

Introduction aux problèmes de calage de paramètres.

J.-B. Blanchard^{1,a}, G. Damblin^{1,b}, M. Baudin^{2,c}

¹DES-Service de thermo-hydraulique et de mécanique des fluides (STMF), CEA, Université Paris-Saclay, F-91191, Gif-sur-Yvette, France

²EDF R&D, Quai Watier, 78400 Chatou, France

Résumé Cette note introduit les notions de calage de paramètres. Plusieurs techniques sont présentées, dépendant des hypothèses sur le modèle et sur les incertitudes, allant de la méthode des moindres carrés aux techniques de calage bayésiennes.

1 Introduction

Dans le contexte de la qualification d'un modèle, ou code reproduisant un comportement physique, le calage¹ est une étape importante de l'analyse de quantification d'incertitude, mais il ne faut pas le confondre avec la validation [3] :

- validation : détermination du degré pour lequel un modèle est une représentation précise du monde réel, vis-à-vis des utilisations prévues par ce dernier.
- calage : amélioration de l'accord du modèle par rapport à des observations de référence par ajustement des paramètres implémentés dans ce dernier.

La question sous-jacente à la validation est : « Quel degré de confiance accorder au modèle au vue de l'écart entre ses prédictions et la réalité physique ? » tandis que celle du calage est : « Étant donné le modèle, quelles valeurs des paramètres minimisent l'écart entre les observations et les prédictions, sous les hypothèses d'incertitude choisies ? ».

Le modèle est la fonction mathématique $f_\theta : \mathbb{R}^{n_x} \rightarrow \mathbb{R}^{n_y}$. Le vecteur d'entrée \mathbf{x} contient les variables d'entrées du problème et $\theta \in \Theta \subset \mathbb{R}^p$ un vecteur constant inconnu. Le modèle est supposé déterministe (*i.e.* $f_\theta(\mathbf{x})$ est constant à \mathbf{x} et θ fixés), utilisé en boîte noire (*i.e.* calage non-intrusif) et nous poserons $n_Y = 1$ pour simplifier.

L'hypothèse standard du calage probabiliste est que les observations sont la somme de la sortie du modèle et d'une variable aléatoire :

$$z = f_\theta(\mathbf{x}) + \varepsilon, \quad (1)$$

où ε est une variable aléatoire d'espérance nulle, nommée *résidu*. Cette dernière modélise l'écart entre la prédiction du modèle et l'observation. Cet écart résulte en général de plusieurs sources d'incertitude :

- expérimentales, affectant les observations z ;
- de modèle, affectant la fonction f_θ .

a. e-mail : jean-baptiste.blanchard@cea.fr

b. e-mail : guillaume.damblin@cea.fr

c. e-mail : michael.baudin@edf.fr

1. appelé aussi problème inverse [1] ou assimilation de données [2] en fonction du contexte et des méthodes choisies.

Soit $\mathcal{C} = \{(\mathbf{x}_i, z_i)\}_{i \in [1, n]}$ l'ensemble des données de référence où n est le nombre d'observations et en ne supposant aucune incertitude sur \mathbf{x}_i , alors les observations peuvent s'écrire $z_i = f_\theta(\mathbf{x}_i) + \varepsilon_i$ pour $i \in [1, n]$. L'objectif du calage est de déterminer les valeurs de θ permettant de rapprocher les observations d'une part et les prédictions du modèle d'autre part.

Il peut arriver que le problème de calage possède une infinité de solutions [4], comme, par exemple, lorsque le modèle f_θ ne dépend que de la somme $\theta_1 + \theta_2$. Dans ce cas, tous les couples de valeurs (θ_1, θ_2) ayant la même somme $\theta_1 + \theta_2$ produisent la même prédiction en sortie de f_θ . Ainsi le calage est associé au problème de son identifiabilité [5].

Pour illustrer certaines techniques, un cas jouet sera utilisé, en définissant le modèle linéaire $f_\theta(x) = \theta_0 + \theta_1 x$ où $x \in [-2.5, 2.5]$ et $(\theta_0, \theta_1) = (-0.45, 0.35)$. Les observations sont un ensemble de $n = 30$ points de mesure dont l'estimation est bruitée par le bruit gaussien $\varepsilon \sim \mathcal{N}(0, \sigma = 0.4)$ ².

Cette note va tout d'abord introduire les méthodes de moindres carrés en § 2, puis les méthodes bayésiennes en § 3 avant de conclure en § 4.

2 Méthode des moindres carrés

Les méthodes décrites ici font l'hypothèse que le paramètre θ possède une valeur déterministe inconnue. Dans une approche probabiliste du calage, il est d'usage d'utiliser l'hypothèse de normalité des résidus :

$$\varepsilon_i \sim \mathcal{N}(0, \sigma_{\varepsilon_i}^2) \quad \text{pour } i = 1, \dots, n,$$

où la variance $\sigma_{\varepsilon_i}^2$ (supposée connue) peut quantifier les deux sources d'incertitude discutées en § 1. L'équation 1 implique que chaque sortie observée est la réalisation d'une variable aléatoire conditionnée aux paramètres θ :

$$z_i \sim Z_i | \theta := \mathcal{N}(f_\theta(\mathbf{x}_i), \sigma_{\varepsilon_i}^2) \quad (2)$$

2. C'est l'hypothèse d'homoscédasticité : la variance des erreurs stochastiques est la même pour chaque observation ($\sigma_{\varepsilon_i} = \sigma_\varepsilon \forall i \in [1, n]$).

2.1 Maximum de vraisemblance

À partir de l'équation 2, la densité de probabilité des observations de référence \mathcal{C} peut s'écrire :

$$L(\mathbf{z}|\theta) = \prod_{i=1}^n \ell(z_i|\theta) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma_{\varepsilon_i}} e^{-\frac{1}{2}\left(\frac{z_i - f_{\theta}(\mathbf{x}_i)}{\sigma_{\varepsilon_i}}\right)^2} \quad (3)$$

Les observations z_i étant données, la densité de probabilité des réalisations de \mathcal{C} est d'autant plus élevée que le paramètre θ est proche de la solution. Le maximum de vraisemblance consiste à considérer que l'équation 3 est une fonction du vecteur θ afin de maximiser la densité $L(\mathbf{z}|\theta)$, ce qui s'écrit $\hat{\theta}_{MV} = \operatorname{argmax}_{\theta} L(\mathbf{z}|\theta)$. Le logarithme de la vraisemblance mène alors au problème de moindres carrés :

$$\hat{\theta}_{MV} = \operatorname{argmin}_{\theta} \sum_{i=1}^n \left(\frac{z_i - f_{\theta}(\mathbf{x}_i)}{\sigma_{\varepsilon_i}} \right)^2.$$

2.2 Minimisation de fonction coût

Cette approche, usuelle, consiste à minimiser une fonction de coût, définie le plus souvent par des différences quadratiques (d'où le nom *Least-Squares* ou LS) calculées à partir de \mathcal{C} . Il existe différentes formes de fonction :

- classique³ : $\hat{\theta}_{OLS} = \operatorname{argmin}_{\theta} \sum_{i \in [1, n]} (z_i - f_{\theta}(\mathbf{x}_i))^2$;
- pondérée⁴ : $\hat{\theta}_{WLS} = \operatorname{argmin}_{\theta} \sum_{i \in [1, n]} w_i (z_i - f_{\theta}(\mathbf{x}_i))^2$.
- relative : $\hat{\theta}_{relLS} = \operatorname{argmin}_{\theta} \sum_{i \in [1, n]} (z_i - f_{\theta}(\mathbf{x}_i))^2 / z_i^2$.

Ces définitions ne sont pas orthogonales, en effet, si $\alpha \in \mathbb{R}$ et $\{w_i\}_{i \in [1, n]} = \alpha$, alors $\hat{\theta}_{WLS} = \hat{\theta}_{OLS}$ (c'est le cas dans l'hypothèse d'homoscédasticité où $\alpha = \sigma_{\varepsilon}^{-2}$). De même, si $\alpha \in \mathbb{R}$ et $\{w_i = (\alpha \% \times z_i)^{-1}\}_{i \in [1, n]}$ alors $\hat{\theta}_{WLS} = \hat{\theta}_{relLS}$ ⁵. Finalement, si $\{w_i = \sigma_{\varepsilon_i}^{-2}\}_{i \in [1, n]}$, alors $\hat{\theta}_{MV} = \hat{\theta}_{WLS}$ ce qui rend l'estimation simple dans le cas linéaire.

2.3 Cas linéaire

Si le modèle est linéaire, alors $f_{\theta}(\mathbf{x}) = h^T(\mathbf{x})\theta$ où $h(\mathbf{x})$ est le vecteur des régresseurs. Si besoin, on peut ajouter un composante "virtuelle" $\theta_0 = 1$ pour intégrer un terme constant dans la régression. La *matrice de conception* est

$$H = [h(\mathbf{x}_1), \dots, h(\mathbf{x}_n)]^T \in M_{n, p}(\mathbb{R}).$$

Ses colonnes définissent le sous-espace sur lequel est projeté f_{θ} . En posant $\Omega = \operatorname{diag}(w_1, \dots, w_n)$, la matrice des poids, la solution est [6] :

$$\hat{\theta}_{WLS} = (H^T \Omega H)^{-1} H^T \Omega \mathbf{z} \quad (4)$$

3. le O est ajouté pour *Ordinary*

4. le W est ajouté pour *Weighted*

5. cette forme est utile quand les erreurs sont multiplicatives.

quand $n \geq p$.

Cet estimateur est, selon le théorème de Gauss-Markov, optimal et sans biais si le résidu est non-biaisé et dans l'hypothèse d'homoscédasticité. Cette conclusion est vraie quelle que soit la distribution de ε , mais l'hypothèse de normalité permet de vérifier la cohérence des estimations, *a posteriori*, en analysant la distribution des résidus. De plus, même si les estimateurs sont ponctuels, l'hypothèse gaussienne fait que l'estimateur est lui-même gaussien, d'espérance $\mathbb{E}[\hat{\theta}] = \theta$ et de variance $\mathbb{V}[\hat{\theta}] = (H^T \Omega H)^{-1}$. Cette dernière est utile pour définir l'intervalle de confiance de niveau $(1-\alpha)\%$:

$$IC_{1-\alpha}^i = \left[\hat{\theta}_i - z_{1-\alpha/2} \sqrt{\mathbb{V}[\hat{\theta}_i]}, \hat{\theta}_i + z_{1-\alpha/2} \sqrt{\mathbb{V}[\hat{\theta}_i]} \right]$$

pour $i \in [1, p]$. Cet intervalle se comprend ainsi : en répétant cette estimation avec un grand nombre de jeux de référence de taille n , la valeur θ_i sera dans $IC_{1-\alpha}^i$ $(1-\alpha)\%$ du temps.

Le problème d'identifiabilité introduit en § 1, consiste à vérifier que la matrice \mathbf{H} est de rang plein, *i.e.* $\operatorname{rank}(\mathbf{H}) = p$ et qu'elle est bien conditionnée, ce qui veut dire que la solution du problème de moindres carrés linéaires est unique et peu sensible aux erreurs d'observation.

2.4 Cas général

Si le modèle linéaire n'est pas une bonne approximation de $\hat{\theta}$, il faut utiliser des techniques d'optimisation numérique pour résoudre le problème de moindres carrés. Les algorithmes sont alors multiples, dépendant principalement :

- du nombre de paramètres à caler ;
- du comportement mathématique attendu (continuité, présence de minima locaux, ...) ;
- du temps de calcul nécessaire aux estimations.

Ces techniques sont discutées plus en détails dans [7].

La figure 1 présente le résultat des méthodes de moindres carrés appliquées au cas jouet. L'hypothèse d'homoscédasticité implique que la méthode relative ($\hat{\theta}_{relLS}$) n'est pas adaptée. Les paramètres $\hat{\theta}_{OLS}$ et $\hat{\theta}_{WLS}$ sont cohérents avec les valeurs vraies, tout comme ceux obtenus par maximum de vraisemblance : $(\hat{\theta}_0, \hat{\theta}_1) = (-0.45 \pm 0.07, 0.37 \pm 0.05)$ ⁶.

3 Méthodes Bayésiennes

La probabilité d'un événement peut être décrite comme la limite de sa fréquence d'occurrence ou comme la quantification d'un jugement personnel sur la réalisation de cet événement. Ainsi, en lançant une pièce, la probabilité $\mathbb{P}[\text{pile}]$ est soit le résultat moyen d'un grand nombre de lancers (factuelle, mais dépendant de la statistique), soit la conviction que cette dernière est équilibrée ou non (jugement *a priori*).

6. Les incertitudes sont données à 1σ donc pour $\alpha = 31.7\%$

Soit (W, Z) un vecteur aléatoire de loi jointe $f_{(W,Z)}(w, z)$ et de lois marginales $f_W(w)$ et $f_Z(z)$. La règle de Bayes stipule alors que :

$$f_{W|Z}(w|z) = \frac{f_{Z|W}(z|w) \times f_W(w)}{f_Z(z)}, \quad (5)$$

où $f_{W|Z}(w|z)$ (respectivement $f_{Z|W}(z|w)$) est la densité de probabilité de W sachant que z est réalisé (et inversement) ⁷.

L'équation 5 implique que la densité de probabilité de la variable aléatoire θ est :

$$\pi_{post}(\theta|\mathbf{z}) = \frac{L(\mathbf{z}|\theta)\pi_{prior}(\theta)}{\pi(\mathbf{z})} \propto L(\mathbf{z}|\theta)\pi_{prior}(\theta). \quad (6)$$

Le terme $L(\mathbf{z}|\theta)$ est la vraisemblance des observations connaissant θ , $\pi_{prior}(\theta)$ est la densité de probabilité de θ *a priori*, appelée *prior*, $\pi(\mathbf{z})$ est la vraisemblance marginale des observations, considérée constante ⁸. La loi *a priori* est dite *propre* si elle est intégrable, *impropre* sinon ⁹. Les écritures sont allégées en notant $\pi(\theta|\mathbf{z})$ pour $\pi_{post}(\theta|\mathbf{z})$ et $\pi(\theta)$ pour $\pi_{prior}(\theta)$. Le choix du *prior* est crucial et doit reposer sur des contraintes physiques du modèle, des avis d'expert ou tout autre information disponible. Si ce n'est pas le cas, il est possible d'utiliser des lois *a priori* non informatives pour que le calage repose essentiellement sur les observations.

3.1 Calage linéaire gaussien

Dans le cas du modèle linéaire gaussien (utilisant les notations de § 2.3 et en définissant comme $\{w_i = \sigma_{\epsilon_i}^{-2}\}_{i \in [1,n]}$ les poids de Ω le *prior* est $\theta \sim \mathcal{N}(m_\theta, \Sigma_\theta)$ et le *posterior* $\pi(\theta|\mathbf{z}) \sim \mathcal{N}(m_\theta^{post}, \Sigma_\theta^{post})$ où

$$m_\theta^{post} = \left(\Sigma_\theta^{-1} + H^T \Omega H \right)^{-1} \left(m_\theta^T \Sigma_\theta^{-1} + \mathbf{z}^T \Omega H \right)^T \quad (7)$$

et

$$\Sigma_\theta^{post} = \left(\Sigma_\theta^{-1} + H^T \Omega H \right)^{-1}. \quad (8)$$

L'équation 7 est l'estimateur du maximum *a posteriori* (ou estimateur *MAP*).

Le *prior* non informatif de Jeffreys peut être utilisé : c'est un *prior* plat impropre ($\pi(\theta) \propto 1$) [8], dont le *posterior*, dans le cadre linéaire, est gaussien et dont les paramètres sont ceux des équations 7 et 8 sans les contributions Σ_θ (donc équivalents aux $\hat{\theta}$ et $\mathbb{V}[\hat{\theta}]$ de § 2.3).

7. On parle alors de loi conditionnelle.

8. Elle ne dépend pas des valeurs de θ mais de sa loi *a priori* ($\pi(\mathbf{z}) = \int_\theta L(\mathbf{z}|\theta)\pi_{prior}(\theta)d\theta$) et joue un rôle de normalisation.

9. Un *prior* impropre s'obtient comme une limite de *prior* propre (en faisant tendre la variance vers l'infini par exemple). Il faut toutefois vérifier que le *posterior* obtenu, est propre.

3.2 Assimilation de données

L'assimilation de données est un cadre d'analyse pour la combinaison d'un modèle et d'observations afin d'estimer l'état d'un système. Cela peut aussi servir au calage des paramètres en utilisant à la fois les mesures, l'estimation *a priori* (appelée « ébauche ») et leurs incertitudes [2, 9].

Les notations diffèrent de celles utilisées jusqu'ici, mais elles sont conventionnelles : \mathbf{x}_t représente les paramètres idéaux (ou vrais), \mathbf{x}_b l'ébauche (ou valeurs *a priori*), \mathbf{y}_o les observations et \mathbf{x}_a les paramètres optimaux (résultat de l'analyse). Dans le cas statique, la simulation et l'observation sont rassemblées en une fonction H (linéaire ou non) transformant des entrées \mathbf{x} en un résultat comparable aux observations. La différence $\mathbf{y} - H(\mathbf{x})$ est appelée *innovation* si $\mathbf{x} = \mathbf{x}_b$ et *résidus d'analyse* si $\mathbf{x} = \mathbf{x}_a$.

Les incertitudes pour cette méthode sont les suivantes :

- l'ébauche diffère de la valeur vraie par $\boldsymbol{\epsilon}_b$ de moyenne $\bar{\boldsymbol{\epsilon}}_b$ et de matrice de covariance \mathbf{B} ,
- les observations diffèrent des prédictions $H(\mathbf{x}_t)$ par $\boldsymbol{\epsilon}_o$, de moyenne $\bar{\boldsymbol{\epsilon}}_o$ et de matrice de covariance \mathbf{R} ,

où $\mathbf{B} = \mathbb{E}[(\boldsymbol{\epsilon}_b - \bar{\boldsymbol{\epsilon}}_b)(\boldsymbol{\epsilon}_b - \bar{\boldsymbol{\epsilon}}_b)^T]$ et $\mathbf{R} = \mathbb{E}[(\boldsymbol{\epsilon}_o - \bar{\boldsymbol{\epsilon}}_o)(\boldsymbol{\epsilon}_o - \bar{\boldsymbol{\epsilon}}_o)^T]$.

Par la suite on fait les hypothèses suivantes : $\bar{\boldsymbol{\epsilon}}_b = \bar{\boldsymbol{\epsilon}}_o = \mathbf{0}$, les matrices \mathbf{R} et \mathbf{B} sont définies positives et sont les matrices de covariance respectivement des erreurs d'observation et de l'ébauche (i.e. du *prior*). Les § 3.2.1 et 3.2.2 introduisent deux techniques par le prisme de l'inversion.

3.2.1 Filtre de Kalman (en inversion linéaire gaussienne)

Si la fonction H est linéaire (i.e. égale à sa jacobienne \mathbf{H}), l'estimation par filtrage est l'estimateur *Best Linear Unbiased Estimator* (BLUE) pour lequel le gain optimal \mathbf{K} et la matrice de covariance de l'analyse \mathbf{A} s'écrivent :

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1} \text{ et } \mathbf{A} = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{B}.$$

En assimilation par filtrage, le vecteur \mathbf{x}_a se déduit de \mathbf{x}_b :

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}(\mathbf{y} - \mathbf{H}(\mathbf{x}_b)) \quad (9)$$

où \mathbf{K} est la matrice de gain de Kalman. La matrice de Kalman est dit optimale car elle minimise l'erreur d'analyse.

L'assimilation de données par filtre de Kalman est comparable à une approche bayésienne dans laquelle les erreurs d'observations et le *prior* sur l'ébauche sont supposés gaussiens. Les équations 7 et 9 sont en effet mathématiquement équivalentes : l'estimateur du filtre de Kalman est l'estimateur *MAP* et sa matrice de covariance est donnée par l'équation 8. Il y a toutefois une différence notable dans la pratique : en assimilation de données, au lieu d'être traitées en un seul calcul, les estimations sont souvent mises à jour itérativement en égrainant les observations.

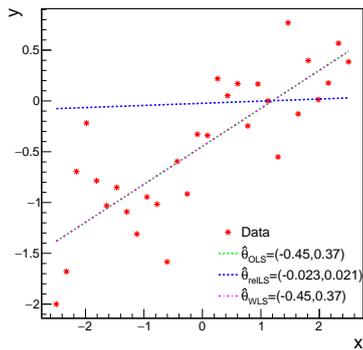


FIGURE 1: Minimisation des moindres carrés. Observations (rouge) et fonctions linéaires utilisant $\hat{\theta}_{OLS}$ (vert), $\hat{\theta}_{relLS}$ (bleu) et $\hat{\theta}_{WLS}$ (violet).

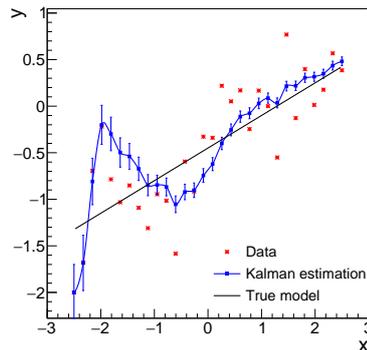


FIGURE 2: Calage bayésien par filtre de Kalman. Observations (rouge) et estimations par filtrage de Kalman (bleu) ainsi que la vraie fonction (noire).

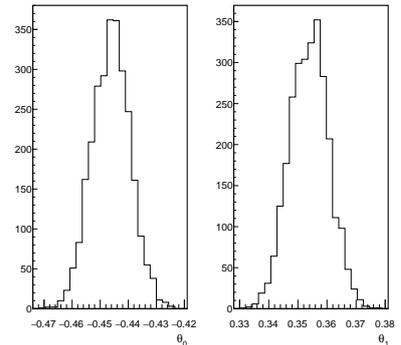


FIGURE 3: Distributions des paramètres estimés par l'algorithme de Metropolis-Hastings : θ_0 à gauche et θ_1 à droite.

3.2.2 3DVAR (en inversion non linéaire gaussienne)

Dans le cas général où la fonction H est non linéaire et $\boldsymbol{\varepsilon}_b$ et $\boldsymbol{\varepsilon}_o$ sont gaussiens, l'estimateur *MAP* mène à l'assimilation de données par la méthode variationnelle (3DVAR).

Pour tous $\boldsymbol{\rho}, \boldsymbol{\tau} \in \mathbb{R}^k$ et toute matrice de covariance $\mathbf{G} \in M_{k,k}(\mathbb{R})$, la distance de Mahalanobis entre $\boldsymbol{\rho}$ et $\boldsymbol{\tau}$ est $\|\boldsymbol{\rho} - \boldsymbol{\tau}\|_{\mathbf{G}} = \sqrt{(\boldsymbol{\rho} - \boldsymbol{\tau})^T \mathbf{G}^{-1} (\boldsymbol{\rho} - \boldsymbol{\tau})}$. En assimilation par la méthode 3DVAR, la fonction de coût à minimiser est :

$$J(\mathbf{x}) = \|\mathbf{y}_o - H(\mathbf{x})\|_{\mathbf{R}}^2 + \|\mathbf{x} - \mathbf{x}_b\|_{\mathbf{B}}^2$$

où $\|\cdot\|_{\mathbf{R}}$ et $\|\cdot\|_{\mathbf{B}}$ sont les normes de Mahalanobis associées respectivement aux matrices \mathbf{R} et \mathbf{B} .

Le minimum de la fonction J est l'estimateur *MAP*. Toutefois, contrairement au cas linéaire, la distribution de l'analyse \mathbf{x}_a est inconnue : pour la déterminer une approche plus générale par chaîne de Markov doit être employée (c.f. § 3.3). Une alternative consiste à linéariser la fonction H au point de l'estimateur *MAP*, menant à l'*approximation de Laplace*.

Dans le cas de problème dynamique, la fonction précédente est parfois étendue en 4DVAR [2]. Lorsqu'on considère un grand nombre de paramètres ou que les hypothèses gaussiennes ne sont plus vérifiées, ces méthodes ne peuvent plus s'appliquer. Une méthode de plus en plus utilisée dans ce contexte est le filtre de Kalman d'ensemble [9].

Les observations du cas jouet sont utilisées itérativement (par filtrage de Kalman) ce qui est montré dans la figure 2 où les prédictions sont de plus en plus proche des observations.

3.3 Cadre général : chaîne de Markov

Quand le modèle n'est pas linéaire en θ ou que les *priors* ne sont pas gaussiens, l'estimation des lois *a posteriori* n'est plus possible analytiquement. Dans ce cas, il est possible

d'obtenir des approximations par des méthodes appelées « Markov-Chain Monte Carlo » (MCMC). Une chaîne de Markov discrète est un processus aléatoire pour lequel le futur ne dépend que de l'instant présent, i.e. pour $t \in \mathbb{N}$: $\{X_{t+1} | X_t, \dots, X_0\} = \{X_{t+1} | X_t\}$. La définition peut être généralisée aux processus continus sur \mathbb{R}^p par l'introduction de noyaux de transition.

Une des chaînes de Markov classique est l'algorithme de Metropolis-Hastings dont le but est de générer un échantillon suivant une densité de probabilité cible. Pour cela, il utilise une méthode de génération à partir d'une réalisation déjà acceptée et d'une probabilité de mouvement permettant de s'assurer que la transition respecte les conditions de réversibilité [10]. Plusieurs critères permettent de juger de la qualité de l'échantillon obtenu.

La figure 3 montre les distributions résultantes des estimations des paramètres. Ces distributions sont proches de la loi gaussienne, de loi $\hat{\theta}_0 \sim \mathcal{N}(-0.45, 0.0068^2)$ et $\hat{\theta}_1 \sim \mathcal{N}(0.35, 0.0068^2)$ ce qui est en accord avec les valeurs utilisées pour la génération des observations.

4 Conclusion

Cette note a introduit le concept de calage des paramètres d'un modèle. Plusieurs techniques ont été abordées, des plus simples, comme les méthodes de moindres carrés, aux plus élaborées, comme les méthodes bayésiennes.

La table suivante présente la hiérarchie des méthodes, de la plus simple à la plus complexe et coûteuse.

	Linéaire	Non linéaire
\emptyset prior	LS linéaires	LS non-linéaires
prior gaussien	BLUE	3DVAR
prior non-gaussien	MCMC	MCMC

D'autres méthodes permettent de mieux prendre en compte la grande dimension (par exemple via le Monte-Carlo hamiltonien), d'effectuer un calage sans accès à la vraisemblance (par exemple les méthodes ABC), d'estimer la variance des erreurs σ^2 (dans le cas d'homoscédasticité)...

Références

1. A. Tarantola, *Inverse problem theory*. SIAM, 2005.
2. M. Asch, M. Bocquet, and M. Nodet, *Data assimilation. Methods, algorithms and applications*. SIAM, 2016.
3. T. G. Trucano, L. P. Swiler, T. Igusa, W. L. Oberkampf, and M. Pilch, "Calibration, validation, and sensitivity analysis : What's what," *Reliability Engineering & System Safety*, vol. 91, no. 10-11, pp. 1331–1357, 2006.
4. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems*. SIAM, 1996.
5. E. Walter and L. Pronzato, "Identification of parametric models," *Communications and control engineering*, vol. 8, 1997.
6. N. H. Bingham and J. M. Fry, *Regression. Linear Models in Statistics*. Springer, 2010.
7. J.-B. Blanchard, "Introduction aux problèmes d'optimisation numérique déterministe," *I3P book of notice*, 2020.
8. C. Bioche, *Approximation de lois impropres et applications*. PhD thesis, 2015. Thèse de doctorat dirigée par Druilhet, Pierre.
9. G. Evensen, *Data Assimilation - The Ensemble Kalman Filter*. SIAM, 2009.
10. S. Chib and E. Greenberg, "Understanding the metropolis-hastings algorithm," *The american statistician*, vol. 49, no. 4, pp. 327–335, 1995.