

Calibration de modèles physiques par inversion probabiliste sur plusieurs sources de données

Proposition de stage

Niveau d'étude : Stage de M2

Lieu : Service de Génie Logiciel pour la Simulation (SGLS), CEA Saclay

Laboratoire d'accueil : SGLS/Laboratoire d'Intelligence Artificielle et de sciences des Données (LIAD)

Durée du stage : 6 mois

Date de démarrage : à partir de Mars 2027

Possibilité de continuation en thèse après le stage : non

Encadrement :

- Clément Gauchy (clement.gauchy@cea.fr)

Le stage se déroulera sur le site du CEA de Saclay, des réunions régulières de suivi du stage seront prévues avec des collègues du CEA/DAM. Ce stage se déroulera également dans le cadre d'une collaboration scientifique entre le CEA et l'IFPEN.

Contexte du stage

Dans le cadre du développement d'outils de simulation, il est nécessaire de calibrer des paramètres de modèles physiques sur des données expérimentales en tenant compte des différentes sources d'incertitudes. Une première source d'incertitude vient de l'aspect expérimental des mesures physiques, et une deuxième source d'incertitude provient d'un écart entre le phénomène physique d'intérêt et le modèle physique choisi, que l'on appelle alors *erreur de modèle*. La prise en compte et la détermination de cette erreur de modèle est un sujet de recherche actif avec plusieurs modélisations possibles. En toute généralité, on considère qu'on dispose de n mesures du phénomène physique d'intérêt notées $(z_i)_{1 \leq i \leq n} \in \mathbb{R}$. À ces mesures physiques correspondent des observations des conditions expérimentales notées $(x_i)_{1 \leq i \leq n} \in \mathbb{R}^d$ qui sont des entrées du modèle physique noté g . Ce modèle dépend également de paramètres $\theta \in \mathbb{R}^p$ à calibrer. L'erreur de modèle peut être modélisée par une structure latente sur le paramètre θ Sargsyan et al. [2019]. Ainsi le modèle statistique se réécrit comme

$$\begin{cases} z_i = g(x_i, \theta_i) + \varepsilon_i \\ \theta_i \stackrel{iid}{\sim} \mathcal{N}(\mu, \Gamma) \\ \varepsilon_i \sim \mathcal{N}(0, \sigma_{\varepsilon_i}^2) \end{cases} \quad (1)$$

Dans ce type de modèles statistiques, θ_i est le *paramètre individuel* et (μ, Γ) les *paramètres de population*. L'estimation des paramètres de population peuvent s'effectuer par un algorithme Espérance Maximisation (EM) stochastique Barbillon et al. [2011] ou par une inférence bayésienne Fu et al. [2015].

On s'intéressera durant ce stage au cas spécifique où les paramètres individuels θ_i peuvent s'écrire sous la forme $\theta_i = (\theta_i^{(1)}, \theta_i^{(2)})$ et les mesures expérimentales $z_i = (z_i^{(1)}, z_i^{(2)})$ et telles que $(z_i^{(1)})$ soit informatif sur $\theta_i^{(1)}$ uniquement et $(z_i^{(2)})$ soit informatif sur le couple $(\theta_i^{(1)}, \theta_i^{(2)})$. De telles situations peuvent arriver sur des problèmes de physique ou de biologie Jacob et al. [2017]. On considérera durant le stage une loi de comportement élasto-plastique :

$$s(e) = \begin{cases} E e, & e \leq e_0, \\ E e_0 + K (e - e_0), & e > e_0, \end{cases} \quad (2)$$

avec e la déformation et s la contrainte. $s_0 = Ee_0$ est la limite d'élasticité et K le module tangent. Le modèle statistique s'écrit :

$$\begin{cases} z_i^{(1)} = E_i e_i^{(1)} + \varepsilon_i^{(1)}, & e_i^{(1)} \leq e_0 \\ z_i^{(2)} = E_i e_0 + K_i (e_i^{(2)} - e_0) + \varepsilon_i^{(2)}, & e_i^{(2)} > e_0 \\ E_i \sim \mathcal{N}(\mu^{(1)}, \Gamma^{(1)}) \\ K_i \sim \mathcal{N}(\mu^{(2)}, \Gamma^{(2)}) \\ \varepsilon_i^{(1)} \sim \mathcal{N}(0, \sigma_{\varepsilon_i^{(1)}}^2) \\ \varepsilon_i^{(2)} \sim \mathcal{N}(0, \sigma_{\varepsilon_i^{(2)}}^2) \end{cases} \quad (3)$$

$z_i^{(1)}$ et $z_i^{(2)}$ sont appelées respectivement *premier module* et *second module*. Les paramètres individuel de l'Équation (1) sont $\theta_i = (E_i, K_i)$. Le modèle physique g est ici la loi de comportement élasto-plastique définie à l'Équation (2). Les déformations $e_i^{(1)}$ et $e_i^{(2)}$ sont les conditions expérimentales $x_i = (e_i^{(1)}, e_i^{(2)})$ imposées à la i -ème éprouvette. e_0 est supposée connue. La mesure $z_i^{(1)}$ vient d'un essai élastique, donc le premier module n'informe que sur E_i . La mesure $z_i^{(2)}$ vient d'un essai plastique sur la même éprouvette i , qui est informatif sur le couple (E_i, K_i) .

Les objectifs du stage sont les suivants

- L'adaptation d'algorithmes d'Espérance Maximisation utilisés pour estimer les paramètres de population dans l'Équation (1)
- Étudier la robustesse des différents algorithmes proposés à la mauvaise spécification du second module (exemple : mauvaise connaissance de e_0 , loi de comportement différente sur le domaine plastique, ...).

Profil du candidat

Le profil recherché est celui d'un étudiant ou d'une étudiante ayant un niveau M2 de mathématiques appliquées avec une spécialisation en statistiques/machine learning. Des connaissances en physique seront appréciées mais non nécessaires.

Conditions de travail

Gratification de 1400€ mensuel, prise en charge par le CEA du pass Navigo à hauteur de 75%, Accès au CEA Saclay par bus de ville ou métro (bus 4606, 9119, 9111 et la ligne 18 au départ de Massy, ligne 4609 entre les Ulis et le CEA via Orsay) ou via navette bus privée affrétée par le CEA desservant de nombreuses communes d'île de France. Restaurant d'entreprise avec tarif avantageux pour les stagiaires ($\approx 2,75\text{€}/\text{repas}$).

Contact

Un CV, une lettre de motivation, et un relevé de notes de M1 sont à envoyer à clement.gauchy@cea.fr

Références

- Khachik Sargsyan, Xun Huan, and Habib N. Najm. Embedded model error representation for bayesian model calibration. *International Journal for Uncertainty Quantification*, 9(4) :365–394, 2019. ISSN 2152-5080.
- P. Barbillon, G. Celeux, G. Grimaud, Y. Lefebvre, and E. De Rocquigny. Nonlinear methods for inverse statistical problems. *Computational Statistics & Data Analysis*, 55(1) :132–142, 2011. ISSN 0167-9473. doi : 10.1016/j.csda.2010.05.030.
- Shuai Fu, Gilles Celeux, Nicolas Bousquet, and Mathieu Couplet. Bayesian inference for inverse problems occurring in uncertainty analysis. *International Journal for Uncertainty Quantification*, 5(1), 2015.

Pierre E. Jacob, Lawrence M. Murray, Chris C. Holmes, and Christian P. Robert. Better together? statistical learning in models made of modules. 2017. URL <https://arxiv.org/abs/1708.08719>.