EPFL

ME : À quoi ça sert ? Chimie computationnelle

Environ. Sci. Technol. **2005**, *39,* 8434−8441

Improving Data Quality for Environmental Fate Models: A Least-Squares Adjustment Procedure for Harmonizing Physicochemical Properties of Organic Compounds

URS SCHENKER, MATTHEW MACLEOD, MARTIN SCHERINGER,* AND KONRAD HUNGERBÜHLER

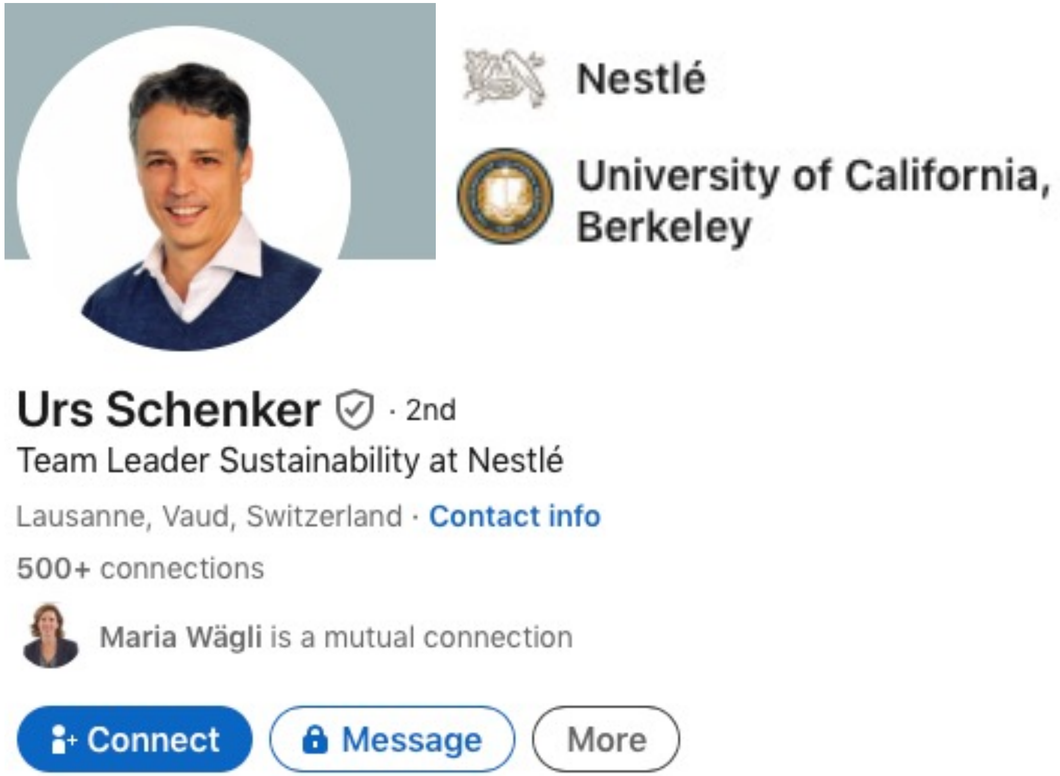
Safety and Environmental Technology Group, Swiss Federal Institute of Technology, ETH Hönggerberg, CH-8093 Zürich, Switzerland

Introduction

The distribution of o
media is a key fact
profiles. For a syste
partitioning can be
solubilities: solubil
the vapor pressure (
and solubility in oct
is dominated by org
octanol. (*R* is the ga
ature.) If ideal solute
the ratios of these so
coefficients: the ai
dimensionless form
octanol−water part
air partition coeffic

Accurate measur
ficients is a challen
chemicals that hav
analytical technique

ME : À quoi ça sert ? Chimie computationnelle *real life* après *ME* ...



- MSc EPFL 2003 (SIE)
- PhD ETHZ

(The function Ω is called the Lagrangian, and $2k$ is the Lagrange multiplier.)

$$\Omega = {v_1}^2 + {v_2}^2 + {v_3}^2 - 2k(v_1 - v_2 + v_3 - w) \qquad (11)$$

The minimizing condition is that the partial derivatives of Ω with respect to all v_i are equal to 0. This condition is fulfilled by

$$v_1 = -v_2 = v_3 = k = \frac{w}{3} \qquad (12)$$

… retour à ME -- **Exo 7+8 !**

Compensation conditionnelle
programmation pour l'analyse

- Usage des matrices auxiliares – pour *simplifier* des calculs
  - residus compensés
    ```
    Aux1 = Kll @ np.transpose(B)
    Aux2 = Aux1 @ np.linalg.inv(B @ Aux1)
    vcomp = Aux2 @ w
    ```
  - covariance des résidus compensés
    ```
    Kvcomp = Aux2 @ np.transpose(Aux1)
    ```
  - covariance des observations comp.
    ```
    Kllcomp = Kll-Kvcomp
    ```

$$\hat{\mathbf{v}} = \underbrace{\overbrace{\boxed{\mathbf{Q}_{\ell\ell}\mathbf{B}^T}}^{\text{Aux1}}\left(\mathbf{B}\overbrace{\boxed{\mathbf{Q}_{\ell\ell}\mathbf{B}^T}}^{\text{Aux1}}\right)^{-1}}_{\text{Aux2}} \cdot \mathbf{w}$$

$$\mathbf{Q}_{\hat{v}\hat{v}} = \overbrace{\mathbf{Q}_{\ell\ell}\mathbf{B}^T\left(\mathbf{B}\mathbf{Q}_{\ell\ell}\mathbf{B}^T\right)^{-1}}^{\text{Aux2}} \overbrace{\boxed{\mathbf{B}\mathbf{Q}_{\ell\ell}}}^{(\text{Aux1})^T}$$

$$\mathbf{Q}_{\hat{\ell}\hat{\ell}} = \mathbf{Q}_{\ell\ell} - \mathbf{Q}_{\hat{v}\hat{v}}$$

- Usage de fonctions auxiliaires – pour *simplifier* l'interprétation
  - écart-types, matrice de corrélation
    ```
    [sigmavcomp, Rvcomp] = covmat2cormat(Kvcomp)
    print('\n', 'sigmavcomp =', '\n', sigmavcomp)
    print('\n', 'Rvcomp =', '\n', Rvcomp)
    plot_heatmap(Rvcomp, "Matrice de corrélation des résidus compensés")
    ```

$$\sigma_{v_i}, \mathbf{R} \longleftarrow \mathbf{Q}_{\hat{v}\hat{v}}$$

Agenda ME 7-2 : Description des résultats

- Sigma *a posteriori* ($\hat{\sigma}_0$)
  - Pourquoi calculer?
  - Comment utiliser?

- Quotient global d'erreur moyenne ($\hat{\sigma}_0/\sigma_0$)
  - Pourquoi calculer?
  - Comment utiliser?
  - Application au triangle

- D'autre application
  - Gas parfait
  - Exemple sur Moodle résolu

Ecarte-type *a posteriori*

- Rappel (statistique de base)
  - Incertitude de la moyenne $\bar{y}$
    - Estimateur non-biaisé

$$s = \sigma_{\bar{y}} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{(n-1)}}$$

- Compensation conditionnelle
  - Différences
    - d'autre type des conditions
    - les observations - pas avec la même précision

$$\hat{\sigma}_0 = \sqrt{\frac{\hat{\mathbf{v}}^T \mathbf{P} \hat{\mathbf{v}}}{r}}$$

  - Similitudes
    - minimum – somme des résidus carrés
    - surdétermination – (pour la moyenne $r = n - 1$)

$$\Omega = \mathbf{v}^T \mathbf{P} \mathbf{v} \longrightarrow \text{min.}$$

Ecarte-type *a posteriori*

- Principe idée de dérivation
  - Résultat $\hat{\Omega} = \hat{\mathbf{v}}^T \mathbf{P} \hat{\mathbf{v}}$

- On espère obtenir $E\{\hat{\Omega}\} = E\left\{trace\left(\hat{\mathbf{v}}^T \mathbf{P} \hat{\mathbf{v}}\right)\right\} = \ldots$
  - On fait usage de
    - la trace - un opérateur commutatif
    - **P** – pas stochastique → à l'extérieur d'espérance
    - $E\left\{\hat{\mathbf{v}}^T \hat{\mathbf{v}}\right\} = \mathbf{K}_{\hat{v}\hat{v}}$
    - $\mathbf{K}_{\hat{v}\hat{v}} = \sigma_0^2 \cdot \mathbf{Q}_{\hat{v}\hat{v}}$
    - $\mathbf{Q}_{\hat{v}\hat{v}} = \mathbf{Q}_{\ell\ell}\mathbf{B}^T \left(\mathbf{B}\mathbf{Q}_{\ell\ell}\mathbf{B}^T\right)^{-1} \mathbf{B}\,\mathbf{Q}_{\ell\ell}$
    - $\mathbf{P} = \mathbf{Q}_{\ell\ell}^{-1}$
    - $\cdots = \sigma_0^2 \cdot trace\left(\left(\mathbf{B}\mathbf{Q}_{\ell\ell}\mathbf{B}^T\right)^{-1} \cdot \mathbf{B}\mathbf{Q}_{\ell\ell}\mathbf{B}^T\right) = \sigma_0^2 \cdot trace\left(\mathbf{I}_r\right)$
    - $= \sigma_0^2 \cdot r$
  - Si le résultat = espérance $\Longrightarrow \hat{\sigma}_0 = \sqrt{\dfrac{\hat{\mathbf{v}}^T\,\mathbf{P}\hat{\mathbf{v}}}{r}}$

Agenda ME 7-2 : Description des résultats

- Sigma *a posteriori* ($\hat{\sigma}_0$)
  - Pourquoi calculer?
  - Comment utiliser?

- Quotient global d'erreur moyenne ($\hat{\sigma}_0/\sigma_0$)
  - Pourquoi calculer?
  - Comment utiliser?
  - Application au triangle

- D'autre application
  - Gas parfait
  - Exemple sur Moodle résolu

EPFL

Quotient global $\frac{\hat{\sigma}_0}{\sigma_0}$

- Erreur moyenne
  - sans dimension!
- Indique
  - sont le(s) modèle(s) et les observations adéquate?
  - Avec sigma0 / poids unitaires

$$\frac{\hat{\sigma}_0}{\sigma_0} = \sqrt{\frac{\hat{\mathbf{v}}^T \, \mathbf{K}_{\ell\ell}^{-1} \hat{\mathbf{v}}}{r}}$$

> 1 — trop optimiste

< 1 — trop pessimiste

- Comment ?
  - $\Longrightarrow \hat{\sigma}_0 = \sqrt{\frac{\hat{\mathbf{v}}^T \, \mathbf{P} \hat{\mathbf{v}}}{r}}$
  - $\hat{\sigma}_0 = \sqrt{\frac{\hat{\mathbf{v}}^T \, \mathbf{Q}_{\ell\ell}^{-1} \hat{\mathbf{v}}}{r}}$
  - $\hat{\sigma}_0 = \sqrt{\frac{\hat{\mathbf{v}}^T \, \sigma_0^2 \mathbf{K}_{\ell\ell}^{-1} \hat{\mathbf{v}}}{r}}$

$$\mathbf{Q}_{\ell\ell} = \frac{1}{\sigma_0^2} \cdot \mathbf{K}_{\ell\ell}$$

Compensation conditionnelle
programmation pour l'analyse

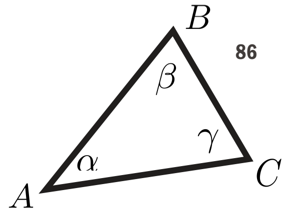
- Observations de types différents
  - on renonce souvent à l'utilisation de cofacteurs
  - Implicitement, $\sigma_0 = 1$
  - dans cette situation
    - `P = np.linalg.inv(Kll)` $\mathbf{P} = \mathbf{K}_{\ell\ell}^{-1}$

- Quetient d'erreur moyenne
  - théorie chapitre 3.4 et 3.5
  - résultats de la minimisation globale
  - le rapport global des écarts-types a posteriori / a priori pour tous les types de measures
    - `Qglob = np.sqrt((np.transpose(vcomp) @ P @ vcomp)/r)`

$$\frac{\hat{\sigma}_0}{\sigma_0} = \sqrt{\frac{\hat{\mathbf{v}}^T \, \mathbf{K}_{\ell\ell}^{-1} \, \hat{\mathbf{v}}}{r}}$$

Analyse - exemple de triangle

$$\ell_\alpha = 035.471\,\text{gon}$$
$$\ell_\beta = 107.383\,\text{gon}$$
$$\ell_\gamma = 057.122\,\text{gon}$$



- Avant

$$\begin{bmatrix} \hat{v}_\alpha \\ \hat{v}_\beta \\ \hat{v}_\gamma \end{bmatrix} = \begin{bmatrix} w/3 \\ w/3 \\ w/3 \end{bmatrix} = \begin{bmatrix} 8 \\ 8 \\ 8 \end{bmatrix} \text{mgon}$$

$$\hat{\sigma}_0 = \sqrt{\frac{\hat{\mathbf{v}}^T \mathbf{P} \hat{\mathbf{v}}}{r}} = \sqrt{\frac{w^2}{3}} = \frac{|w|}{\sqrt{3}} = \frac{13.8}{1.732} = 13.8\ \text{mgon}$$

- je le compare à quoi?
- sigma0 n'est suivant pas définit explicitement …
- … alors je fais quoi?

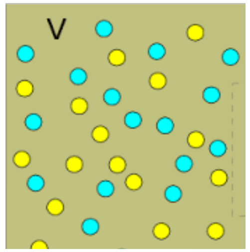
Agenda ME 7-2 : Description des résultats

- Sigma *a posteriori* ($\hat{\sigma}_0$)
  - Pourquoi calculer?
  - Comment utiliser?

- Quotient global d'erreur moyenne ($\hat{\sigma}_0/\sigma_0$)
  - Pourquoi calculer?
  - Comment utiliser?
  - Application au triangle

- D'autre application
  - Gas parfait
  - Exemple sur Moodle résolu

Gaz parfait - suite



- 2em étape : modèle stochastique des résultats
  - `Sigmavcomp` et `Rvcomp`, `sigmalcomp` et `Rlcomp`, `QuotientGlobal`
- 3em étape : autre modèle stochastique $K_{\ell\ell}$
  - Modifier σ_p , σ_V ou σ_T et observer les changements sur $\hat{v}$, $R_{\hat{v}\hat{v}}$, $\hat{\sigma}_0/\sigma_0$

- Fonction des observations compensées $\mathbf{y} = \mathbf{f}(\hat{\ell})$
  - Exemple : estimer la constate du gaz : $c = P \cdot V/T$ et σ_c
  - Option : estimer c et σ_c par pondération

    Alternative partielle à la pondération des observations :

    - calculer $c_i = \ell_{p_i} \cdot \ell_{V_i}/\ell_{T_i}$ et σ_{c_i} pour chaque état, puis pondérer

    Cas spécial: erreurs relatives constantes pour $P, V, T \longrightarrow$ moyenne arithmétique

Exemple sur Moodle

- $\Delta h_{CW} = 1719$ m. : la différence d'altitude entre la cabane et le Weisshorn obtenue en soustrayant les deux altitudes sur la carte topographique.
- $\Delta h_{CF} = 188$ m. : la différence d'altitude entre la cabane et le pied de la face mesurée avec le baromètre.
- $\Delta h_{FW} = 1534$ m. : la différence d'altitude entre le pied de la face et le sommet du Weisshorn mesurée en laissant tomber la montre baromètre.
- $H_C = 2785$ m. : l'altitude de la cabane mesurée avec la montre GNSS.
- $H_W = 4507$ m. : l'altitude du sommet du Weisshorn mesuré avec la montre GNSS.

Ils considèrent les écarts-types suivants pour leur mesures :

$$\sigma_{\Delta h_{CW}} = 0.5 \text{ m}, \quad \sigma_{\Delta h_{CF}} = 1.0 \text{ m}, \quad \sigma_{\Delta h_{FW}} = 3.0 \text{ m}, \quad \sigma_{H_p} = 2.0 \text{ m}$$

Synthèse ME 7-2 : Description des résultats

- Préférer sigma *a priori* (σ_0) ou *a posteriori* ($\hat{\sigma}_0$) ?
  - Préanalyse : pas de valeur *a posteriori* (seulement σ_0)
  - Nouveau procédé : pas de valeur *a priori*
  - Quell valeur exprime la meilleure connaissance ?

- Quotient d'erreur moyenne
  - Ecart-type *a posteriori* pour $P = K_{\ell\ell}^{-1}$, donc pour $\sigma_0 = 1$
  - Dans la forme quadratique, les unités des résidus et des poids s'annulent
  - Utile surtout pour des mesures de types différents
  - Exemple connu : quotient global ($\hat{\sigma}_0/\sigma_0$) le gaz parfait

- Triangle avec mesure des angles <u>et</u> de côtés
  - Exemple du polycopié, pp. 57-60