Exercice 8 : Chimie computationnelle

Contexte

Urs Schenker a suivi le cours de méthodes d'estimation en 1999–2000 et obtenu son diplôme EPFL en 2003. Il a réalisé son doctorat à l'ETHZ et séjourné dans plusieurs universités américaines (MIT, Berkeley). Depuis mai 2009, il est spécialiste en analyse du cycle de vie (*LCA specialist*) au centre de recherche de Nestlé à Vers-chez-les-Blanc (NRC – Lausanne).
Au cours de son doctorat au sein du *Safety and Environmental Technology Group*, Institute for Chemical and Bioengineering, ETH Hönggerberg, il fut le premier auteur de l'article suivant, paru dans la revue *Environmental Science and Technology*, 39 (21), 8434–8441, 2005.
Improving Data Quality for Environmental Fate Models: A Least-Squares Adjustment Procedure for Harmonizing Physicochemical Properties of Organic Compounds

Résumé

Pour modéliser l'évolution de polluants environnementaux, il est nécessaire de connaître leurs propriétés physico-chimiques, en particulier, la solubilité dans l'eau, la solubilité dans l'octanol, la constante de Henry et les coefficients de partition octanol-air et octanol-eau. Les valeurs trouvées dans la littérature présentent souvent des incohérences. Les auteurs proposent une nouvelle méthode pour améliorer leur cohérence en tenant compte de certaines conditions thermodynamiques qui les lient. L'article considère 16 composants organiques et démontre l'amélioration de la nouvelle méthode par rapport aux pratiques usuelles.

Principes

Le quotient de fugacité (f) correspond au rapport de la solubilité de la phase solide (S_S) et de la phase liquide refroidie (S_L) : $f = S_S/S_L$. Ce quotient intervient pour relier la solubilité dans l'air (S_A), dans l'eau (S_W) et dans l'octanol (S_O), avec les relations suivantes :

$$S_{AL} = \frac{S_{AS}}{f}, \qquad S_{WL} = \frac{S_{WS}}{f}, \qquad S_{OL} = \frac{S_{OS}}{f}. \tag{1}$$

En supposant un bon contact entre le soluté et le solvant, les coefficients de partition air–eau (K_{AW}), octanol–eau (K_{OW}) et octanol–air (K_{OA}) sont en équilibre.
Idéalement :

$$\log(K_{AW}) - \log(K_{OW}) + \log(K_{OA}) = 0. \tag{2}$$

Si l'on dispose des trois solubilités et des trois coefficients, on peut écrire les conditions thermodynamiques suivantes. Idéalement :

$$\log(S_{AL}) - \log(S_{WL}) \; - \; \log(K_{AW}) = 0, \tag{3}$$
$$\log(S_{WL}) - \log(S_{OL}) \; + \; \log(K_{OW}) = 0, \tag{4}$$
$$-\log(S_{AL}) + \log(S_{OL}) \; - \; \log(K_{OA}) = 0. \tag{5}$$

Hypothèses de l'exercice

L'article ne mentionne que les valeurs compensées de solubilité et de coefficients de partitions. Il ne donne pas explicitement d'écarts-types. Les hypothèses suivantes sont donc adoptées pour la résolution de l'exercice :

1. Les observations suivantes pour le PCB-3 sont utilisées pour l'exercice. Ce ne sont pas des mesures originales mais elles sont réalistes :

PCB-3	Mesure	Unité		Mesure	Unité
S_{AS}	$126 \cdot 10^{-5}$	[mol/m^3]	S_{AL}	$178 \cdot 10^{-6}$	[mol/m^3]
S_{WS}	$145 \cdot 10^{-3}$	[mol/m^3]	S_{WL}	$200 \cdot 10^{-4}$	[mol/m^3]
S_{OS}	$696 \cdot 10^{1}$	[mol/m^3]	S_{OL}	$102 \cdot 10^{1}$	[mol/m^3]
$\log K_{AW}$	2.16	[−]			
$\log K_{OA}$	6.66	[−]			
$\log K_{OW}$	4.82	[−]			

Table 1: Valeurs fictives mesurées et unités associées

2. La précision des coefficients de partition est supposée uniforme :

$$\sigma_{\log K} = 0.1 \quad [\mathrm{mol/m^3}]$$

3. La précision des mesures de solubilité dépend de l'observation selon la relation suivante :

$$\sigma_{S_i} = 0.03 \cdot S_i$$

Formulation des conditions

Pour effectuer une compensation conditionnelle des observations, il est nécessaire de former des conditions indépendantes reliant les observations. Les équations 1, 2, 3, 4 et 5 relient l'ensemble des observations entre elles. Former des conditions avec :

1. les équations thermodynamiques et les rapports de solubilité.
2. les équations thermodynamiques et les coefficients de partition.
3. les équations thermodynamiques, les coefficients de partition et les rapports de solubilité.

Astuces pour la résolution en Python

- Le fichier `fct_chim_trous.py` contient une fonction `fct_chimie` à compléter qui permet de calculer un vecteur $\mathbf{w}$ d'écarts de fermeture à partir d'un vecteur d'observations $\boldsymbol{\ell}$. Elle peut être utile pour calculer des dérivées numériques.
- Le fichier `covmat2corrmat.py` permet d'extraire un vecteur d'écarts-types et une matrice de corrélation associés à une matrice de covariance donnée en argument de la fonction `covmat2corrmat`.
- Le fichier `jacobian.py` contient une fonction `compute_jacobian` qui permet de calculer la matrice des dérivées partielles $\mathbf{B}$ par dérivation numérique.
- Le fichier `matrix_heatmap.py` contient une fonction `plot_heatmap` qui permet d'afficher un graphe sous forme de carte de chaleur d'une matrice. Il peut être utile pour analyser des corrélations entre paramètres.

Démarche générale

Il vous est demandé d'effectuer une compensation conditionnelle des observations en Python en suivant la marche à suivre présentée ci-dessous.

Calcul de la compensation conditionnelle

1. Former le vecteur $\boldsymbol{\ell}_{(n,1)}$ des observations.
2. Quelle est la surdétermination r ?
3. Poser le modèle fonctionnel en formant les conditions reliant l'ensemble des observations:
$$\mathbf{f}_{(r,1)}(\boldsymbol{\ell} - \hat{\mathbf{v}}) = \mathbf{0}_{(r,1)}$$
4. Calculer les écarts de fermeture $\mathbf{w} = \mathbf{f}(\boldsymbol{\ell})$ et analysez les résultats obtenus.
5. Former la matrice $\mathbf{B}_{(r,n)}$ des dérivées partielles, il est possible de la remplir de 2 manières en Python, choisir la plus adaptée.
6. Former la matrice $\mathbf{K}_{\ell\ell}$ de covariance des observations.
7. Former la matrice des cofacteur des observations:
$$\mathbf{Q}_{\ell\ell} = \frac{1}{\sigma_0^2} \cdot \mathbf{K}_{\ell\ell}$$
8. Calculer le vecteur des résidus compensés:
$$\hat{\mathbf{v}} = \mathbf{Q}_{\ell\ell} \cdot \mathbf{B}^T \cdot (\mathbf{B} \cdot \mathbf{Q}_{\ell\ell} \cdot \mathbf{B}^T)^{-1} \cdot \mathbf{w}$$
9. Calculer le vecteur des observations compensées:
$$\hat{\boldsymbol{\ell}} = \boldsymbol{\ell} - \hat{\mathbf{v}}$$
10. Calculer le vecteur $\hat{\mathbf{w}}$ des écarts de fermeture compensés.

Analyse des résultats

1. Analyser et comparer les vecteur $\boldsymbol{\ell}$ et $\hat{\boldsymbol{\ell}}$.
2. Analyser et comparer les vecteur $\mathbf{w}$ et $\hat{\mathbf{w}}$.
3. Analyser les résidus $\hat{\mathbf{v}}$ en calculant:
$$\mathbf{Q}_{\hat{v}\hat{v}} = \mathbf{Q}_{\ell\ell} \cdot \mathbf{B}^T \cdot (\mathbf{B} \cdot \mathbf{Q}_{\ell\ell} \cdot \mathbf{B}^T)^{-1} \cdot \mathbf{B} \cdot \mathbf{Q}_{\ell\ell}$$
Que peut-on utiliser pour simplifier l'analyse de la $\mathbf{Q}_{\hat{v}\hat{v}}$?
4. Analyser les observations compensées en calculant:
$$\mathbf{Q}_{\hat{\ell}\hat{\ell}} = \mathbf{Q}_{\ell\ell} - \mathbf{Q}_{\hat{v}\hat{v}}$$
Que peut-on utiliser pour simplifier l'analyse de la $\mathbf{Q}_{\hat{\ell}\hat{\ell}}$?
5. Analyser la précision en calculant l'écart-type de l'unité de poids:
$$\hat{\sigma}_0 = \sqrt{\frac{\hat{\mathbf{v}}^T \cdot \mathbf{P} \cdot \hat{\mathbf{v}}}{r}}$$
Avec $\mathbf{P} = \mathbf{Q}_{\ell\ell}^{-1}$.