

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

22 juin 2005

José Ragot et Didier Maquin

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.

Email {jragot,dmaquin}@ensem.inpl-nancy.fr

## Résumé

réconciliation de données, équilibrage de bilan de production, estimation d'état, grosses erreurs, diagnostic.

## 1 Introduction

Habituellement, lorsqu'on procède à la critique et à la validation des mesures collectées sur un processus physique, on teste l'adéquation des mesures vis-à-vis du modèle du processus (phase de détection des incohérences) puis on ajuste ces mesures au moyen de termes correctifs de façon à vérifier le modèle (phase d'estimation de l'état du processus). L'amplitude de ces termes correctifs est ensuite analysée, les valeurs les plus importantes pouvant être révélatrices de mesures aberrantes (phase de localisation d'erreurs). Si ces erreurs se révèlent importantes, les mesures correspondantes doivent être supprimées et on effectue généralement un nouvel ajustement des mesures restantes. Cette façon de procéder suppose donc, outre la connaissance des mesures, celle du modèle du processus qui sert de contrainte à l'estimation et aussi celle de la précision des mesures qui inter-viennent dans les fonctions poids affectant les différents termes correctifs.

Depuis plusieurs décennies, les techniques de réconciliation de données ont été largement développées et appliquées dans de nombreux secteurs de l'industrie [6], [11]. C'est le cas notamment des secteurs chimiques [13], [21], [32] et minéralurgiques [5], [9], [33] où, en particulier, les méthodes d'équilibrage de bilan-matière et de bilan-énergie sont bien implantées et ont fait l'objet de logiciels de portée générale [35] ou adaptée à des applications particulières (Bilco [7] et Algosys [2] dans le domaine minéral, Aspen dans le domaine chimique par exemple).

Dans cette présentation, on rappelle les fondements du problème de réconciliation de données, en évoquant ensuite les difficultés d'application compte tenu d'hypothèses et de contraintes technologiques d'applicabilité. Deux techniques récentes de réconciliation sont exposées.

## 2 Principe de base de la réconciliation des données

On considère un processus caractérisé par le vecteur des variables  $x$  (de dimension  $v$ ) et représenté par le modèle linéaire :

$$Mx = 0 \quad x \in \mathcal{R}^v \quad M \in \mathcal{R}^{n.v} \quad (1)$$

Cette formulation permet notamment de décrire le processus à partir d'équations de bilan-matière ; la matrice  $M$  (à coefficients connus) encore appelée matrice d'incidence, traduit les connexions entre les processus unitaires qui constituent le processus global. La représentation (1) est cependant générale et peut aussi s'appliquer à des systèmes dynamiques discrétisés [24]. A partir des capteurs, on dispose de mesures  $x_m$  des variables  $x$  qui sont liées, de façon additive, aux erreurs  $\epsilon$  de mesure par la relation :

$$x_m = x + \epsilon \quad (2)$$

On peut mettre en évidence la présence de ces erreurs en évaluant les résidus de bilan :

$$r = Mx_m \quad (3)$$

qui ne sont pas nuls en général. Cependant, ce non respect des équations de bilan peut être du uniquement à quelques mesures erronées qu'il convient de localiser. Face à cette incohérence, il convient alors d'estimer l'état du procesus, et pour cela, connaissant son modèle et les mesures, on peut utiliser le principe du maximum de vraisemblance si la distribution des erreurs de mesures est connue. Dans le cas d'une distribution normale de matrice de variance-covariance  $V$ , cela revient à minimiser, sous la contrainte (5), la fonction quadratique (4) :

$$\Phi = \frac{1}{2} \| \hat{x} - x_m \|_{V^{-1}}^2 \quad (4)$$

$$M\hat{x} = 0 \quad (5)$$

On obtient (sous réserve que la matrice  $M$  soit de plein rang) comme expression de l'estimé [13] :

$$\hat{x} = (I - VM^T(MVM^T)^{-1}M) x_m \quad (6)$$

On déduit aisément de l'estimation (6) deux grandeurs importantes, le vecteur des termes correctifs  $e$  et le critère résiduel  $\Phi$  [23] :

$$e = x_m - \hat{x} \quad (7)$$

$$\Phi_r = x_m^T M^T (MVM^T)^{-1} M x_m \quad (8)$$

Le lecteur sera attentif à ces deux expressions qui dépendent directement des mesures et par conséquent fournissent des moyens de test de la qualité de ces mesures, l'un en étudiant de façon locale la correction (7) apportée à chaque mesure, l'autre en analysant globalement une fonction des corrections (8). Notamment, on peut démontrer que si les erreurs de mesure  $\epsilon$  sont des réalisations de variables à distributions normales centrées, les termes correctifs  $e$  suivent une loi normale centrée et le critère résiduel  $\Phi_r$  suit une loi du *khi2* ; le non respect de ces lois de distribution peut être révélateur de la présence de valeurs de mesures aberrantes qu'il convient alors de détecter et de localiser.

### 3 Les difficultés

La formulation précédente permet de corriger les mesures  $x_m$  en proposant une estimation  $\hat{x}$  vérifiant exactement le modèle du processus. Lorsque ce modèle traduit les lois de conservation de la matière ou de l'énergie, ce résultat a une grande importance pratique car fournit des grandeurs cohérentes utilisables pour l'amélioration de la conduite du système. La portée de l'approche précédente reste cependant limitée voire simpliste puisque des hypothèses assez restrictives ont été implicitement utilisées : modèle du système parfaitement connu, modèle linéaire, modèle statique, mesures toutes disponibles, mesures à précisions identiques ... Dans ce qui suit, on aborde ces difficultés en donnant la façon de poser le problème à traiter, le lecteur étant invité à consulter les références indiquées.

#### 3.1 Mesures incomplètes : observabilité

La réconciliation de données repose sur l'estimation de l'état du système à partir de mesures. Au préalable à cette estimation, il est donc nécessaire de se poser le problème de l'observabilité du système c'est-à-dire de savoir si à partir des mesures disponibles, l'estimation de l'état du système est techniquement possible. Pour cela, l'équation de mesure (2) est généralisée sous la forme :

$$x_m = Cx + \epsilon \quad C \in \mathcal{R}^{m.v} \quad (9)$$

la matrice  $C$  indiquant les rangs des variables qui sont effectivement mesurées. Par exemple :

$$C = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

montre que seules les variables 2, 4 et 5 sont mesurées. De plus, les variables sont toujours liées par l'équation issue du modèle du système (1) et on peut remarquer que l'état  $x$  du système est solution de :

$$\begin{pmatrix} M \\ C \end{pmatrix} x = \begin{pmatrix} 0 \\ I \end{pmatrix} x_m + \begin{pmatrix} 0 \\ -I \end{pmatrix} \epsilon \quad (10)$$

Ainsi, pour que l'état  $x$  soit estimable (en ne prenant pas en compte les erreurs de mesure), il est nécessaire que la matrice  $\begin{pmatrix} M \\ C \end{pmatrix}$  soit de plein rang. Dans le cas contraire, l'état du système ne peut être déterminé (par manque d'informations en nombre suffisant). De façon plus complète, l'étude de la structure de la matrice  $\begin{pmatrix} M \\ C \end{pmatrix}$  donne des renseignements intéressants sur la partie de  $x$  estimable, la partie de  $x$  qui peut être déduite des mesures, la partie de  $x$  qui peut être estimée de différentes manières (notion de degré de redondance). De façon plus générale, cette analyse conduit à séparer les variables (et les équations du système) selon des critères d'observabilité, de redondance, de déductibilité [23], [22], [4].

Par extension, remarquons que la matrice  $C$  peut prendre en compte la présence de redondances matérielles, c'est-à-dire le fait qu'une même grandeur peut être mesurée par plusieurs capteurs. L'analyse d'observabilité s'étend sous certaines conditions aux systèmes non linéaires et en particulier aux systèmes multi-linéaires rencontrés en chimie et en minéralurgie traduisant des bilans matière en flux total et en flux partiel (par espèce chimique ou minérale).

#### 3.2 Précision des mesures inconnue

Dans la section 2, on a montré comment estimer l'état du système à partir des mesures et de leur variance (6). Cette variance traduit la précision des mesures effectuées. Il arrive que cette dernière

ne soit pas connue et si l'utilisateur n'a aucune connaissance a priori lui permettant d'en fixer l'ordre de grandeur, il convient alors d'estimer cette variance, à condition toutefois de disposer d'un nombre suffisant de mesures. Ce problème un peu délicat nécessite alors de reformuler le problème d'estimation en revenant à la définition de la fonction de vraisemblance des mesures. Celle-ci est construite à partir d'hypothèse sur la distribution des erreurs de mesure et, dans le cas gaussien où la matrice de variance (inconnue) des erreurs est notée  $V$ , la fonction de vraisemblance des estimés  $\hat{x}$  s'écrit :

$$\mathcal{V} = \frac{1}{(\sqrt{2\pi})^m |V|^{m/2}} \exp\left(-\frac{1}{2} \|\hat{x} - x_m\|_{V^{-1}}^2\right) \quad (11)$$

Le problème de réconciliation des données consiste maintenant à déterminer  $\hat{x}$  maximisant la fonction de vraisemblance  $\mathcal{V}$  sous la contrainte de modèle :

$$M\hat{x} = 0 \quad (12)$$

D'un point de vue technique, il faut maximiser  $\mathcal{V}$  par rapport aux inconnues  $\hat{x}$  et  $V$  tout en respectant la contrainte (12). Ce problème a été traité et l'on peut se référer à [15], [31] pour une démonstration complète. Ainsi, même si la variance des mesures n'est pas connue, cette extension permet de réconcilier les données dont la précision est a priori inconnue.

### 3.3 Modèles incertains

Les techniques précédentes ont toutes été présentées dans le cadre des systèmes parfaitement connus. En pratique, les modèles des systèmes n'en sont qu'une image plus ou moins précise et dans certains cas, les systèmes étant à paramètres non constants ou mal connus, les modèles précédents ne sont pas toujours adaptés. Pour prendre en compte une telle situation, considérons le cas simple où seule la matrice des contraintes  $M$  présente des incertitudes supposées inconnues mais bornées :

$$M(\theta)x = 0 \quad -\delta \leq \theta \leq \delta \quad (13)$$

Le problème consiste alors à minimiser le critère (4) sous la contrainte (13), ce qui fournit, en général, un ensemble de solutions admissibles compte tenu des valeurs prises par le paramètre  $\theta$ .

A titre d'exemple, rappelons la relation entre pour-cent solide  $\alpha$  d'une pulpe homogène et sa densité  $d$  :

$$\alpha = C\left(1 - \frac{1}{d}\right)$$

$C$  étant caractéristique des densités des matériaux constituant ce mélange. Un "problème" typique de réconciliation des mesures est la suivant : avec  $\alpha = 0.51$ ,  $d = 1.50$  et  $1.50 \leq C \leq 1.56$ , quelle estimation cohérente peut-on donner pour le pour-cent solide et la densité ? Le problème peut être formalisé de la façon suivante. On note  $\hat{d}$  et  $\hat{\alpha}$  les estimations recherchées de la densité et du pour-cent solide. Compte tenu du formalisme général énoncé à la section 2, ces estimées sont solution du problème :

$$\{\hat{d}, \hat{\alpha}\} = \arg \min_{d, \alpha} \left( (d - d_m)^2 + (\alpha - \alpha_m)^2 \right) \quad (14a)$$

$$\hat{\alpha} = C\left(1 - \frac{1}{\hat{d}}\right) \quad (14b)$$

$$C_{min} \leq C \leq C_{max} \quad (14c)$$

Pour ne pas alourdir la présentation, la solution de ce problème n'est pas développée ; indiquons une possibilité qui consiste à résoudre (14a) sous la contrainte (14b) pour chaque valeur de  $C$  satisfaisant

(14c) et à faire ensuite la synthèse des estimés obtenus ce qui donne en fait un ensemble de solutions admissibles.

Bien que l'idée de modèle imprécis est fort ancienne, les travaux sur la réconciliation de mesures vis-à-vis de modèles incertains sont assez récents [16], [17].

### 3.4 Biais de mesure et valeurs aberrantes

Les capteurs qui équipent un système peuvent être sujets à dérèglement et dans ce cas fournissent des mesures  $x_m$  biaisées. Si la réconciliation des mesures est faite en appliquant directement la formule (6), l'estimée qui en résulte est altérée par ces biais. Ce point a fait l'objet de nombreux travaux visant à corriger les estimées de ces biais [10], [11], [33]. L'une des approches proposées résout ce problème de façon séquentielle schématisée par les étapes suivantes :

- on estime l'état du système à partir de (6) en utilisant toutes les mesures  $x_m$
- on forme les termes correctifs  $\tilde{x} = |\hat{x} - x_m|$
- on détecte dans les composantes de  $\tilde{x}$  si le terme le plus grand est anormal
- après suppression de la mesure ainsi suspectée, on réitère le processus d'estimation

La procédure est efficace, la seule difficulté étant de définir le seuil de détection au dessus duquel un terme correctif est déclaré anormalement élevé. A cette procédure séquentielle, on peut préférer une procédure plus globale prenant en compte dès sa formalisation, de la présence d'erreurs de faibles amplitudes et de biais auxquels les estimées ne doivent pas être sensibles (voir section 5).

### 3.5 Systèmes non linéaires

Le système physique et les capteurs qui lui sont associés sont maintenant décrits par des relations non linéaires :

$$\begin{aligned} f(x) &= 0 \\ x_m &= g(x) + \epsilon \end{aligned}$$

Le principe de la réconciliation énoncé dans le cas des systèmes linéaires reste valable. L'estimé est obtenu en minimisant le critère  $\Phi$  (15) sous la contrainte (16) :

$$\Phi = \frac{1}{2} \| g(\hat{x}) - x_m \|_{V^{-1}}^2 \quad (15)$$

$$f(\hat{x}) = 0 \quad (16)$$

La résolution de ce problème peut être complexe selon la nature des fonctions  $f$  et  $g$ ; des techniques itératives [20] permettent de construire la solution  $\hat{x}$  à partir d'une estimée initial "bien choisie". Cependant, dans le domaine des processus de traitement de minerais, lorsqu'on s'intéresse aux bilans de production, les lois de conservation de la matière en flux total et en flux partiel (par constituant) peuvent s'exprimer de façon multilinéaire ; la particularité de ces structures peut être mise à profit pour proposer des solutions au problème précédent. Des articles spécialisés [20], [23] ont largement développé ce point de vue dans les domaines du génie chimique et du traitement des minerais.

### 3.6 Systèmes dynamiques

Les techniques qui viennent d'être rappelées dans le cas statique sont à l'origine des stratégies "modernes" de diagnostic ; la génération d'équations de redondance a été notamment étendue au cas des systèmes dynamiques

$$\dot{x}(t) = f(x(t))$$

$$x_m(t) = x(t) + \epsilon(t)$$

La encore le principe de la réconciliation des mesures vis-à-vis du modèle s'applique directement, mais cette fois en utilisant toutes les mesures de l'horizon  $[0 \tau]$ . L'estimé  $\hat{x}$  doit rendre minimal le critère :

$$\Phi = \int_0^\tau \| \hat{x}(t) - x_m(t) \|^2 dt$$

sous la contrainte :

$$\dot{\hat{x}}(t) = f(\hat{x}(t)) \quad t \in [0 \tau]$$

Sous la forme indiquée, il faut néanmoins convenir que la recherche du minimum de  $\Phi$  est assez délicate à conduire. Cependant, l'usage de technique de discrétisation du temps permet de proposer des solutions tout à fait acceptable d'un point de vue numérique et temps de calcul. On trouve dans les références [23], [1], [21] la façon de résoudre ce problème et des exemples d'application.

### 3.7 Placement optimal de capteurs

Au paragraphe 3.1, nous avons évoqué le problème d'observabilité qui, concrètement, permet de répondre à la question : l'état du système est-il estimable à partir des mesures disponibles ? Dans la négative, il est alors intéressant de savoir quelles variables mesurer pour que l'état du système devienne estimable. Sous cette forme, le problème n'est pas très bien posé, car il suffirait de répondre en disant que si toutes les variables étaient mesurées, alors l'estimation serait toujours possible (voir 6). Il est plus judicieux de poser ce problème avec des contraintes de coût. En effet, le nombre de capteurs est en général limité pour des raisons économiques ou technique ; dans ce cas, il s'agit de choisir le nombre minimal de capteurs et de définir les variables à mesurer pour que l'état du système soit complètement estimable. La figure 1 ci-dessous illustre ce problème. Le réseau comporte 15 voies de circulation de matière et seulement les flux 1, 2, 4, 6, 8 et 10 sont mesurés. Que peut-on déduire de la connaissance de ces mesures ?

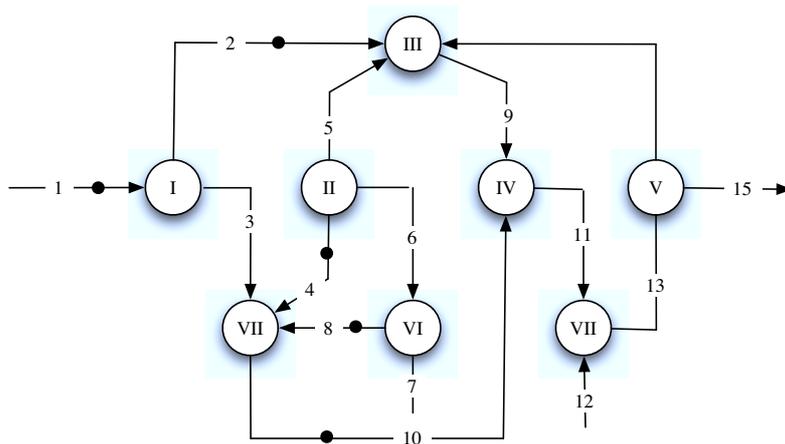


FIG. 1 – Placement de capteurs

On constate que les flux mesurés ne sont pas directement liés par des équations de redondance. On peut aussi déduire de l'analyse de ce réseau que les flux 1, 2, 4, 8 et 10 sont liés par une équation de redondance obtenue en agrégeant les unités I et VII. De plus la mesure du flux 6 n'intervient pas dans

cette redondance et les flux 3, 5 et 7 se déduisent des flux mesurés ; les flux 9, 11, 12 et 13 ne peuvent être déduits des flux connus. Le placement de capteurs consiste ici à définir le nombre de capteurs supplémentaires à utiliser et à quelle position les placer de façon à pouvoir déduire tous les autres flux. Parmi les solutions envisageables, en voici quelques unes :  $\{14, 15\}$ ,  $\{11, 12\}$ . Il y a en fait 12 possibilités de placer 2 capteurs supplémentaires pour rendre observable l'ensemble du réseau. Des techniques d'optimisation permettent de définir ces solutions possibles et, avec des contraintes supplémentaires, de choisir une solution particulière.

D'un point de vue général, ce problème admet de nombreuses variantes et en particulier celle visant à définir le nombre et la position des capteurs en intégrant les coûts de maintenance des capteurs et leur degré de fiabilité [34], [18].

## 4 Techniques intervalles

On introduit maintenant une technique "nouvelle" de réconciliation utilisant un principe de résolution uniquement par satisfaction de contraintes de type inégalité.

Réexaminons le problème de la réconciliation des mesures avec la vision plus heuristique que peut en avoir un utilisateur non averti des techniques d'estimation statistique. En l'absence de connaissance théorique sur la distribution des erreurs de mesure, on peut souvent traduire la recherche de l'état  $\hat{x}$  en terme d'intervalles de confiance, ceux-ci étant généralement issus de considérations expérimentales. On peut alors formuler mathématiquement le problème de réconciliation en termes de respect de contraintes inégalités, les amplitudes des corrections des mesures et celles des résidus d'équations de bilan étant contraintes à rester dans le domaine défini par les inégalités :

$$\begin{aligned} \underline{e}_x &\leq \hat{x} - x_m \leq \bar{e}_x \\ \underline{r}_x &\leq M\hat{x} \leq \bar{r}_x \end{aligned} \quad (17)$$

Les bornes inférieures  $\underline{e}_x$  et supérieures  $\bar{e}_x$  seront choisies en fonction de connaissances empiriques sur l'état du processus et en particulier sur les domaines de variation probable des différentes grandeurs. Comme ces bornes traduisent l'amplitude des corrections que l'on apporte aux mesures, elles peuvent dépendre de la précision de ces mesures. Les bornes  $\underline{r}_x$  et  $\bar{r}_x$  sont liées aux degrés de satisfaction des contraintes de bilan et dépendent aussi des importances relatives attribuées aux différentes équations de bilan (ces bornes peuvent être choisies nulles si l'on souhaite satisfaire exactement les contraintes). L'ensemble des inégalités doubles (18) peut s'exprimer sous la forme d'inégalités simples par rapport aux inconnues  $\hat{x}$  :

$$\begin{aligned} -\hat{x} + x_m + \underline{e}_x &\leq 0 \\ \hat{x} - x_m + \bar{e}_x &\leq 0 \\ -M\hat{x} + \underline{r}_x &\leq 0 \\ M\hat{x} - \bar{r}_x &\leq 0 \end{aligned} \quad (18)$$

faisant ainsi apparaître une structure directement exploitable par les logiciels de résolution de systèmes d'inéquations ; le lecteur pourra consulter, par exemple, les routines de la bibliothèque LMI (Linear Matrix Inequalities) de Matlab.

### Extension au cas des bilans en flux partiel

Généralement, les flux considérés sur un processus sont caractérisés par leurs débits (massiques) et la teneur en différents constituants ou espèces minérales ; dans la suite, on considère le cas où un seul constituant est pris en compte du point de vue teneur, mais la généralisation au cas multi-constituants ne pose pas de difficultés. Il convient alors, pour effectuer le traitement le plus complet possible, de prendre en compte les équations de conservation de masse totale et de masse par constituant. On obtient des équations de bilan en flux total et en flux partiel pouvant s'écrire, dans le cas de mesures sans erreurs :

$$Mx = 0 \quad (19)$$

$$Mx * y = 0$$

Le bilan en flux partiel utilise l'opérateur \* qui permet de faire le produit de deux vecteurs composante par composante. Les mesures des grandeurs  $x$  et  $y$  étant en général entachées d'erreurs, il convient de procéder à une estimation des flux et des teneurs. En généralisant le formalisme proposé précédemment (18), cela revient à chercher les estimations  $\hat{x}$  et  $\hat{y}$  des flux totaux et des concentrations telles que :

$$\underline{e}_x \leq \hat{x} - x_m \leq \bar{e}_x \quad (20)$$

$$\underline{r}_x \leq M\hat{x} \leq \bar{r}_x$$

$$\underline{e}_y \leq \hat{y} - y_m \leq \bar{e}_y$$

$$\underline{r}_{xy} \leq M\hat{x} * \hat{y} \leq \bar{r}_{xy}$$

Il apparaît clairement que ce système est bilinéaire en  $\hat{x}$  et  $\hat{y}$  et que sa résolution pose quelques difficultés. On peut proposer deux techniques pour résoudre le système (20), la première (sous-optimale) permet de se ramener à des inéquations linéaires par rapport aux variables, la deuxième (optimale) procède par itérations successives pour satisfaire l'ensemble des inéquations conservées sous leurs formes non-linéaires [23].

### Application à des données granulométriques

On considère les données relatives à un séparateur de particules minérales. Le modèle utilisé pour les équations de conservation de matière prend en compte les débits d'alimentation  $Q_a$ , de surverse  $Q_o$ , de sous-verse  $Q_u$ , les granulométries  $a_i, o_i, u_i$  et la courbe de partage  $p$  du séparateur. On dispose alors de trois modèles de bilan en flux total, en flux par tranche granulométrique (pour 4 tranches) et en flux de surverse à partir de la courbe de partage :

$$Q_a - Q_o - Q_i = 0 \quad (21)$$

$$Q_a a_i - Q_o o_i - Q_u u_i = 0 \quad i = 1..N \quad (22)$$

$$Q_o o_i = p_i Q_a a_i \quad i = 1..N \quad (23)$$

Les mesures des débits et des granulométries apparaissent dans la colonne 2 de la table 1. Les coefficients de partage du séparateur (pour les quatre intervalles granulométriques considérés) sont :

$$p = (0.881 \quad 0.50 \quad 0.119 \quad 0.018)$$

Une analyse préalable des mesures est effectuée. Elle consiste à tester la cohérence des mesures vis-à-vis des différentes équations de bilan. Comme le montrent les résultats de la table 2, les résidus sont non nuls ; les bilans ne sont donc pas équilibrés et on procède alors à une réconciliation des mesures selon la technique précédemment exposée. Pour cela on a fixé d'une part les bornes inférieure et

supérieure des estimations (colonnes 3 et 5 de la table 2) et d'autre part les seuils de satisfaction des bilans (0.05 pour le bilan total, 0.05 pour le bilan granulométrique, 0.25 pour le bilan issu de la courbe de partage).

TAB. 1 – Grandeurs réconciliées

Variable	Mesure	Borne inférieure	Estimation	Borne supérieure	Taux correction
$Q_a$	19.973	16.977	20.018	22.969	-0.2
$Q_o$	8.599	7.739	8.565	9.459	0.4
$Q_u$	11.422	10.279	11.453	12.564	-0.3
$a_1$	0.186	0.149	0.178	0.223	4.5
$o_1$	0.344	0.275	0.357	0.412	-3.9
$u_1$	0.047	0.033	0.043	0.061	7.6
$a_2$	0.217	0.174	0.216	0.260	0.4
$o_2$	0.271	0.217	0.260	0.325	4
$u_2$	0.169	0.119	0.183	0.220	-8.2
$a_3$	0.238	0.190	0.232	0.285	2.2
$o_3$	0.063	0.050	0.064	0.075	-1.6
$u_3$	0.347	0.243	0.359	0.451	-3.4
$a_4$	0.141	0.113	0.144	0.169	-2.1
$o_4$	0.012	0.009	0.010	0.015	12.2
$u_4$	0.250	0.175	0.244	0.325	2.5

TAB. 2 – Test de cohérence des mesures brutes

Résidus de bilan total	-0.048			
Résidus de bilan partiel	0.220	0.073	0.248	-0.142
Résidus de bilan de partage	0.314	-0.163	0.025	-0.052

L'équilibrage est réalisé selon la technique précédente, ce qui revient à résoudre l'ensemble des inégalités :

$$\begin{aligned}
 \underline{e} &\leq \hat{Q}_a - Q_a \leq \bar{e} \\
 \underline{e} &\leq \hat{Q}_o - Q_o \leq \bar{e} \\
 \underline{e} &\leq \hat{Q}_u - Q_u \leq \bar{e} \\
 \underline{e} &\leq \hat{a}_i - a_i \leq \bar{e} && i = 1..4 \\
 \underline{e} &\leq \hat{o}_i - o_i \leq \bar{e} && i = 1..4 \\
 \underline{e} &\leq \hat{u}_i - u_i \leq \bar{e} && i = 1..4 \\
 \underline{r} &\leq \hat{Q}_a - \hat{Q}_o - \hat{Q}_i \leq \bar{r} \\
 \underline{r} &\leq \hat{Q}_a \hat{a}_i - \hat{Q}_o \hat{o}_i - \hat{Q}_u \hat{u}_i \leq \bar{r} && i = 1..N \\
 \underline{r} &\leq \hat{Q}_o \hat{o}_i - p_i \hat{Q}_a \hat{a}_i \leq \bar{r} && i = 1..N
 \end{aligned}$$

Après équilibrage de bilan, on peut analyser les grandeurs réconciliées. Les résidus de bilan sont donnés à la table 3 et les estimations des débits et des granulométries apparaissent dans la table 1. Le lecteur pourra constater que l'ensemble des contraintes est satisfait à  $10^{-2}$  près.

TAB. 3 – Test de cohérence des estimations

Résidus de bilan total (*10 <sup>-5</sup> )	-4.9			
Résidus de bilan partiel (*10 <sup>-2</sup> )	-0.12	-0.13	0.02	0.04
Résidus de bilan partage (*10 <sup>-2</sup> )	7.2	-6.6	0.72	-3.8

## 5 Approche robuste : distribution contaminée

Une des difficultés majeures de la réconciliation des données est due à la présence de valeurs aberrantes dans les mesures. Lors de l'estimation (6), ces valeurs aberrantes se retrouvent disséminées dans toutes les estimées ce qui n'est pas l'effet recherché. Comme indiqué dans la section 3.4, l'un des principes adoptés consiste à procéder à la réconciliation en dépit de la présence de ces valeurs aberrantes, à analyser les résultats obtenus pour localiser ces valeurs aberrantes, puis à réitérer l'opération de réconciliation après suppression des mesures suspects.

Une façon plus séduisante de procéder consiste à modifier l'hypothèse de distribution des erreurs de mesure. On construit alors une fonction de répartition des erreurs prenant en compte d'une part les erreurs de faibles amplitudes et d'autre part les erreurs de fortes amplitudes. Ainsi, pour une erreur  $\epsilon$ , cette fonction se présente sous la forme :

$$p(\epsilon) = wp_1(\epsilon) + (1 - w)p_2(\epsilon) \quad (24)$$

$$p_1(\epsilon) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{1}{2}\left(\frac{\epsilon}{\sigma_1}\right)^2\right) \quad (25)$$

$$p_2(\epsilon) = \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{1}{2}\left(\frac{\epsilon}{\sigma_2}\right)^2\right) \quad (26)$$

en choisissant  $\sigma_2 \gg \sigma_1$ , de façon à décrire les deux distributions. Le paramètre  $w$  réalise une pondération entre les deux distributions, c'est-à-dire entre les erreurs de faibles amplitudes et les biais.

Cette loi est adoptée pour toutes les erreurs de mesure. On peut alors montrer, que si l'on maximise la fonction de vraisemblance des grandeurs vraies construite à partir de la loi de distribution des erreurs de mesure, on obtient une estimée se présentant structurellement sous une forme tout à fait comparable à celle obtenue précédemment (6), les poids  $W$  ayant maintenant une nouvelle définition :

$$\hat{x} = (I - W_{\hat{x}}A^T(AW_{\hat{x}}A^T)^{-1}A)\tilde{x} \quad (27a)$$

$$W_{\hat{x}}^{-1} = \text{diag}_{i=1..v} \left( \frac{\frac{w p_{\hat{x},1,i}}{\sigma_{x,1}^2} + \frac{(1-w) \hat{p}_{\hat{x},2,i}}{\sigma_{x,2}^2}}{w p_{\hat{x},1,i} + (1-w)p_{\hat{x},2,i}} \right) \quad (27b)$$

$$p_{\hat{x},1,i} = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{1}{2}\left(\frac{\hat{x}_i - x_{m,i}}{\sigma_1}\right)^2\right) \quad (27c)$$

$$p_{\hat{x},2,i} = \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{1}{2}\left(\frac{\hat{x}_i - x_{m,i}}{\sigma_2}\right)^2\right) \quad (27d)$$

On note que le système (27) n'est pas explicite par rapport à la solution  $\hat{x}$ . En fait,  $\hat{x}$  (27a) dépend des poids  $W$  (27b) qui eux mêmes dépendent de  $\hat{x}$  (27c) et (27d). Une procédure itérative doit être mise en œuvre pour résoudre ce système.

En dépit d'une écriture qui peut sembler lourde, la procédure proposée se généralise aux problèmes de réconciliation de données mettant en jeu des bilans en flux total et en flux partiel comme proposé à la section 4, où les distributions dites contaminées portent à la fois sur les variables de flux  $x$  et les variables de concentration  $y$ . Une démonstration complète peut être trouvée dans [].

L'exemple proposé illustre, en simulation, ce cas à propos d'un système comportant 9 nœuds et 16 voies de circulation de la matière (figure 2). Les mesures disponibles sont des débits massiques et des

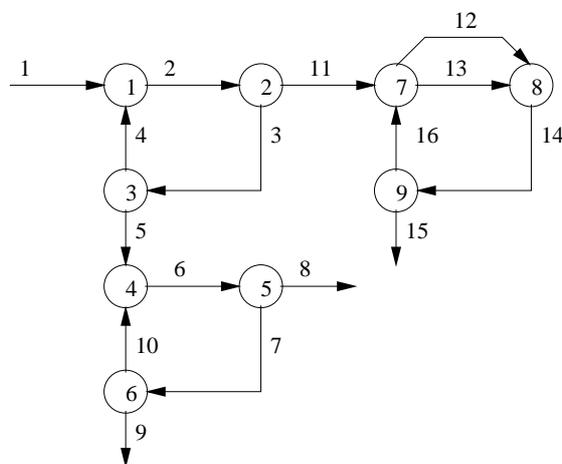


FIG. 2 – Flowsheet

concentrations en une espèce minérale ; les équations de conservation de masse sont regroupées dans la table (4). Le tableau (5) regroupe l'ensemble des mesures et des résultats d'estimation. Comme cet exemple a été entièrement simulé, ce tableau fournit les données vraies (colonnes 3 et 7) et les mesures (colonnes 4 et 8) que l'on peut comparer aux estimations. On notera ici que le flux de la voie 5 n'est pas connu ainsi que la concentration de la voie 2 ; par contre, elles ont été estimées. Pour cet exemple, les mesures des flux voies 3, 7 et 16 ont été volontairement biaisées afin de simuler des biais de capteur ; il en est de même pour les concentrations mesurées des voies 1, 9 et 12.

Deux méthodes ont été appliquées. La méthode des moindres carrés (sigle *LS*) et la méthode dite robuste (sigle *RLS*) utilisant le principe des distributions contaminées de façon à prendre directement en compte la présence de mesures aberrantes.

L'examen du tableau révèle que :

- les grandeurs non mesurées ont été estimées
- pour la méthode *RLS* les flux 3, 7 et 16 ont été fortement corrigés
- pour la méthode *LS* tous les flux ont été corrigés
- pour la méthode *RLS* les concentration 1, 9 et 12 ont été fortement corrigés
- pour la méthode *LS* toutes les concentrations ont été corrigées

On conclut, sur cet exemple, que l'utilisation d'une double distribution conduit à une méthode robuste vis-à-vis des biais de mesure. Cela se traduit par une estimation "correcte" des variables biaisées sans affecter de façon sensible les estimations des autres variables. Cela est particulièrement révélateur sur la figure (3) qui visualise les corrections (en valeur absolue) des flux et des concentrations obtenues par les deux méthodes. Sur cette figure, les abscisses représentent les numéros des voies de mesure (de 1 à 16) et les ordonnées sont relatives aux valeurs absolues des termes correctifs. Les variables qui avaient été volontairement biaisées sont bien corrigées en utilisant l'approche *RLS* ; par contre, l'approche *LS*

node1	$x_1 - x_2 - x_4 = 0$	$x_1y_1 - x_2y_2 - x_4y_4 = 0$
node2	$x_2 - x_3 - x_{11} = 0$	$x_2y_2 - x_3y_3 - x_{11}y_{11} = 0$
node3	$x_3 - x_4 - x_5 = 0$	$x_3y_3 - x_4y_4 - x_5y_5 = 0$
node4	$x_5 - x_6 - x_{10} = 0$	$x_5y_5 - x_6y_6 - x_{10}y_{10} = 0$
node5	$x_6 - x_7 - x_8 = 0$	$x_6y_6 - x_7y_7 - x_8y_8 = 0$
node6	$x_7 - x_9 - x_{10} = 0$	$x_7y_7 - x_9y_9 - x_{10}y_{10} = 0$
node7	$x_{11} - x_{12} - x_{13} - x_{16} = 0$	$x_{11}y_{11} - x_{12}y_{12} - x_{13}y_{13} - x_{16}y_{16} = 0$
node8	$x_{12} - x_{13} - x_{14} = 0$	$x_{12}y_{12} - x_{13}y_{13} - x_{14}y_{14} = 0$
node9	$x_{14} - x_{15} - x_{16} = 0$	$x_{14}y_{14} - x_{15}y_{15} - x_{16}y_{16} = 0$

TAB. 4 – Equations du système de la figure 2.

corrige l'ensemble des variables (débits et concentrations, *c'est-à-dire que les biais ont été répartis sur l'ensemble des*

Var.	$x_{inf}$	données vraies	mes.	RLS est.	LS est.	données vraies	mes.	RLS est.	LS est.
	$\underline{x}$	$x$	$\tilde{x}$	$\hat{x}$	$\hat{x}$	$y$	$\tilde{y}$	$\hat{y}$	$\hat{y}$
1	50	57.72	57.74	58.35	59.47	<b>6.23</b>	8.96	<b>6.36</b>	<b>7.08</b>
2	50	67.71	67.05	66.61	68.29			7.04	7.67
3	50	<b>52.98</b>	63.91	<b>53.74</b>	<b>57.07</b>	6.37	6.60	6.52	7.26
4	5	7.99	7.94	8.25	8.82	11.65	11.84	11.86	11.56
5	10			45.49	48.25	5.43	5.45	5.55	6.45
6	50	55.71	55.89	55.97	59.36	6.40	6.26	6.43	7.03
7	30	<b>32.13</b>	39.49	<b>32.49</b>	<b>35.50</b>	7.24	7.10	7.19	8.32
8	20	23.58	23.44	23.48	23.86	5.26	5.42	5.37	5.10
9	22	21.40	21.35	22.00	24.39	<b>5.62</b>	8.78	<b>5.74</b>	<b>7.78</b>
10	5	10.73	10.45	10.48	11.11	10.47	10.31	10.23	9.52
11	5	12.73	13.03	12.86	11.22	9.07	9.25	9.22	9.75
12	5	17.05	16.56	16.76	18.59	<b>8.80</b>	11.85	<b>8.69</b>	<b>10.01</b>
13	1	2.42	2.38	2.32	2.65	22.02	22.81	21.79	21.16
14	5	19.47	18.90	19.08	21.24	10.45	10.23	10.29	11.40
15	5	12.73	13.07	12.86	11.22	9.07	9.13	9.22	9.75
16	0	<b>6.74</b>	11.63	<b>6.22</b>	<b>10.02</b>	13.05	12.46	12.48	13.24

TAB. 5 – Mesures et estimations (pour  $y$ , les valeurs sont multipliées par 100)

## 6 Conclusion

S'il y a quelques années, l'automatisation des processus de production avait pour objectif principal l'augmentation de leur productivité par la mise en oeuvre de commandes performantes, aujourd'hui, la prise en compte des aspects maintenance, gestion technique ou sûreté de fonctionnement devient une nécessité. La complexité de plus en plus importante des systèmes automatisés de production les rend vulnérables aux défaillances qui sont à l'origine de coûts importants (risque d'accidents, de pollution, diminution de la production, non-qualité,...). Cette vulnérabilité justifie l'introduction, dans leur système de contrôle-commande, de procédures de surveillance, de diagnostic de l'état de fonctionnement, de validation de données en particulier. On a exposé ici succinctement les principes de la validation de données, en rappelant le principe de base de la réconciliation, en donnant les idées majeures sur l'analyse des résultats de la réconciliation, en énonçant les difficultés de mise en oeuvre, puis en donnant un aperçu sur quelques extensions en cours de validation. On peut imaginer que les prochains

développements scientifiques concernent probablement la prise en compte, dans le cahier des charges de conception des systèmes, des contraintes imposées pour le diagnostic de celui-ci (quantité, localisation, qualité des capteurs et actionneurs).

## Références

- [1] Abu-el-zeet Z.H., Roberts, P.D. and Becerra, V.M. Bias detection and identification in dynamic data reconciliation. In European Control Conference, Porto, Portugal, September 4-7, 2001.
- [2] Algosys inc. 1389, Avenue Galilée. Québec (Québec) Canada G1P 4G4.
- [3] Arora N. and Biegler, L.T. Redescending estimator for data reconciliation and parameter estimation. *Computers and Chemical Engineering*, 25 (11/12), 1585-1599, 2001.
- [4] Bagajewicz, M.J. *Process plant instrumentation : design and upgrade*. Lancaster, PA : Technomic Publishing Company, 2000.
- [5] Bazin C., D. Hodouin. Importance of covariance in mass balancing of particle size distribution data. *Minerals Engineering*, 14 (8), 2001.
- [6] Brown D., F. Marechal, G. Heyen, J. Paris. Application of Data Reconciliation to the Simulation of System Closure Options in a Paper Deinking Process. European Symposium on Computer aided Process Engineering-13, ed. A. Kraslawski and I. Turunen, 14, p. 1001-1006, Elsevier, 2003.
- [7] Caspeo - France. Campus BRGM, Avenue Claude Guillemin, BP 6009, OrléansCédex 2, France.
- [8] B. Ghosh-Dastider, J.L. Schafer. Outlier detection and editing procedures for continuous multivariate data. Working paper 2003-07, RAND, Santa Monica OPR, Princeton University, 2003.
- [9] Hodouin D., A. Mirabedini, S. Makni, C. Bazin, Reconciliation of mineral processing data containing correlated measurement errors, *International Journal of Mineral Processing*, 54 (3-4), p. 201-215, 1998.
- [10] Jordache C, Ternet D, Brown S. Efficient gross error elimination methods for rigorous on-line optimization, Escape 11, Kolding (Denmark), 2001.
- [11] 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.
- [12] Langenstein M., J. Jansky : Process data validation in CCGT and nuclear power plants, Paper O03/2, SmiRT 16, Washington DC, August 2001
- [13] Mah R.S.H., Stanley, G.M. and Downing, D. Reconciliation and rectification of process flow and inventory data. *Ind. Eng. Chem., Process Des. Dev.*, 15 (1), 175-183, 1976.
- [14] Mandel D., A. Abdollahzadeh, D. Maquin, J. Ragot. Data reconciliation by inequality balance equilibration. *International Journal of Mineral Processing*, 53, p. 157-169, 1999.
- [15] Maquin D., S. Narasimhan, J. Ragot. Data validation with unknown variance matrix. 9th European Symposium on Computer Aided Process Engineering, Budapest, Hungary, May 31 - June 2, 1999.
- [16] Maquin D. and Ragot J. Validation de données issues de systèmes de mesure incertains. *Journal Européen des Systèmes Automatisés*, 37 (9), 1163-1179, 2003.
- [17] Maquin D., O. Adrot, J. Ragot. Data reconciliation with uncertain models. *ISA Transactions*, 39, p. 35-45, 2000.

- [18] Maquin D., M. Luong, J. Ragot. About the design of measurement systems and fault accommodation. *Engineering Simulation, special issue on Engineering Diagnostics*, G.E. Pukhov ed., vol. 13, p. 1009-1024, 1996.
- [19] D. Mandel, A. Abdollahzadeh, D. Maquin, J. Ragot. Data reconciliation by inequality balance equilibration. *International Journal of Mineral Processing*, 53, p. 157-169, 1998.
- [20] McBrayer K.A. , T.A. Soderstrom, T.F. Edgar, R.E. Young. The application of nonlinear dynamic data reconciliation to plant data. *Computer and Chemical Engineering*, 22 (12), pp. 1907-1911, 1998.
- [21] Mingfang K., C. Bingzhen, L. Bo. An integral approach to dynamic data rectification. *Computer and Chemical Engineering*, 24, pp. 749-753, 2000.
- [22] Narasimhan S. , Jordache C. *Data Reconciliation and Gross Error Detection*. Houston, Gulf Publishing Company 2000.
- [23] Ragot J., Darouach M., Maquin D., Bloch G., Validation de données par équilibrage de bilan. *Traité des nouvelles technologies, série diagnostic et maintenance*, HERMES, juin 1990.
- [24] Ragot J., Maquin D., Sauter D. Data validation using orthogonal filters. *IEE Proceedings D, control theory and applications*, vol. 139, n° 1, p. 47-52, 1992.
- [25] Ragot J., Luong M. Maquin D. Observability of systems involving flows circulation. *International Journal of Mineral Processing*, 47, p. 125-140, 1996.
- [26] Ragot J., Maquin D. and Adrot O. LMI approach for data reconciliation. In 38th Conference of Metallurgists, Symposium Optimization and Control in Minerals, Metals and Materials Processing, Quebec, Canada, August 22-26, 1999.
- [27] Ragot J., Maquin D., Alhaj-Dibo M. Linear mass balance equilibrium : a new approach for an old problem. *ISA Transactions*, vol. 43, 2004.
- [28] 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.
- [29] Ragot J., Chadli M., Maquin D. Mass balance equilibration : a bilinear case with a robust approach using contaminated distribution. *AIChE Journal*, vol. 51, no 5, pp. 1569-1575, 2005
- [30] Rollins D.K., Davis J.F. Gross error detection when variance-covariance matrices are unknown. *AIChE Journal* 39 (8), pp. 1335 - 1341, 1993.
- [31] Romagnoli J. A., Sanchez M. C. *Data Processing and Reconciliation for Chemical Process Operations*. San Diego, Academic Press 2000
- [32] Singh, S.R., Mittal, N.K. and Sen, P.K. A novel data reconciliation and gross error detection tool for the mineral processing industry. *Minerals Engineering*, 14 (7), 808-814, 2001.
- [33] Turbatte H.C., Maquin D., Huynh C.T., Ragot J. Conception de l'architecture d'un système d'instrumentation. Application à un procédé pétrochimique. *Revue Européenne Diagnostic et Sécurité de Fonctionnement*, vol. 4, n° 1, pp. 53-72, 1994.
- [34] VALI III, User Guide, Belsim S.A., Liege, Belgium, December 2003.
- [35] Veverka V.V. and F. Madron. *Material and Energy Balancing in the Process Industries. From Microscopic Balances to Large Plants*. Elsevier, Amsterdam, 1997.
- [36] Wang D. and Romagnoli J.A. Robust data reconciliation based on a generalized objective function. In *15th IFAC World Congress on Automatic Control*, Barcelona, Spain, July 21-26, 2002.

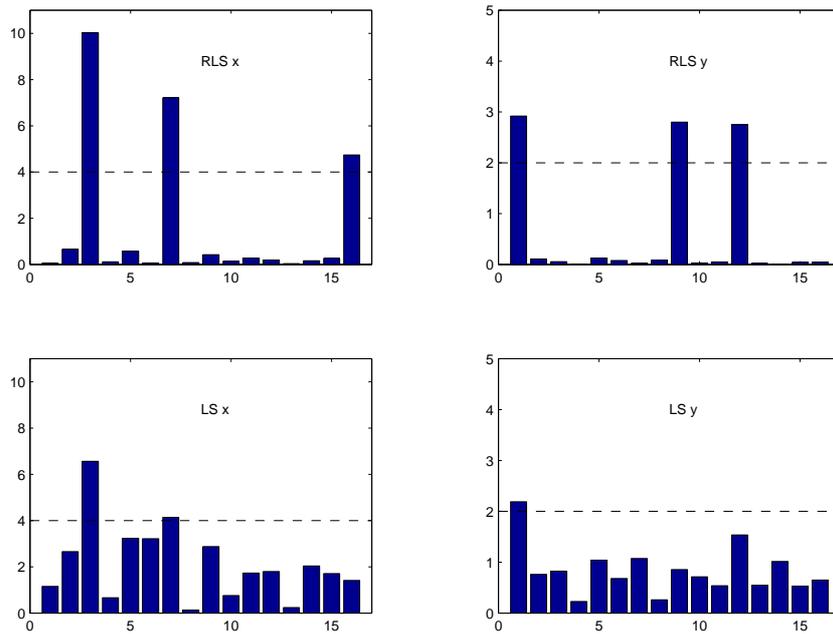


FIG. 3 – Termes correctifs pour RLS et LS : les lignes en pointillés indiquent la position du seuil adopté pour la detection des grosses erreurs