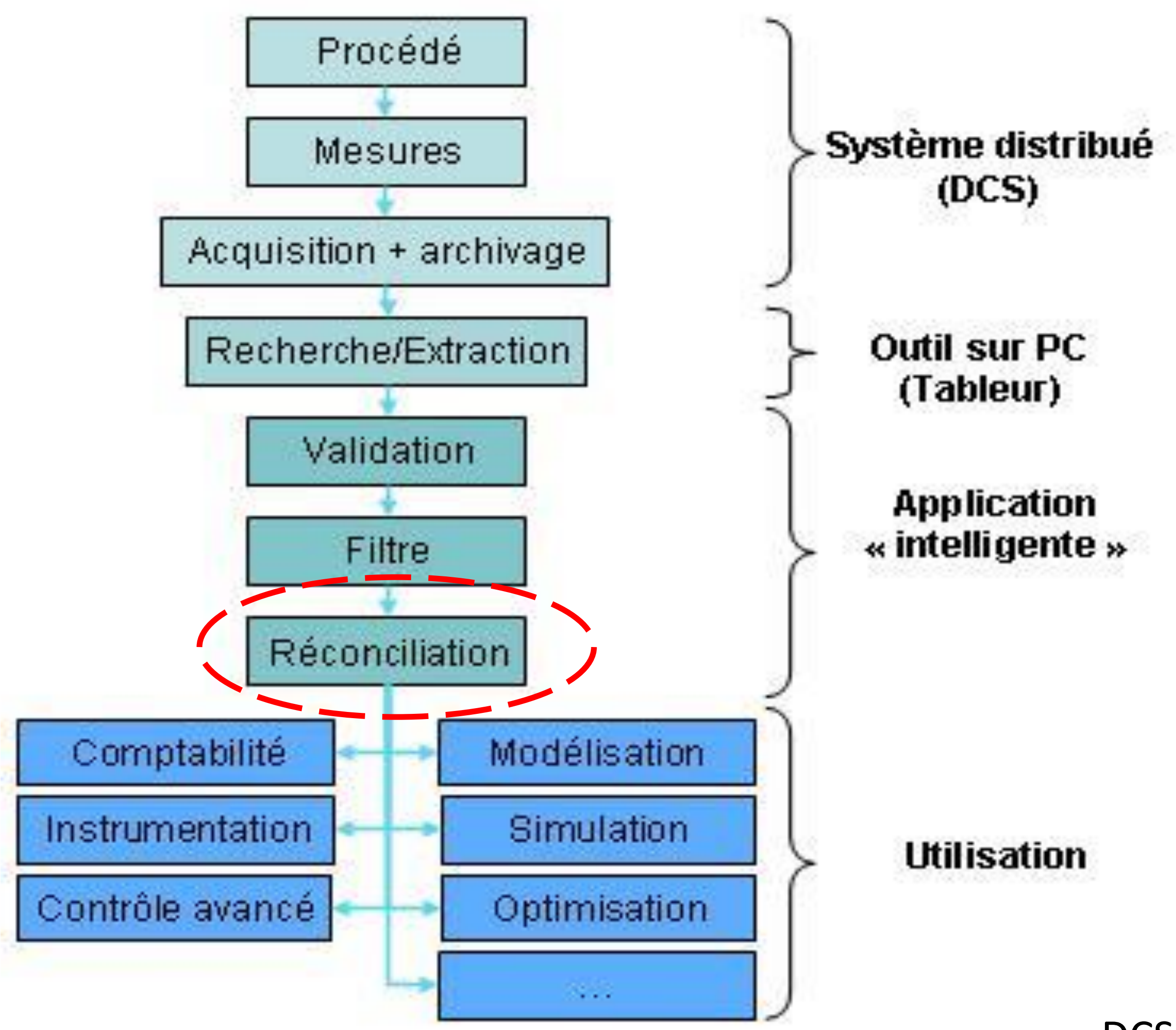
Réconciliation de données par bilans de matière

La reconciliation de données au coeur de la gestion de l'information des procédés



(Adapté de OSISoft)

DCS = Distributed Control System

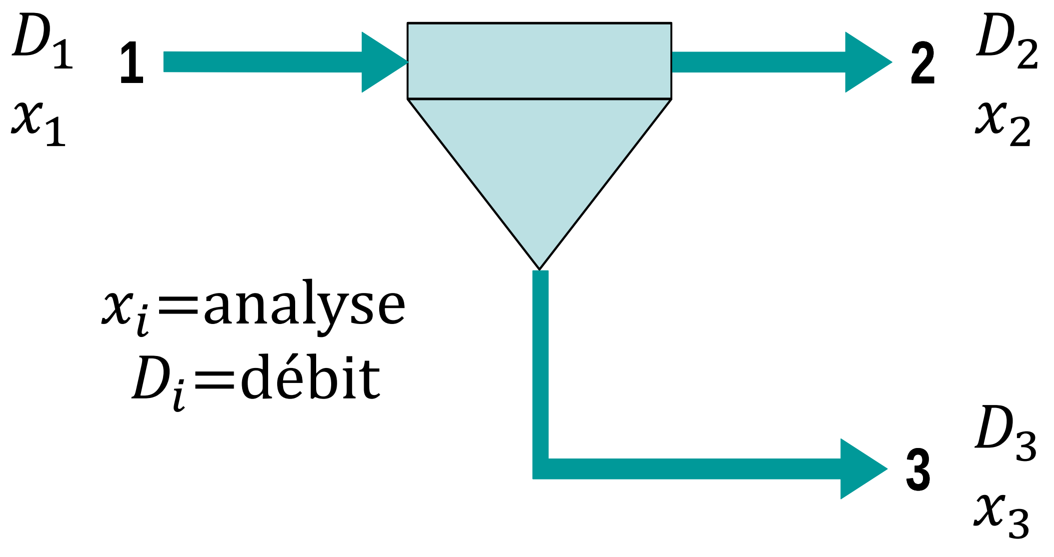
Pourquoi faire de la réconciliation de données ?

- Une mesure n’est jamais une représentation exacte de la réalité
- Toute mesure est entachée d’une erreur pouvant provenir:
  - du procédé et de ces instabilités
  - d’un échantillonnage non représentatif
  - de la technique de mesure elle-même
  - de l’appareil de mesure (mauvaise calibration)
  - etc...
- Le manque de cohérence des données mesurées en usine fait en sorte que les bilans ne ferment jamais (erreur de 5 à 30%!)
  - les opérateurs finissent par ne se fier qu’à certaines analyses…

➢ La réconciliation de données par bilan de matière vise à rendre les mesures plus significatives

➢ Son avantage réside dans la **cohérence** des valeurs « réconciliées » qui satisfont aux contraintes de conservation de la matière/énergie

Formule des deux produits



6 valeurs à obtenir, mais la connaissance de 4 permet le calcul des deux manquantes au moyen des bilans suivants:

$$\begin{aligned} D_1 &= D_2 + D_3 \\ D_1 x_1 &= D_2 x_2 + D_3 x_3 \end{aligned} \qquad (1)$$

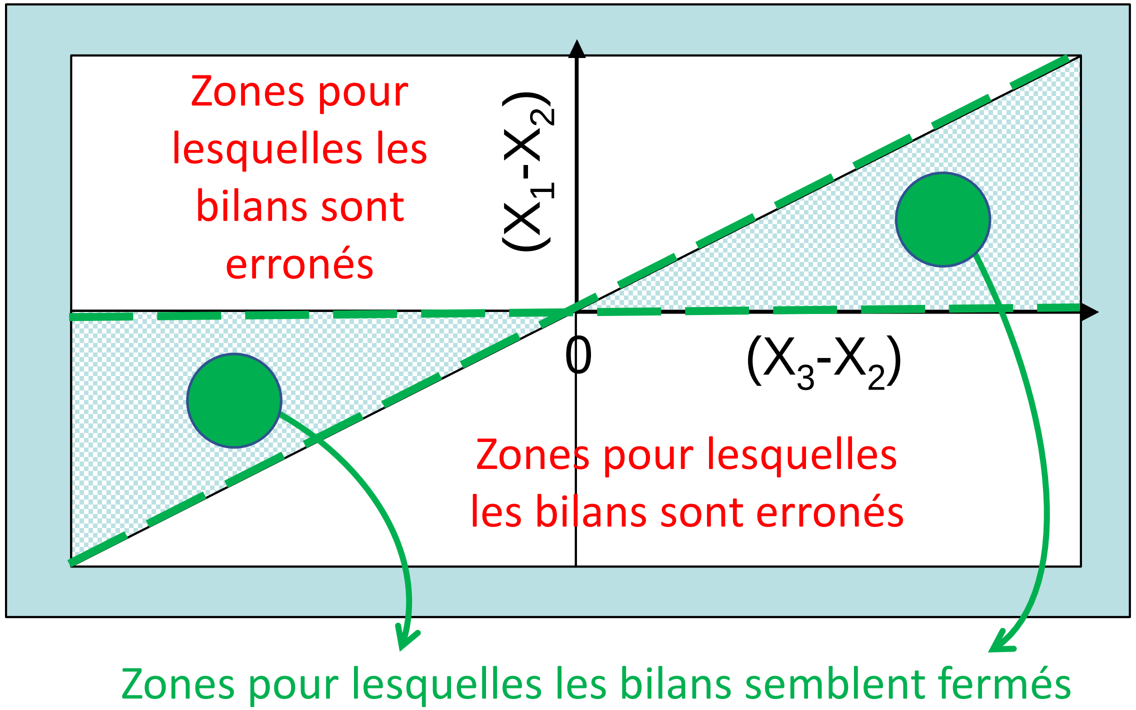
Toutefois, la **redondance** peut être exploitée pour **améliorer la qualité des mesures** faites et s'assurer d'un bilan de matière cohérent

(1) → $$\frac{D_3}{D_1} = \frac{x_1 - x_2}{x_3 - x_2} \qquad (2)$$

avec comme contraintes : $$\frac{D_3}{D_1} \leq 1 \quad (3) \qquad \text{et} \qquad \frac{D_3}{D_1} \geq 0 \quad (4)$$

Formule des deux produits (suite)

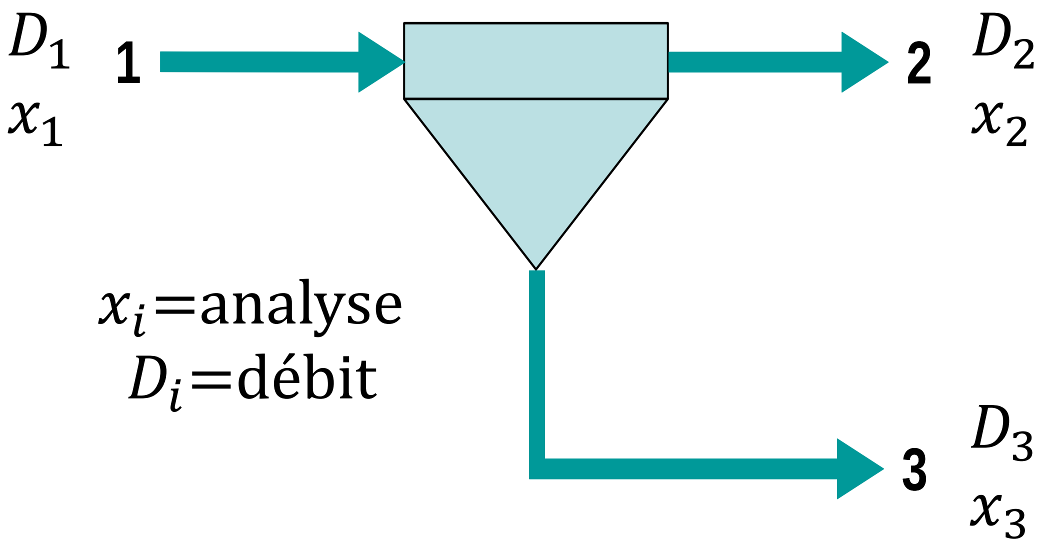
$$\left(x_1 - x_2\right) = \underbrace{\frac{D_3}{D_1}}_{0 \le \text{pente} \le 1}\left(x_3 - x_2\right)$$



Exemple

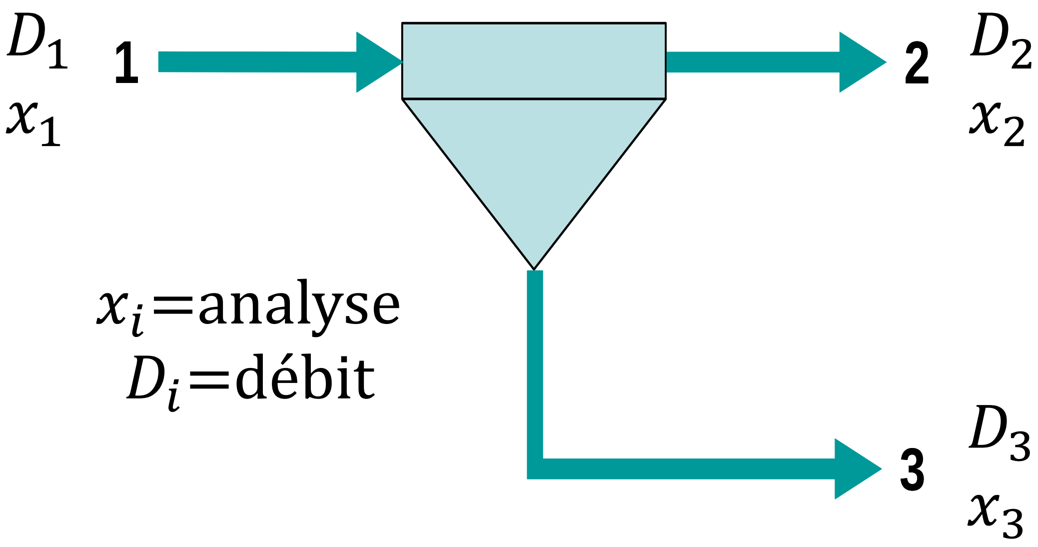
Unité de traitement du zinc:

- Mesures de la concentration de toutes les espèces dans toutes les courants
- Que remarquez-vous?

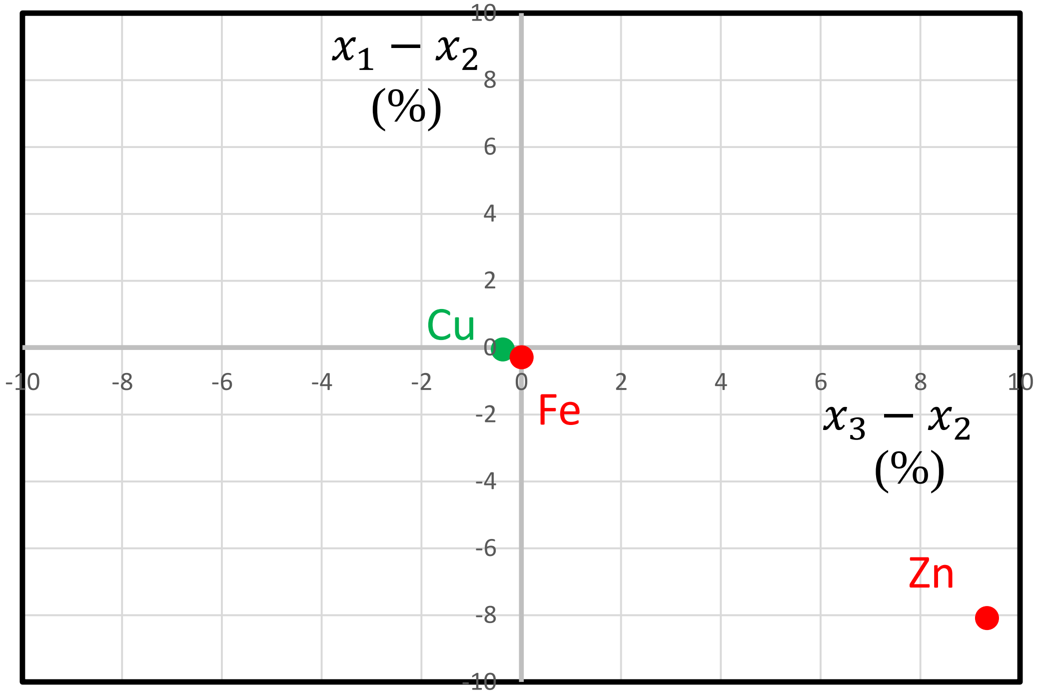


Analyse sur:	x_1 (%)	x_2 (%)	x_3 (%)
Cu	0.96	1.02	0.657
Zn	34.64	42.73	52.07
Fe	14.37	14.66	14.67

Exemple



x_i=analyse
D_i=débit



Valeur des analyses - application de la formule des deux produits

Analyse sur:	x_1 (%)	x_2 (%)	x_3 (%)	$x_1 - x_2$ (%)	$x_3 - x_2$ (%)	$\frac{D_3}{D_1}$
Cu	0.96	1.02	0.657	-0.060	-0.363	0.165
Zn	34.64	42.73	52.07	-8.090	9.340	**-0.866**
Fe	14.37	14.66	14.67	-0.290	0.010	**-29.0**

Analyses non-cohérentes ne respectant pas la conservation de la matière

La redondance est une notion centrale de la réconciliation de données

- C'est en faisant un usage judicieux de la redondance que l'on peut espérer arriver à rendre cohérents les bilans de matière
- Deux formes de redondance:
  - ➢redondance spatiale: essentielle à la résolution des équations de bilan
  - ➢Redondance temporelle: permet une estimation de la qualité des mesures et de l'atteinte d'un regime transitoire (à travers la variance σ^2)

Reconciliation de données au moyen de l'optimisation sous contraintes

- Nous souhaitons trouver les valeurs ajustées sur les débits (D_i) et les analyses (x_i) minimisant l'erreur au sens des moindres carrés entre les valeurs mesurées et les nouvelles valeurs ajustées tel que:

$$\min_{D_{i,ajusté},\, x_{i,ajusté}} \left[\sum_i \frac{\left(D_{i,mesuré} - D_{i,ajusté}\right)^2}{\sigma^2_{D_{i,mesuré}}} + \sum_i \frac{\left(xD_{i,mesuré} - xD_{i,ajusté}\right)^2}{\sigma^2_{x_{i,mesuré}}} \right]$$

tout en satisfaisant les contraintes de la conservation de la matière:

$$(D_{i,ajusté})_{in} - (D_{i,ajusté})_{out} = 0 \qquad (xD_{i,ajusté})_{in} - (xD_{i,ajusté})_{out} = 0$$

On reconnait ici un problème pouvant être résolu par la méthode des multiplicateurs de Lagrange

x_i=analyse $\quad$ D_i=débit $\quad$ σ^2=variance sur les mesures ou évaluation subjective de qualité des débits/analyses

Multiplicateurs de Lagrange : Rappel

- Supposons devoir trouver le maximum de la function suivante:

$$f(x, y) = x^2 y$$

- Avec la condition que les coordonnées x et y soient située sur la périphérie d'un cercle de rayon unitaire:

$$x^2 + y^2 = 1$$

- Comment procéder?

Reconciliation de données au moyen de l'optimisation sous contraintes (suite)

$$\min_{D_{i,ajusté},\, x_{i,ajusté}} \left[\sum_i \frac{\left(D_{i,mesuré} - D_{i,ajusté}\right)^2}{\sigma^2_{D_{i,mesuré}}} + \sum_i \frac{\left(xD_{i,mesuré} - xD_{i,ajusté}\right)^2}{\sigma^2_{xD_{i,mesuré}}} \right]$$

avec $\left(D_{i,ajusté}\right)_{in} - \left(D_{i,ajusté}\right)_{out} = 0$ et $\left(xD_{i,ajusté}\right)_{in} - \left(xD_{i,ajusté}\right)_{out} = 0$

équivalent

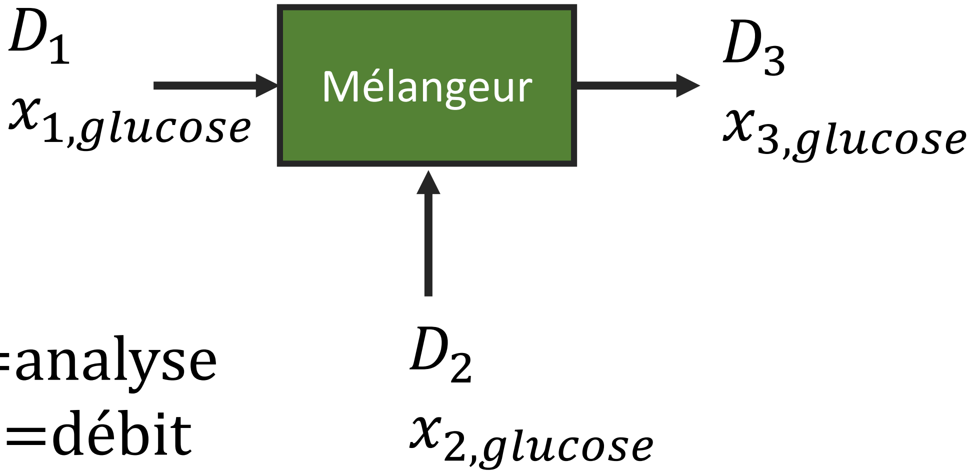
$$\min_{\substack{D_{i,a} \\ (D_{i,a})_{in} - (D_{i,a})_{out} = 0}} \left[\sum_i \frac{\left(D_{i,mesuré} - D_{i,a}\right)^2}{\sigma^2_{D_{i,mesuré}}} \right]$$

suivi de

$$\min_{\substack{x_{i,a} \\ (xD_{i,a})_{in} - (xD_{i,a})_{out} = 0}} \left[\sum_i \frac{\left(D_{i,a}\, x_{i,mesuré} - D_{i,a} x_{i,a}\right)^2}{\sigma^2_{xD_{i,mesuré}}} \right]$$

On reconnait ici des problèmes pouvant être résolus par la méthode des <u>multiplicateurs de Lagrange</u>

Exemple



x_i=analyse
D_i=débit

Ces données sont-elles cohérentes d'un point de vue des bilans de matière ?

- La formule des deux produits mène ici à:

$$\frac{D_1}{D_3} = \frac{x_3 - x_2}{x_1 - x_2} = -1.25 \leq 0$$

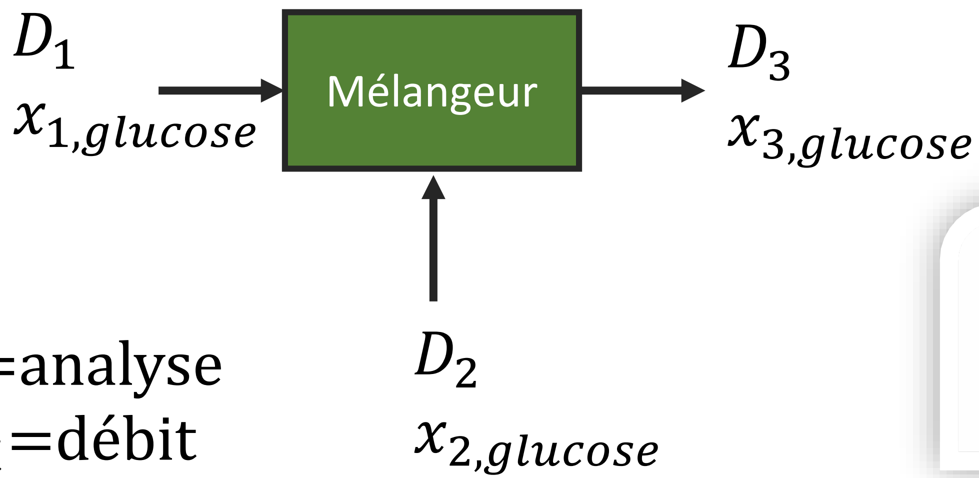
- De plus, nous avons:

$$D_1 + D_2 = 13.5 + 16.1 \neq 33.2 = D_3$$

Données sur l'unité de mélange:

Courant:	x_i (% mass.)	σ_{x_i} (% mass.)	D_i (t/h)	σ_{D_i} (t/h)
1	14.1	0.3	13.5	0.5
2	21.2	1.0	16.1	0.5
3	30.1 ?	6.0	33.2	0.3

Exemple



x_i=analyse
D_i=débit

Ces données sont-elles cohérentes d'un point de vue des bilans de matière ?

- La formule des deux produits mène ici à:

Added: Let the means be μ and ν, and the variances be σ^2 and τ^2. Then the variance of XY is, by the above argument, equal to

$$(\sigma^2 + \mu^2)(\tau^2 + \nu^2) - \mu^2\nu^2.$$

This simplifies to $\sigma^2\tau^2 + \sigma^2\nu^2 + \tau^2\mu^2$.

Données sur l'unité de mélange:

Courant:	x_i (% mass.)	σ_{x_i} (% mass.)	D_i (t/h)	σ_{D_i} (t/h)	$\sigma_{x_i D_i}$ (% mass. × t/h)
1	14.1	0.3	13.5	0.5	8.13
2	21.2	1.0	16.1	0.5	19.28
3	30.1	6.0	33.2	0.3	199.41

Courant:	x_i (% mass.)	σ_{x_i} (% mass.)	D_i (t/h)	σ_{D_i} (t/h)
1	14.1	0.3	13.5	0.5
2	21.2	1.0	16.1	0.5
3	30.1	6.0	33.2	0.3

- Commençons par trouver les $D_{i,a}$:

$$\min_{\substack{D_{i,a} \\ (D_{i,a})_{in} - (D_{i,a})_{out} = 0}} \left[\sum_i \frac{\left(D_{i,mesuré} - D_{i,a}\right)^2}{\sigma^2_{D_{i,mesuré}}} \right]$$

$$L\left(D_{1,a}, D_{2,a}, D_{3,a}, \lambda\right) = \frac{\left(13.5 - D_{1,a}\right)^2}{0.5^2} + \frac{\left(16.1 - D_{2,a}\right)^2}{0.5^2} + \frac{\left(33.2 - D_{3,a}\right)^2}{0.3^2} + \lambda(D_{1,a} + D_{2,a} - D_{3,a})$$

- On écrit ensuite: $$\frac{\partial L}{\partial D_{1,a}} = 0; \ \frac{\partial L}{\partial D_{2,a}} = 0; \frac{\partial L}{\partial D_{3,a}} = 0 \ \text{ et } \ \frac{\partial L}{\partial \lambda} = D_{1,a} + D_{2,a} - D_{3,a} = 0$$

- Là, nous dérivons et nous obtenons:

- Là, nous dérivons et nous obtenons:

$$\left.\begin{aligned}\frac{\partial L}{\partial D_{1,a}} &= 8\,D_{1,a} - 108 + \lambda = 0\\ \frac{\partial L}{\partial D_{2,a}} &= 8\,D_{2,a} - 128.8 + \lambda = 0\\ \frac{\partial L}{\partial D_{3,a}} &= 22.22\,D_{3,a} - 737.77 - \lambda = 0\\ \text{et } \frac{\partial L}{\partial \lambda} &= D_{1,a} + D_{2,a} - D_{3,a} = 0\end{aligned}\right\} \quad \left.\begin{aligned} D_{1,a} &= 15.03\ t/h\\ D_{2,a} &= 17.62\ t/h\\ D_{3,a} &= 32.65\ t/h\\ \text{et } \lambda &= -12.02\ h/t\end{aligned}\right\} \quad D_{1,a} + D_{2,a} = D_{3,a}$$

- Maintenant réalisons la deuxième minimisation pour trouver les $x_{i,a}$:

$D_{i,mesuré}\ x_{i,mesuré}$

ou

$$\min_{\substack{x_{i,a}\\ (xD_{i,a})_{in}-(xD_{i,a})_{out}=0}} \left[\sum_i \frac{\left(D_{i,a}\,x_{i,mesuré} - D_{i,a}x_{i,a}\right)^2}{\sigma^2_{xD_{i,mesuré}}}\right]$$

$$H(x_{1,a}, x_{2,a}, x_{3,a}, \lambda) = \frac{15.03^2\left(14.1 - x_{1,a}\right)^2}{8.13^2} + \frac{17.62^2\left(21.2 - x_{2,a}\right)^2}{19.28^2} + \frac{32.65^2\left(30.1 - x_{3,a}\right)^2}{199.41^2} + \lambda(15.03x_{1,a} + 17.62x_{2,a} - 32.65x_{3,a})$$

Courant:	x_i (% mass.)	σ_{x_i} (% mass.)	D_i (t/h)	σ_{D_i} (t/h)	$\sigma_{x_iD_i}$ (% mass. × t/h)†
1	14.1	0.3	13.5	0.5	8.13
2	21.2	1.0	16.1	0.5	19.28
3	30.1	6.0	33.2	0.3	199.41

- On écrit ensuite: $\frac{\partial H}{\partial x_{1,a}} = 0;\ \frac{\partial H}{\partial x_{2,a}} = 0;\ \frac{\partial H}{\partial x_{3,a}} = 0$ et $\frac{\partial H}{\partial \lambda} = 15.03x_{1,a}+17.62x_{2,a}-32.65x_{3,a} = 0$

- Là, nous dérivons et nous obtenons:

$$\left.\begin{aligned}
\frac{\partial H}{\partial x_{1,a}} &= 6.83\,x_{1,a} - 96.38 + 15.03\lambda = 0\\
\frac{\partial H}{\partial x_{2,a}} &= 1.67\,x_{2,a} - 35.41 + 17.62\lambda = 0\\
\frac{\partial H}{\partial x_{3,a}} &= 0.05\,x_{3,a} - 1.61 - 32.65\lambda = 0\\
\text{et } \frac{\partial H}{\partial \lambda} &= 15.03x_{1,a}+17.62x_{2,a} - 32.65x_{3,a} = 0
\end{aligned}\right\}
\left.\begin{aligned}
x_{1,a} &= 14.14\%\\
x_{2,a} &= 21.41\%\\
x_{3,a} &= 18.06\%\\
\text{et } \lambda &= -0.02\ h/t/\%
\end{aligned}\right\}
0 \leq \frac{D_1}{D_3} = \frac{x_{3,a}-x_{2,a}}{x_{1,a}-x_{2,a}} = 0.46 \leq 1$$

Données sur l'unité de mélange ajustées vs. originelles:

Courant:	x_i (% mass.)	σ_{x_i} (% mass.)	D_i (t/h)	σ_{D_i} (t/h)	$x_{i,ajusté}$ (% mass.)	$D_{i,ajusté}$ (t/h)
1	14.1	0.3	13.5	0.5	14.1	15.0
2	21.2	1.0	16.1	0.5	21.4	17.6
3	30.1 ?	6.0	33.2	0.3	18.1	32.6

- On peut aussi résoudre directement le problème (non-linéaire) suivant:

$$\min_{D_{i,ajusté}, x_{i,ajusté}} \left[\sum_i \frac{\left(D_{i,mesuré} - D_{i,ajusté}\right)^2}{\sigma^2_{D_{i,mesuré}}} + \sum_i \frac{\left(xD_{i,mesuré} - xD_{i,ajusté}\right)^2}{\sigma^2_{xD_{i,mesuré}}} \right]$$

avec $(D_{i,ajusté})_{in} - (D_{i,ajusté})_{out} = 0$ et $(xD_{i,ajusté})_{in} - (xD_{i,ajusté})_{out} = 0$

- Comment ?

1. Toujours à partir des multiplicateurs de Lagrange:

$$L\left(D_{i,a}, x_{i,a}, \lambda_1, \lambda_2\right) = \left[\sum_i \frac{\left(D_{i,mesuré} - D_{i,ajusté}\right)^2}{\sigma^2_{D_{i,mesuré}}} + \sum_i \frac{\left(xD_{i,mesuré} - xD_{i,ajusté}\right)^2}{\sigma^2_{xD_{i,mesuré}}} \right] + \lambda_1\left((D_{i,ajusté})_{in} - (D_{i,ajusté})_{out}\right) + \lambda_2\left((xD_{i,ajusté})_{in} - (xD_{i,ajusté})_{out}\right)$$

2. On derive $L\left(D_{i,a}, x_{i,a}, \lambda_1, \lambda_2\right)$ par rapport à $\{D_{i,a}, x_{i,a}, \lambda_1, \lambda_2\}$ avec i ∈ {1,2,3}, on égale les dérivées à zéro et on obtient un système de 8 équations, cette fois-ci, non-linéaires à résoudre pour $\{D_{i,a}, x_{i,a}, \lambda_1, \lambda_2\}$ avec i ∈ {1,2,3}
3. On utilise par exemple la méthode de Newton matricielle pour résoudre :

$$\boldsymbol{x} = \boldsymbol{x}^{(0)} \quad \text{(estimé initial)}$$

$$\mathcal{J}\left(\boldsymbol{x}^{(i)}\right)\Delta\boldsymbol{x} = -\boldsymbol{F}\left(\boldsymbol{x}^{(i)}\right)$$

$$\boldsymbol{x}^{(i+1)} = \boldsymbol{x}^{(i)} + \Delta\boldsymbol{x}$$

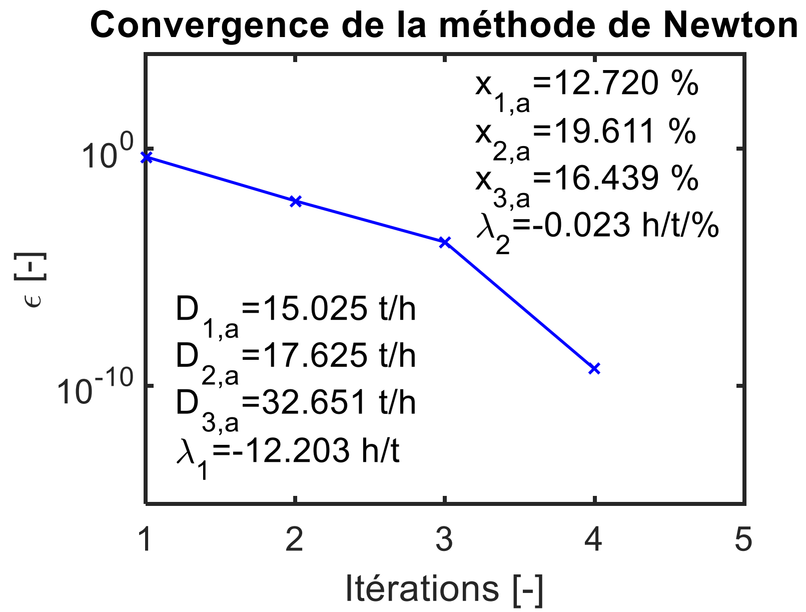
$$\left\|\boldsymbol{x}^{(i+1)} - \boldsymbol{x}^{(i)}\right\| \leq \varepsilon$$

$$\boldsymbol{x}^{(i)} = \boldsymbol{x}^{(i+1)}$$

avec : $\boldsymbol{x} = \{D_{i,a}, x_{i,a}, \lambda_1, \lambda_2\}^T$ avec i ∈ {1,2,3}

$$\boldsymbol{F}(\boldsymbol{x}) = \left\{ \frac{\partial L(\boldsymbol{x})}{\partial x_1}, \frac{\partial L(\boldsymbol{x})}{\partial x_2}, \cdots, \frac{\partial L(\boldsymbol{x})}{\partial x_8} \right\}^T$$

$\mathcal{J}(\boldsymbol{x})$ la matrice jacobienne dont les termes sont définis tel que: $\mathcal{J}_{ij} = \frac{\partial F_i(\boldsymbol{x})}{\partial x_j} = \frac{\partial}{\partial x_j}\left(\frac{\partial L(\boldsymbol{x})}{\partial x_i}\right)$



Données sur l'unité de mélange ajustées (via les approches linéaire et non-linéaire) vs. originelles:

Courant:	x_i (% mass.)	D_i (t/h)	$x_{i,ajusté}$ (% mass.)	$D_{i,ajusté}$ (t/h)	$x_{i,ajusté}$ (% mass.)	$D_{i,ajusté}$ (t/h)
1	14.1	13.5	14.1	15.0	12.7	15.0
2	21.2	16.1	21.4	17.6	19.6	17.6
3	30.1	33.2	18.1	32.6	16.4	32.6

avec
$D_{i,a}\ x_{i,mesuré}$

(voir p.13)

avec
$D_{i,mesuré}\ x_{i,mesuré}$

(voir p.13)

Pondération des mesures – sources d'erreur

- La base de l'estimation de l'erreur associée à chaque mesure réside dans la connaissance des sources d'erreur
- Sources d'erreur:

Représentativité	Quantité prélevée vis-à-vis du procédé
Échantillonnage	Échantillonnage à un moment d'instabilité Échantillons intervertis Échantillons mal prélevés (erreur procédurale) Échantillons contaminés
Analyse	Précision de l'analyseur (calibration ?) Erreur ou précision de la technique, du technicien Contamination entre échantillons
Humaine	Mauvaise entrée dans les registres Erreur de recopie pour la communication

Sources souvent peu en cause de part la precision des analyseurs utilisés

Sources où resident souvent l'erreur

- Dans les cas où on ne peut estimer chaque mesure, il est bon d'associer une cote (bonne, moyenne ou faible) à chaque mesure. On peut ensuite associer à chaque cote, basée sur l'expérience, une valeur en pourcentage (5, 10 et 15% par exemple).