}$$

où l'on notera que la matrice $\hat{C}_\ell = \hat{P}_\ell\hat{P}_\ell^T$ n'est pas égale à la matrice identité.

Détermination du nombre de composantes ℓ

Qin et Dunia (2000) ont proposé de déterminer la valeur du nombre ℓ de composantes à retenir par minimisation de la variance de l'erreur de reconstruction. La reconstruction consiste à estimer une variable à l'aide du modèle

ACP et des autres variables, i.e. à partir des relations de redondance existant entre cette variable et les autres.

Rappelons comment reconstruire une variable. Soit $\mathbf{x}_j(k) = [x_1(k) \dots z_j(k) \dots x_m(k)]^T$ le vecteur de mesure $\mathbf{x}(k)$ à l'instant k dont la $j^{\text{ème}}$ composante est reconstruite de la façon suivante [12] :

$$\mathbf{x}_j(k) = G_j\,\mathbf{x}(k) \tag{6}$$

et

$$G_j^T = [\ \xi_1 \cdots g_j \cdots \xi_m\ ]\ ,\ g_j^T = \frac{\left[c_{-j}^T\ 0\ c_{+j}^T\right]}{1 - c_{jj}} \tag{7}$$

où $\xi_j = [0 \dots 1 \dots 0]^T$ est la $j^{\text{ème}}$ colonne de la matrice identité, c_j est la $j^{\text{ème}}$ colonne de $\hat{C}_\ell$ et les indices $+j$ et $-j$ désignent, respectivement, les vecteurs formés par les $(j-1)$ premiers et les $(m-j)$ derniers éléments du vecteur c_j.

La variance de l'erreur de reconstruction de la $j^{\text{ème}}$ composante de $\mathbf{x}(k)$ est donnée par :

$$\rho_j(\ell) = \frac{\tilde{\xi}_j^T \Sigma \tilde{\xi}_j}{\left(\tilde{\xi}_j^T \tilde{\xi}_j\right)^2} \tag{8}$$

où $\tilde{\xi}_j = \left(I - \hat{C}_\ell\right)\xi_j$.

Le nombre de composantes principales à retenir s'obtient en minimisant par rapport à ℓ le critère :

$$J(\ell) = \sum_{j=1}^{m} \frac{\rho_j(\ell)}{\xi_j^T \Sigma \xi_j} \qquad \ell = 1, \dots, m-1 \tag{9}$$

les contributions des variables au critère étant pondérées par leurs variances.

Ayant défini le modèle *ACP* , et en particulier le nombre de composantes principales significatives à retenir, nous examinons maintenant son utilisation pour la détection et la localisation de défauts de capteurs.

Génération de résidus indicateurs de défauts

Considérons un nouveau vecteur de mesure $\mathbf{x}(k)$ dont on souhaite éprouver la consistance. Evaluons le vecteur des dernières composantes principales (1) :

$$\tilde{\mathbf{t}}_{m-\ell}(k) = \tilde{P}_{m-\ell}^T\,\mathbf{x}(k) \tag{10}$$

Soient $\mathbf{x^o}(k)$ le vecteur des valeurs vraies, $\epsilon(k)$ le vecteur des bruits de mesure supposé blanc et ξ_f la direction du défaut. En présence d'un défaut d'amplitude $d(k)$ quelconque, on peut expliciter le vecteur de mesure sous la forme :

$$\mathbf{x}(k) = \mathbf{x^o}(k) + \epsilon(k) + \xi_f d(k) \tag{11}$$

Le vecteur des résidus (10) s'écrit alors :

$$\tilde{\mathbf{t}}_{m-\ell}(k) \;=\; \underbrace{\tilde{P}_{m-\ell}^T\,\mathbf{x^o}(k)}_{=0} + \tilde{P}_{m-\ell}^T\,\epsilon(k) + \tilde{P}_{m-\ell}^T\,\xi_f d(k) \tag{12}$$

En absence de défauts, l'espérance mathématique du résidu est nulle. Par contre, en présence de défauts, l'espérance mathématique du résidu n'est plus nulle et le défaut affecte toutes les composantes du vecteur des résidus :

$$Esp(\tilde{\mathbf{t}}_{m-\ell}(k)) = \tilde{P}_{m-\ell}^T\,\xi_f d(k)$$

On peut noter le rôle important joué par la matrice $\tilde{P}_{m-\ell}^T\,\xi_f$ qui indique la façon dont les résidus sont influencés par le défaut considéré. L'examen du rang de cette matrice et de ses colonnes est révélateur de l'influence de ce défaut.

Localisation de défauts par reconstruction

Pour localiser la ou les variables du vecteur $x(k)$ qui sont en cause, c'est-à-dire retrouver la direction ξ_f, nous allons tout à tour reconstruire chaque variable de ce vecteur à partir du modèle ACP établi précédemment et des autres variables [13],[17], [18].

Etudions la propagation d'un défaut sur le vecteur des reconstructions $\mathbf{x}_j(k)$ dans la direction j, c'est-à-dire :

$$\mathbf{x}_j(k) = G_j(\mathbf{x^o}(k) + \epsilon(k)) + G_j\,\xi_f d(k) \tag{13}$$

qui se réduit naturellement à :

$$\mathbf{x}_j(k) = G_j(\epsilon(k)) + G_j\,\xi_f d(k) \tag{14}$$

Si la variable en défaut est la variable reconstruite $(j = f)$ alors, compte tenu de la définition de G_j, on peut montrer que :

$$G_j\,\xi_f = 0 \tag{15}$$

A partir de l'équation (15), on constate que si la reconstruction se fait dans la direction du défaut (c'est-à-dire quand $j = f$), l'effet du défaut est éliminé sur $\mathbf{x}_j$ puisque :

$$Esp(\mathbf{x}_j(k)) = G_j\,\xi_f d(k) = 0$$

Ainsi, l'analyse des amplitudes des différents résidus obtenus par reconstruction dans toutes les directions (j=1..m) est révélatrice de la présence de défauts et permet de déterminer la composante de la donnée affectée par ce défaut.

Remarque – Cette méthode peut être utilisée pour la localisation de défauts multiples en reconstruisant simultanément les variables supposées en défauts [13].

2.2 Extension : ACP robuste

L'*ACP* est essentiellement basée sur la mise en évidence de relations linéaires ou quasi-linéaires entre les variables et présente un caractère d'optimalité uniquement au sens d'un critère portant sur l'erreur quadratique d'estimation en valeur moyenne (*MSE*). Il est bien connu que l'estimation basée sur l'utilisation de critère de type *MSE* est moins robuste aux valeurs aberrantes que celle issue d'autres critères comme celui de l'erreur en valeur absolue.

Pour préciser ce point, rappelons que l'approche classique de l'*ACP* utilise un calcul préliminaire de la moyenne des données et de leur matrice de covariance ; la moyenne et la variance sont sensibles à la présence de valeurs aberrantes, et les résultats obtenus s'avèrent souvent inexploitables car trop biaisés par l'influence de ces valeurs aberrantes. Afin de réduire la sensibilité de l'*ACP* aux valeurs aberrantes, nous avons privilégié une technique consistant à réaliser l'*ACP* directement sur les données éventuellement contaminées par des valeurs aberrantes en recherchant des directions principales insensibles à ces valeurs aberrantes. Dans [6], les auteurs définissent une matrice de covariance "locale" qui tend à privilégier la contribution d'observations proches au détriment d'observations éloignées, l'éloignement étant dû à la présence de valeurs aberrantes. Cette matrice est définie de la façon suivante :

$$\tilde{\Sigma} = \frac{\displaystyle\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} w_{i,j} \left(\mathbf{x}(i) - \mathbf{x}(j)\right) \left(\mathbf{x}(i) - \mathbf{x}(j)\right)^T}{\displaystyle\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} w_{i,j}} \tag{16}$$

où les poids $w_{i,j}$ sont eux-mêmes définis par :

$$w_{i,j} = \exp\left(-\frac{\beta}{2} \left(\mathbf{x}(i) - \mathbf{x}(j)\right)^T \Sigma^{-1} \left(\mathbf{x}(i) - \mathbf{x}(j)\right)\right) \tag{17}$$

β étant un paramètre à régler pour obtenir effectivement une réduction de l'influence des observations éloignées. l'ACP peut alors être conduite sur cette "nouvelle" matrice de covariance réputée robuste vis-à-vis des valeurs aberrantes grâce à la présence de poids adaptés $w_{i,j}$.

3 Réconciliation des données

3.1 Principe de base

La réconciliation de données a pour but de rendre compatibles les mesures effectuées sur un système avec son modèle. A ce titre, les méthodes de réconciliation se rapprochent parfaitement des méthodes d'estimation d'état. Une conséquence importante de la réconciliation est la détection de valeurs de mesures aberrantes. En effet, les valeurs réconciliées peuvent être comparés aux mesures ; les écarts constatés peuvent être analysés, les plus grands d'entre eux pouvant témoigner de la présence de mesures aberrantes.

La validation de données peut être effectuée sur la base de modèles plus ou moins complexes selon la connaissance dont on dispose sur une installation. Comme nous l'avons mentionné précédemment, l'établissement de bilans matière élémentaires sur des durées d'observation suffisamment importantes conduit à des modèles statiques linéaires. En pratique, dans ce cas, les grandeurs sujettes à validation sont des cumuls ou des moyennes des mesures des grandeurs physiques sur une fenêtre d'observation temporelle. La prise en compte de bilans énergie ou de bilans par espèces chimiques ou minérales, par exemple, nécessite l'usage de modèles non linéaires. La validation de données peut également être effectuée, à chaque instant, en utilisant des modèles dynamiques ; en fait, les méthodes développées dans le cadre des modèles statiques peuvent être ré-utilisées pour les modèles dynamiques. Pour cette raison, les principes généraux de la validation de données sont exposés pour le cas des modèles statiques. On considère donc un système pouvant être décrit par :

– un ensemble de contraintes statiques représenté par une fonction vectorielle (linéaire ou non linéaire) d'un vecteur d'état inconnu :

$$F(X^*) = 0 \tag{18}$$

où $F(X^*)$ est de dimension n et X^*, le vecteur d'état du système, est de dimension v ;

– une équation d'observation liant un vecteur de mesure Z aux états :

$$Z = H(X^*) + \varepsilon \tag{19}$$

où Z est de dimension m, $H(X^*)$ caractérise le système de mesure et ε, de dimension m, est un vecteur aléatoire d'erreurs de mesure caractérisées par des lois de distributions de probabilité connues.

Historiquement, la loi de Laplace-Gauss a précisément été élaborée à partir de l'observation statistique d'erreurs de mesure.

Pour une variable aléatoire continue scalaire, la densité de probabilité d'une telle loi s'écrit :

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) \tag{20}$$

où μ et σ représentent respectivement l'espérance et l'écart-type de la distribution. Cette expression se généralise dans le cas d'une variable aléatoire vectorielle $X \in \mathbb{R}^m$ sous la forme :

$$f(X) = \frac{1}{(2\pi)^{m/2}|V|^{1/2}} \exp\left(-\frac{1}{2}\|X - \mathrm{Esp}(X)\|^2_{V^{-1}}\right) \tag{21}$$

où, par définition, l'écriture $\|U\|_T^2$ correspond à U^TTU.

Pour un échantillon de N observations indépendantes $Z_i = H(X^*) + \varepsilon_i$, pour $i = 1, \ldots, N$, la fonction de vraisemblance s'écrit comme le produit des densités de probabilité des erreurs :

$$v = \frac{1}{(2\pi)^{Nm/2}|V|^{k/2}} \prod_{i=1}^{N} \exp\left(-\frac{1}{2}\|Z_i - H(X^*)\|_{V^{-1}}^2\right) \tag{22}$$

Au sens du maximum de vraisemblance, le meilleur estimé $\hat{X}$ est celui qui maximise la fonction de vraisemblance tout en respectant les contraintes du modèle $F(\hat{X}) = 0$.

Compte tenu de la forme de cette fonction de vraisemblance et du fait que la fonction logarithme est une fonction monotone, on peut maximiser le logarithme de la fonction de vraisemblance qui possède son maximum pour les mêmes valeurs d'argument que la fonction de vraisemblance :

$$\ln(v) = -\frac{N}{2}(\ln(2\pi^m|V|)) - \frac{1}{2}\sum_{i=1}^{N}\|Z_i - H(X^*)\|_{V^{-1}}^2 \tag{23}$$

En supprimant le terme constant, le problème d'estimation revient donc à chercher le minimum, par rapport à X^*, du critère :

$$\phi = \frac{N}{2}\ln(|V|) + \frac{1}{2}\sum_{i=1}^{N}\|Z_i - H(X^*)\|_{V^{-1}}^2 \tag{24}$$

sous les contraintes $F(X^*) = 0$

Selon le degré de connaissance de la matrice de variance-covariance V des erreurs de mesure, la solution peut être explicitée davantage. L'hypothèse la plus communément employée consiste à supposer que la matrice de variance des erreurs de mesures est connue. En effet, celle-ci peut être déduite de la connaissance des précisions avec lesquelles les mesures sont effectuées. Fréquemment, cette matrice est également supposée diagonale, c'est-à-dire que l'on formule l'hypothèse que les erreurs de mesure sont statistiquement indépendantes. Le problème précédent [24] se ramène alors à la recherche du minimum, par rapport à la grandeur inconnue X^*, du critère :

$$\phi = \frac{1}{2}\sum_{i=1}^{N}\|(Z_i - H(X^*)\|_{V^{-1}}^2 \tag{25}$$

sous les contraintes $F(X^*) = 0$

Il s'agit donc d'un problème classique d'optimisation sous contraintes égalité [37]. La difficulté de résolution de ce problème dépend étroitement de la structure et de la dimension des équations du modèle, de la structure des contraintes et du nombre d'observations. Dans le cas de contraintes et d'équations d'observation linéaires, les solutions sont analytiques [40]. Dans tous les autres cas, il faut utiliser les techniques de calcul hiérarchisé, de linéarisation, de changement de variables ou même des solutions approchées. De nombreux travaux présentent ces techniques, par exemple [44], [27].

3.2 Extension : réconciliation robuste

De nombreux développements complètent le principe de base que nous venons de rappeler. Ils sont nés de la nécessité de pouvoir appliquer ce principe à des situations et des données réelles, situations qui ne respectent pas toujours les hypothèses de base précédemment utilisées. C'est ainsi que des extensions ont permis de s'intéresser aux systèmes dynamiques [10], à la présence d'erreurs de mesure bornées [35], [39], [38], à la présence de paramètres mal connus [4], à l'estimation de la précision des mesures [26], à la réconciliation des données de façon simultanée à l'estimation des paramètres du système [31], à la localisation de défauts de mesure [33] [34], à la conception de réseaux de capteurs pour la surveillance des systèmes [24],[5].

Dans ce paragraphe, le point particulier de la robustesse de la réconciliation vis-à-vis des mesures aberrantes est abordé. Pour introduire cette problématique, rappelons que la réconciliation repose sur la minimisation d'un critère formé à partir des écarts pondérés entre estimés et mesures, les pondérations étant proportionnnelles aux variances des erreurs de mesure. La validité et le caractère optimal de cette approche sont éminemment liés à l'hypothèse forte de normalité des erreurs de mesure. Dans la pratique, cette hypothèse peut être mise en défaut en présence de grosses erreurs qui constituent des valeurs aberrantes. On apporte donc plus de réalisme en posant le problème de réconciliation de la façon suivante. A partir de mesures $\tilde{x}_i$ estimer les grandeurs vraies x_i^* d'un système à modèle linéaire, sachant que :

$$\tilde{x}_i = x_i^* + \epsilon_i + g_i, \qquad i = 1..v \tag{26a}$$

$$\sum_{j=1}^{v} a_{i,j}x_j^* = 0, \qquad i = 1..n \tag{26b}$$

les mesures étant affectées par des erreurs de faibles et de grosses amplitudes ϵ_i et g_i, ces dernières n'étant pas connues a priori. Pour prendre ces erreurs en compte, plusieurs approches ont été envisagées.

Partant de l'hypothèse que le nombre de grosses erreurs est faible, la première technique procède par réconciliation des mesures par moindres carrés (donc avec l'hypothèse de normalité des erreurs), puis détecte et localise les grosses erreurs (analyse des termes correctifs), et enfin reconduit la procédure de réconciliation en affectant une variance très grande aux mesures pour lesquelles des grosses erreurs ont été localisées. Le défaut majeur de cette a‘pproche réside dans la première réconciliation qui peut s'avérer fortement erronnée en raison de la présence des grosses erreurs ; ceci peut rendre ensuite difficile la localisation des grosses erreurs en analysant les termes correctifs.

La seconde approche cherche à prendre en compte directement la présence des grosses erreurs au moyen d'une loi de distribution plus appropriée. La classe des *M-estimateurs* fournit des estimées robustes vis-à-vis des grosses erreurs. Le critère d'estimation est pris sous la forme :

$$\Phi_r = -\sum_{i=1}^{N} \log\left(f(\epsilon_i)\right) \tag{27}$$

où f est habituellement une distribution dite *contaminée*, c'est-à-dire prenant en compte la présence des erreurs de faibles amplitudes et les grosses erreurs [41]. Par exemple, avec des distributions normales :

$$f(\epsilon_i) = \mu p_{1,i}(\epsilon_i) + (1-\mu)p_{2,i}(\epsilon_i) \tag{28}$$

$$p_{1,i}(\epsilon_i) = \frac{1}{\sqrt{2\pi}\sigma_{1,i}} \exp\left(-\frac{1}{2}\left(\frac{x_i^* - \tilde{x}_i}{\sigma_{1,i}}\right)^2\right) \tag{29}$$

$$p_{2,i}(\epsilon_i) = \frac{1}{\sqrt{2\pi}\sigma_{2,i}} \exp\left(-\frac{1}{2}\left(\frac{x_i^* - \tilde{x}_i}{\sigma_{2,i}}\right)^2\right) \tag{30}$$

Les écart-types $\sigma_{2,i}$ seront choisis plus grands que $\sigma_{1,i}$ afin de marquer la disparité entre les deux types d'erreur. Dans la pratique, on pourra choisir une dépendance linéaire entre les écart-types des deux distributions $\sigma_{2,i} = A.\sigma_{1,i}$. Le coefficient μ, encore appelé paramètre de mélange traduit la proportion de grosses erreurs dans les mesures. C'est un coefficient connu a priori ou dans le cas contraire qui peut être estimé.

Finalement, avec le critère (27) et compte tenu des définitions (28, 29, 30) et des contraintes (26), la réconciliation des données résulte donc d'une procédure d'optimisation sous contraintes. La forme du critère à optimiser limite ici les approches analytiques qui doivent être substituées par des approches numériques souvent itératives [41], [42].

3.3 Applications

La réconciliation de mesures est un moyen largement répandu et appliqué au niveau industriel pour mettre en évidence la cohérence ou l'incohérence des mesures, et dans le deuxième cas, pour corriger ces incohérences.

Dans l'article [3], les auteurs présentent un exemple d'application de la réconciliation de données à des mesures fournies par les capteurs installés sur un four de durcissement de boulettes d'oxyde de fer. Les résultats montrent que la technique utilisée permet d'estimer des débits de gaz dans des conduites de plusieurs mètres de diamètre à partir de mesures de pression et de température, en plus de permettre d'identifier des mesures imprécises ou biaisées de capteurs. Les résultats peuvent aussi être utilisés pour le développement de capteurs virtuels pour relier la puissance de ventilateurs au débit de gaz poussé par ces derniers.

On trouve des applications semblables dans le domaine de l'industrie papetière [19], [5], de l'industrie nucléaire [11], de l'industrie minérale [2], [25], de l'hydraulique [10], [31], [28], [43], de l'énergie [21], de la sidérurgie [49], [38], du génie chimique [34], [39], [44]. Toutes ces applications mettent en évidence l'intérêt des procédures de réconciliation de données qui fournissent une critique des données et des mesures, qui restituent à l'utilisateur des données cohérentes pouvant ensuite être utilisées dans des procédures de contrôle des processus.

4 Diagnostic à base d'observateurs

Cette section présente quelques méthodes de diagnostic de systèmes exploitables lorsque l'opérateur dispose d'un modèle du système à superviser. Ce modèle est un ensemble de relations mathématiques faisant intervenir les variables d'entrée et de sortie du système, ainsi que des variables internes caractéristiques de l'état du système. Un tel modèle est obtenu, soit à partir des lois de comportement des éléments constitutifs du système, soit par identification à partir de la mesure en fonctionnement des signaux d'entrée et de sortie.

Une fois le modèle du système établi, il est possible de l'utiliser pour avoir une meilleure connaissance de l'état dans lequel est le système. Généralement l'opérateur a accès aux mesures des grandeurs d'entrée et de sortie, en revanche, les grandeurs internes peuvent être inaccessibles à la mesure. Afin de pallier ce manque, on peut chercher à construire un filtre, qui permet de déduire la valeur des grandeurs internes à partir des grandeurs mesurées, on parle alors d'*observateur*.

Le principe général du diagnostic à base d'observateurs est que le modèle du système représente son fonctionnement sain, c'est à dire exempt de tout défaut. Autrement dit, l'estimation obtenue en utilisant un observateur donne les valeurs des grandeurs internes en l'absence de défaut et l'estimation des variables d'état permet, à l'aide du modèle, d'estimer les variables de sorties. Il est alors possible de comparer les sorties mesurées et les sorties estimées. L'éventuel écart entre les grandeurs mesurées et estimées met en évidence un écart entre le fonctionnement réel et le fonctionnement sain du système. En d'autres termes, cet écart peut être considéré comme un *résidu*, c'est à dire un indicateur de l'occurrence d'un dysfonctionnement.

Les dysfonctionnements affectant un système sont souvent modélisés par des signaux inconnus qui peuvent être de deux natures. D'une part les signaux de *défauts*, qui représentent le mauvais fonctionnement d'un composant, par exemple la déviation d'un capteur ou la variation d'un paramètre due à l'usure mécanique d'une pièce. D'autre part, des signaux de *perturbations* induits par les différentes sources d'incertitudes, par exemple des erreurs commises lors de la modélisation ou des imprécisions de mesures. L'objectif du diagnostic est de pouvoir mettre en évidence la présence des défauts malgré la présence des perturbations.

Dans la suite, nous détaillerons ces techniques dans le cas des systèmes décrits par un modèle dynamique linéaire. Malgré l'apparente simplicité de cette classe de modèles, cette approche permet de décrire une grande variété de processus. Par la suite, le diagnostic d'autres classes de modèles sera évoqué.

4.1 Diagnostic de systèmes linéaires

Synthèse d'observateur

Dans cette partie, il est supposé que le processus à surveiller peut être décrit par un modèle dynamique linéaire défini par :

$$\begin{cases} \frac{dx(t)}{dt} = Ax(t) + Bu(t) \\ y(t) = Cx(t) \\ x(0) = x_0 \end{cases} \tag{31}$$

où $u(t) \in \mathbb{R}^p$ est le vecteur des entrées de commande du système, $x(t) \in \mathbb{R}^n$ est le vecteur des variables internes du système, aussi appelé vecteur d'état, et $y(t) \in \mathbb{R}^m$ est le vecteur des sorties mesurées. Les matrices $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times p}$ et $C \in \mathbb{R}^{m \times n}$ sont des matrices constantes et connues.

L'observateur est un filtre actif qui permet de reconstruire l'état $x(t)$ à partir des grandeurs connues $u(t)$ et $y(t)$. Les estimées de l'état et des sorties sont notées respectivement $\hat{x}(t)$ et $\hat{y}(t)$ et sont définies par le système suivant :

$$\begin{cases} \frac{d\hat{x}(t)}{dt} = A\hat{x}(t) + Bu(t) + K(y(t) - \hat{y}(t)) \\ \hat{y}(t) = C\hat{x}(t) \\ \hat{x}(0) = \hat{x}_0 \end{cases} \tag{32}$$

Cette structure, proposée par [22], peut être intuitivement perçue comme un système évoluant en parallèle selon le modèle connu du système et qui, de plus, comporte un bouclage sur l'erreur de sortie (terme $K(y(t) - \hat{y}(t))$ afin de corriger l'erreur d'estimation de l'état. Cela se vérifie analytiquement, en étudiant l'évolution de l'erreur d'estimation de l'état, $\tilde{x}(t) = x(t) - \hat{x}(t)$. Cette erreur obéit à l'équation suivante, obtenue avec (31) et (32) :

$$\frac{d\tilde{x}(t)}{dt} = (A - KC)\tilde{x}(t) \tag{33}$$

Si la matrice K est choisie telle que $A - KC$ est stable, alors l'erreur d'estimation converge vers 0 quelles que soient les valeurs x_0 (inconnue) et $\hat{x}_0$ (fixée arbitrairement). Autrement dit, même si la valeur initiale de l'état est inconnue, l'observateur donnera une estimation juste de l'état, après un régime transitoire court dépendant du choix de K.

Application au diagnostic

Afin d'illustrer l'utilisation d'un observateur pour le diagnostic, on considère à présent un modèle linéaire décrivant un processus affecté de défauts de capteur $f_c(t)$ définis par :

$$\begin{cases} \frac{dx(t)}{dt} = Ax(t) + Bu(t) \\ y(t) = Cx(t) + C_c f_c(t) \\ x(0) = x_0 \end{cases} \tag{34}$$

Dans ce cas l'erreur d'estimation de l'état $\tilde{x}(t)$ et de la sortie $\tilde{y}(t) = y(t) - \hat{y}(t)$ sont définies par :

$$\begin{cases} \frac{d\tilde{x}(t)}{dt} = (A - KC)\tilde{x}(t) - KC_c f_c(t) \\ \tilde{y}(t) = C\tilde{x}(t) + C_c f_c(t) \\ \tilde{x}(0) = x_0 - \hat{x}_0 \end{cases} \tag{35}$$

En l'absence de défaut, les signaux $\tilde{x}(t)$ et $\tilde{y}(t)$ tendent asymptotiquement vers 0. En revanche, à l'apparition de $f_c(t)$, les erreurs d'estimation de l'état ou des sorties vont s'écarter de 0. Les variables internes étant généralement inconnues de l'opérateur, il est donc impossible de former l'écart $\tilde{x}(t)$; néanmoins il est possible de calculer $\tilde{y}(t)$.

Une valeur de $\tilde{y}(t)$ significativement différente de 0 est un indicateur de l'occurrence d'un défaut, autrement dit permet la *détection* de défaut. Pour obtenir une information plus précise, en particulier être capable de savoir quel défaut est apparu, il faut étudier plus précisément la structure de l'observateur. Par exemple il est possible de choisir le gain K de l'observateur dans le but de rendre une composante du vecteur $\tilde{y}(t)$ particulièrement sensible à un défaut donné, il s'agit du placement de structure propre [**?**].

Les caractéristiques statistiques de $\tilde{y}(t)$ peuvent également être exploitées. En effet, si un comportement aléatoire de $\tilde{y}(t)$ peut s'expliquer par la présence de bruits de mesure (généralement modélisés par un bruit blanc de variance constante), il est anormal que la moyenne ou la variance de $\tilde{y}(t)$ change brutalement. Différentes techniques de test de caractères aléatoires ou de détection de saut de moyenne ou de variance sont exposées dans [37].

Diagnostic à base de bancs d'observateurs

Dans l'étude précédente, un observateur unique est construit à partir de toutes les grandeurs mesurées, donc l'estimation des sorties est affectée par tous les défauts de capteurs. Dans le but de discerner quel est le défaut présent, on préférera construire un banc de m observateurs, où chacun dépend uniquement d'une sortie. Dans ce cas, l'erreur d'estimation des sorties de chaque observateur ne sera sensible qu'à un seul défaut de capteur, il est alors facile de déduire que la mesure en défaut est celle correspondant à une erreur d'estimation des sorties non nulle. Cette méthode est applicable sous des conditions d'observabilité du système relativement fortes, pour relaxer ces conditions, on peut construire un banc d'observateurs où chacun utilise toutes les sorties sauf une. Dans ce cas une logique de décision simple permet de mettre en évidence le défaut apparu [**?**].

4.2 Quelques extensions

Dans ce paragraphe, on présentera quelques méthodes permettant d'étendre les résultats présentés précédemment lorsque le modèle n'est pas connu avec exactitude, ou lorsqu'il n'est pas linéaire.

Prise en compte des incertitudes
Les incertitudes de modélisation ou de mesure sont généralement modélisées sous la forme de signaux d'entrée inconnus $d_a(t)$ et $d_c(t)$ affectant le système :

$$\begin{cases} \frac{dx(t)}{dt} = Ax(t) + Bu(t) + B_d d_a(t) \\ y(t) = Cx(t) + C_c f_c(t) + C_d d_c(t) \\ x(0) = x_0 \end{cases} \tag{36}$$

Dans le cas où les signaux de perturbations sont inconnus, sous certaines hypothèses de découplage, il est possible de décomposer le système en deux parties : une soumise aux entrées inconnues et l'autre indépendante de ces entrées [16] [8]. En utilisant exclusivement le sous système exempt de perturbation, on obtient une erreur d'estimation des sorties ne dépendant que des défauts à détecter.

Si les conditions de découplage parfait ne sont pas vérifiées, il est souhaitable d'atténuer l'influence des perturbations sur l'erreur d'estimation. Il faut alors définir une norme pour quantifier cette influence. La synthèse d'observateur se ramène alors au calcul des gains de l'observateur minimisant la norme de la fonction de transfert des perturbations vers l'erreur d'estimation. Les deux normes couramment utilisées sont les normes H_2 et H_∞, permettant respectivement de minimiser l'énergie de l'erreur d'estimation et l'amplification maximale entre les perturbations et l'erreur d'estimation [45], [14].

Dans le cas où le processus est affecté par des perturbations modélisées par des variables aléatoires (de moyennes nulles, non corrélées entre elles et de variances connues), on utilise généralement le filtre de Kalman [20] [32]. Le filtre de Kalman donne une estimation sans biais (la moyenne de l'estimée est égale à la valeur réelle à estimer) et de variance minimale (la dispersion des estimations autour de la valeur réelle est minimisée).

Si les caractéristiques statistiques des perturbations sont inconnues, mais que leurs plages de variation sont connues, on peut utiliser l'arithmétique des intervalles afin de calculer une enveloppe contenant de manière sûre la valeur réelle de l'état et des sorties. La procédure de diagnostic consiste alors à générer une alarme lorsque les intervalles obtenus pour les sorties possibles (calculées à partir des mesures et des bornes des perturbations) et les sorties estimées sont disjoints [23].

Dans le cas où le modèle n'est pas exactement connu, les incertitudes de modélisation peuvent être compensées par l'utilisation d'un observateur proportionnel intégral [47].
Certains modèles font apparaître simultanément des équations dynamiques et des équations statiques (relations de maillage, bilan de matière), on parle alors de systèmes singuliers [7]. La synthèse d'observateurs pour ce type de systèmes est proche de la synthèse d'observateurs pour systèmes à entrées inconnues en effet la partie statique d'un tel système peut être considérée comme une perturbation [9]. De ce fait de nombreuses techniques de diagnostic ont été étendues à ce type de systèmes [29], [30].

Diagnostic de systèmes non linéaires
Très peu de systèmes sont réellement linéaires, de ce fait le modèle linéaire ne représente correctement le comportement du système qu'autour d'un point de fonctionnement donné. Afin de s'affranchir de cette limitation, on peut déterminer un ensemble de modèles valables chacun en différents points de fonctionnement. Un modèle valide autour de l'ensemble des points de fonctionnement est obtenu par pondération entre tous les modèles, on parle de modèle polytopique ou flou [48]. Dans ce cas, la synthèse d'un observateur est effectuée en mélangeant différents observateurs, dédiés à chacun des modèles linéaires [**?**]. L'étude de l'erreur d'estimation des sorties ainsi obtenue permet d'étendre les techniques de diagnostic aux systèmes flous [1].

Dans le cas des systèmes non-linéaires, des hypothèses sur la nature des non-linéarités (systèmes bilinéaires, présence de non-linéarités lipschitziennes, ...) permettent généralement d'utiliser des techniques de synthèse proches de celles existant pour le cas linéaire afin de construire des observateurs particuliers (observateurs bilinéaires [46], observateurs à modes glissants [15], observateurs adaptatifs [50], ...).

5 Conclusion

La surveillance des systèmes, le diagnostic étant l'un des éléments importants de cette fonction, est sans doute l'un des domaines de recherche auquel une très grande attention est donnée. En effet, après la construction d'un processus d'ordre technologique, il convient d'assurer son exploitation et en particulier de garantir que son fonctionnement est celui pour lequel il a été conçu. Le diagnostic a précisément comme objectif de déterminer l'état de fonctionnement de ce processus. Dans cet exposé, nous nous somme limité à trois aspects importants relatifs à la validation de données et la détection de défauts, sachant qu'il existe bien d'autres facettes à ce problème et notamment les stratégies de pronostic (prédiction de l'évolution des défauts), de détection de changement de modes de fonctionnement, de réaction face aux défauts.

Références

[1] Akhenak A., Chadli M., Ragot J., Maquin D. State Estimation via Multiple Observer with Unknown Input. Application to the three tank System. *Proc. of the* 5^{th} *IFAC SAFEPROCESS*, pp. 245-251, 2003.

[2] Alhaj-Dibo M., Maquin D., Ragot J. Data reconciliation : a robust approach using a contaminated distribution. Control Engineering Practice, 2007. Accepted article to be published.

[3] Bazin C., S Rochon-Tremblay, C Gosselin. Estimation Of Gas Flow Rates and Pellet Temperatures in an Iron Oxide Induration Furnace. Canadian Metallurgical Quarterly 42, pp 301-312, 2003.

[4] Boukhris A., Mourot G. , Ragot J. Estimation d'état et estimation paramétrique simultanée avec technique de régularisation pour les séries temporelles courtes et bruitées. Mediterranean Conference on Electronics and Automatic Control, MCEA'98, Marrakech, Maroc, September 17-19, 1998.

[5] Brown D., Maréchal F., Heyen G., Paris J. Data Reconciliation and Sensor System Design of a Paper Deinking Process. ESCAPE-13, Lappeenranta, 2003.

[6] Caussinus H., S. Hakam, A. Ruiz-Gazen, Projections révélatrices contrôlées, groupements et structures diverses. Rev. Statist. Appl., LI (1), 37-58, 2003.

[7] Dai L. *Singular Control Systems*. Springer, Berlin, 1989.

[8] Darouach M., Zasadzinski M., Xu S.J. Full-order observers for linear systems with unknown inputs. *IEEE Trans. Autom. Contr.*, 39, pp. 606-609, 1994.

[9] Darouach M., Boutayeb M. Design of observers for descriptor systems. *IEEE Trans. Autom. Contr.*, 40, pp. 1323-1327, 1995.

[10] Deltour J.L., Canivet E., Sanfilippo F., Sau J. Data Reconciliation on the Complex Hydraulic System of Canal de Provence. J. Irrig. and Drain. Engrg., 131 (3), pp. 291-297, May/June 2005.

[11] Dragoni A. F. , Giorgini P. Sensor data validation for nuclear power plants through bayesian conditioning and Dempster's rule of combination. Computers and artificial intelligence, 17, (2-3) , pp. 151-168, 1998.

[12] Dunia R., Qin S.J., Edgar T.F., McAvoy T.J. Identification of faulty sensor using principal component analysis. AICHE, 42, pp. 2797-2812, 1996.

[13] Dunia R., Qin S.J., Subspace approach to multidimentional identification and reconstruction, AIChE Journal, Vol. 44, 1998, pp. 1813-1831.

[14] Edelmayer A., Bokor J., Keviczky L. An H_∞ Approach to Robust Detection of Failures in Dynamical Systems, Proc. of the 33rd IEEE Conference on Decision and Control, pp. 3037-3039, 1994.

[15] Edwards C., Spurgeon S.K. , Patton R.J.. Sliding mode observers for fault detection and isolation. Automatica, 36 (4), pp. 541-553, 2000.

[16] Hou M., Muller P.C. Design of observers for linear systems with unknown inputs, *IEEE Trans. on Autom. Contr.*, 37 (6), pp.871–875, 1992

[17] Harkat M.F., Mourot G., Ragot J. Diagnostic de fonctionnement de capteurs d'un réseau de surveillance de la qualité de l'air par analyse en composantes principales. Journal Européen des Systèmes Automatisés, 39 (4), pp. 417-436, 2005.

[18] Harkat M.F., Mourot G., Ragot J. An improved PCA scheme for sensor FDI : application to an air quality monitoring network. Journal of Process Control, 16 (6), pp. 625-634, 2006.

[19] Jacob J., Paris J. Data Sampling and Reconciliation, Application to Pulp and Paper Mills, Appita Journal, Part I : Methodology and Implementation, 56, (1), 25-9 & 52 (2003) ; Part II : Case Studies, 56 (2), pp. 116-21, 2003.

[20] R.E. Kalman, A new approach to linear filtering and prediction problems. ASME Journal of Basic Engineering, 82-D, pp. 34-45,1960.

[21] Kratz F., Mourot G., Maquin D., Loisy F. Detection of measurement errors in nuclear power plants . IFAC Symposium on Control of Power Plants and Power Systems, München, March 9-11, 1992.

[22] Luenberger D. Observers for multivariable systems, *IEEE Trans. on Autom. Contr.*, 11, pp. 190-197, 1966

[23] Lunze S., T. Steffen et U. Riedel. Fault diagnosis of dynamical systems based on state-set observers, *14th International Workshop on Principles of Diagnosis, DX'03*, Washington D.C., USA, 2003.

[24] Luong M., Maquin D., Ragot J. Sensor network design for failure detection and isolation . IFAC Conference on Control of Industrial Systems, Belfort, France, May 20-22, 1997.

[25] Mandel D., Abdollahzadeh A., Maquin D., Ragot J. Data reconciliation by inequality balance equilibration . International Journal of Mineral Processing, 53, pp. 157-169, 1998.

[26] Maquin D., Narasimhan S., Ragot J. Data validation with unknown variance matrix. Computers and Chemical Engineering Supplement, pp. S609-S612, 1999.

[27] Maquin D., Ragot J. Diagnostic des systèmes linéaires. Collection pédagogique d'automatique. Hermès Science Publications, 2000.

[28] Maquin D., Ragot J. Surveillance des réseaux de distribution d'eau potable. In Diagnostic, coordinateur : Bernard Dubuisson, chapitre 11, 30 p. Traité Information, Commande, Communication, IC2, Hermès Science Publications, Paris, 2001.

[29] Marx B., D. Koenig et D. Georges. Robust Fault Diagnosis for Descriptor Systems - a Coprime Factorization Approach, *Proc. of the IFAC SAFEPROCESS 03*, pp. 507-512, 2003.

[30] Marx B., D. Koenig et D. Georges, Robust Fault Diagnosis for Linear Descriptor Systems using Proportional Integral Observers, *Proc. of the 42nd IEEE Conference on Decision and Control*, pp. 457-462, 2003.

[31] Mourot G., Maquin D., Ragot J. Simultaneous state and parameter estimation. Application to data validation for urban sewer network control. 14th IFAC World Congress, Beijing, China, July 5-9, 1999.

[32] M. Najim. Filtrage optimal. Technip volume R7228-S1, pp. 12-18 1998.

[33] S. Narasimhan, C. Jordache. Data reconciliation and gross error detection. Gulf publishing Company, Houston, Texas, 2000.

[34] Ozyurt D.B., Pike R.W. Theory and practice of simultaneous data reconciliation and gross error detection for chemical processes, *Computers and Chemical Engineering*, 28, pp. 381-402, 2004.

[35] Puig V., Saludes J., Quevedo J. Worst-case simulation of discrete linear time-invariant interval dynamic systems, *Reliable Computing*, 9 (4), pp. 251-290, 2003.

[36] Qin S.J., Dunia R. Determining the number of principal components for best reconstruction, Journal of Process Control, 10, pp. 245-250, 2000.

[37] Ragot J., Darouach M., Maquin D. et Bloch G. *Validation de Données et Diagnostic*, Hermès, France, 1990.

[38] Ragot J., Maquin D. and Adrot O. LMI approach for data reconciliation fg. In *38th Conference of Metallurgists, Symposium Optimization and Control in Minerals, Metals and Materials Processing*, Quebec, Canada, August 22-26, 1999.

[39] Ragot J., Maquin D. Reformulation of data reconciliation problem with unknown-but-bounded errors. Industrial and Engineering Chemistry Research, 43 (6), pp. 1530-1536, 2004.

[40] Ragot J., Maquin D, Alhaj-Dibo M. Linear mass balance equilibration : a new approach for an old problem. ISA Transactions, 44 (1), pp. 23-35, 2005.

[41] Ragot J., Chadli M., Maquin D. Mass balance equilibration : a robust approach using contaminated distribution, *AIChE Journal*, 51 (5), pp. 1569-1575, 2005.

[42] Ragot J., Maquin D. Data reconciliation : a robust approach using contaminated distribution. Application to a petrochemical process. 16th IFAC World Congress, Prague, Czech Republic, July 4-8, 2005.

[43] Ragot J., Maquin D. Fault measurement detection in an urban water supply network . Journal of Process Control, 16 (9), pp. 887-902, 2006.

[44] Romagnoli J.A., Sánchez M.C. *Data processing and reconciliation for chemical process operations*. Elsevier Science and Technology Books, Saint Louis, Missouri, Etats-Unis, 2000.

[45] Shaked U., Theodor Y. H_∞-Optimal estimation : a tutorial, *24th Conference on Decision and Control, Tucson, USA*, 44 (2), pp. 253–264, 1992.

[46] Yu D., Shields D.N., Daley S. A bilinear fault detection observer and its application to a hydraulic system. *International Journal of Control*, 64 (6), pp. 1023-1047, 1996.

[47] Soffker D., Yu T.J., Müller P.C. State estimation of dynamical systems with nonlinearities by using proportional integral observers, *Int. J. Systems Science*, 26 (9), pp. 1571-1582, 1995.

[48] Takagi T., Sugeno M. Fuzzy Identification of Systems and its Application to Modeling and Control. *IEEE Trans. Syst., Man, Cybern.*, 15, pp. 116-132, 1985.

[49] Yi H.-S., Kim, J. H. and Han C. Industrial application of gross error estimation and data reconciliation to byproduction gases in iron and steel making plants. In *International Conference on Control, Automation and Systems*, Muju, Korea, October 16-19, 2002.

[50] Zhang Q. Adaptive observer for MIMO linear time varying systems. IEEE Transactions on Automatic Control, 47 (3), pp. 525-529, 2002.