

SUJET POUR LA PARTIE CONTRAT « PRO » - ALTERNANCE, Sept 2018 - mars 2019

« Création de Notebooks en support à une formation au traitement des incertitudes en simulation numérique »

De nombreux phénomènes physiques sont représentés par des équations déterministes qui conduisent, lors de la phase de modélisation numérique, à l'obtention de simulateurs numériques ou codes de calcul. Ces outils logiciels prennent en entrée un grand nombre de paramètres caractéristiques du phénomène étudié. L'ensemble de ces paramètres sont entachés d'une incertitude plus ou moins importante selon le degré de connaissance et de caractérisation du phénomène modélisé. Il est alors important de prendre en compte ces incertitudes et d'étudier, au travers d'une analyse de sensibilité ou de propagation des incertitudes, comment elles se répercutent sur les sorties du code. Cela permet de valider le modèle mathématique, physique ou numérique, d'orienter des efforts de caractérisation de certains paramètres, d'améliorer la compréhension du phénomène modélisé, ou encore, dans le cadre d'études d'analyse de risque menées au CEA, d'estimer des quantiles ou des probabilités associées à un événement « critique » d'un point de vue de la sûreté des réacteurs.

Afin de permettre un large déploiement de la démarche de traitement des incertitudes dans la simulation numérique, le CEA a mis en place des actions de formation aux outils et méthodes. Ainsi, une formation méthodologique d'une semaine destinée aux ingénieurs-chercheurs modélisateurs du CEA est organisée annuellement. En plus de cours théoriques, cette formation intègre des travaux pratiques (TP) sous le logiciel statistique R (<https://cran.r-project.org/>) afin d'illustrer les outils statistiques et les mettre en pratique sur des exemples analytiques. Cette formation est complémentaire des formations opérationnelles à la plateforme URANIE, développée par le CEA pour le traitement des incertitudes.

L'objectif aujourd'hui est de faire évoluer la formation méthodologique sur les deux points suivants :

- **Réalisation de notebooks couplés avec URANIE pour l'ensemble des TP.** Ces notebooks contiendraient à la fois une partie « texte et équations relatives aux méthodes », du code (R ou URANIE) pour exécuter les méthodes ainsi qu'une partie « visualisation des résultats ». Ces Notebooks offriront ainsi un environnement plus complet, convivial et pédagogique de travail. Pour cela, l'application web JUPYTER, déjà utilisée dans le cadre d'autres formations, sera utilisée (<http://jupyter.org/>).
- **Application de l'ensemble des outils sur un unique cas test.** Ce cas test, fil rouge de la formation, se voudrait plus proche des applications industrielles que les exemples analytiques actuellement utilisés. Ce modèle numérique plus « représentatif » des simulateurs utilisés dans nos applications industrielles devra nécessiter un temps de calcul négligeable pour permettre une mise en œuvre rapide des méthodes. Pour cela, plusieurs possibilités sont envisagées parmi lesquelles un simulateur de déformations d'un volcan à

partir des paramètres caractéristiques du réservoir magmatique, appliqué au Piton de la Fournaise. Ce modèle, développé par Valérie Cayol (vulcanologue au CNRS), a récemment été adapté par Durrande et Le Riche [2017] pour illustrer la démarche du traitement des incertitudes en simulation numérique. Le modèle et l'ensemble des codes sont disponibles sous R.

Après une revue des méthodes de traitement des incertitudes présentées dans la formation et une familiarisation aux outils (Jupyter et Uranie), l'étudiant aura en charge la création de ces notebooks illustratifs (et interactifs) ainsi que l'intégration dans ces notebooks du cas test retenu. Sur ce dernier point, il s'agira d'illustrer sur le cas test l'ensemble des étapes de traitement des incertitudes abordées durant la formation : quantification des incertitudes en entrée, création d'un plan d'expériences, réalisation d'une analyse de sensibilité de la sortie du simulateur, ajustement d'un métamodèle entre la sortie et les entrées et propagation des incertitudes pour estimer un quantile ou une probabilité. Pour cela, l'étudiant pourra s'appuyer sur l'ensemble des supports de cours de la formation ainsi que l'ensemble des TP actuellement en R.

Durrande, N. and Le Riche, R. [2017]. *Introduction to Gaussian Process Surrogate Models*. Lecture at 4th MDIS Formater workshop, slides as HAL report no. cel-01618068.

Discipline : Statistiques et probabilités, Informatique

Formation souhaitée : BAC + 4 (Alternance pour dernière année d'Ecole d'ingénieur ou Master 2)

Localisation : Laboratoire d'Etudes et Modélisations des Systèmes (DER/SESI/LEMS), CEA de Cadarache

Encadrant CEA: Amandine MARREL (DER/SESI/LEMS) – [amandine.marrel\[at\]cea.fr](mailto:amandine.marrel@cea.fr)

SUJET POUR LA PARTIE STAGE MASTER2 ASSOCIE, avril 2019 - sept 2019

« Plans d'expériences space-filling sous contraintes »

Dans le cadre du traitement des incertitudes, des outils de planification d'expériences sont utilisés pour explorer au mieux l'espace des paramètres incertains, dans le but par exemple d'ajuster un métamodèle, d'estimer des indices de sensibilité ou encore d'estimer une probabilité ou un quantile. Des plans de type *space-filling* garantissant une bonne couverture de l'espace sont classiquement utilisés en simulation numérique. On peut citer par exemple les suites à discrétance faible ou les plans hypercubes latins (plans LHS introduit par McKay et al. [1979]) maximisant un critère géométrique (Pronzato et Müller [2012], Damblin et al. [2013]).

Cependant, dans certaines études, il existe des contraintes que les variables d'entrée doivent respecter pour que la simulation soit possible (ou physiquement admissible). Même si ces contraintes sont connues (inégalité entre un sous-ensemble des paramètres d'entrée par exemple), cela pose le problème de la construction d'un plan d'expériences de type *space-filling* sous contraintes. En effet, le domaine de variation des paramètres incertains sous contraintes n'est souvent plus un hypercube et peut même ne plus être convexe. Les algorithmes de construction de plans *space-filling* ou de simples plans hypercubes latins ne peuvent plus être appliqués directement. **Dans le cadre du stage, plusieurs pistes pourront être envisagées pour pallier ce problème et construire un plan space-filling de n expériences respectant les contraintes imposées :**

- ⇒ **Approche « Fast Flexible Space-Filling Design »** : lorsque le domaine de respect des contraintes est connu (décrit par un système d'inégalités par exemple) ou facilement identifiable (calcul très rapide pour vérifier si un point appartient au domaine admissible), Lekivetz et Jones [2014] propose de générer un très grand nombre N de point dans le domaine admissible ($N \gg n$) et de réaliser ensuite un clustering des N points en imposant d'avoir n clusters (regroupement des points selon un critère géométrique). A chaque cluster est ensuite affecté un point représentatif du cluster, centroïde du cluster par exemple. On obtient ainsi les n points du design. A noter que dans le cas de domaines non-convexes, Lekivetz et Jones [2014] propose de choisir le point représentatif de chaque cluster selon une distance basée sur le critère *MaxPro* (Maximum Projection Design, Joseph et al. [2015]). Cette méthode illustrée en dimension 5 devrait en pratique être facilement étendue aux cas en dimension plus élevée (simulateurs numériques avec plusieurs dizaines de paramètres en entrée). Cependant, cette méthode ne semble pas garantir de bonnes propriétés statistiques des distributions marginales sur chaque paramètre, contrairement aux plans LHS (à vérifier selon l'utilisation du critère *MaxPro*). Pour pallier ce problème, la méthode permettant de choisir les points représentatifs des clusters pourrait être améliorée.

- ⇒ **Approche basée sur des « plans *maximin* sous contraintes »** : Auffray et al. [2012] ont développé une méthode d'échantillonnage de plans maximisant le critère *maximin* pour des espaces bornés de forme quelconque. L'algorithme proposé, basé sur le recuit simulé, consiste à sélectionner, à chaque étape, une paire de points très proches l'un de l'autre et de remplacer un de ces deux points par un point tiré aléatoirement. L'acceptation ou le rejet du plan contenant le nouveau point dépend d'une probabilité calculée à partir du critère. Comme précédemment, cette approche ne garantit pas de propriétés particulières sur les marginales. Les travaux de Petelet et al. [2009] permettant de générer des LHS sous contraintes pourraient être adaptés avec l'algorithme de Auffray et al. [2012] pour générer des LHS sous contraintes *maximin*. Par ailleurs, le critère *maximin* qui peut être sous-optimal pour les sous-projections (comme illustré par Damblin et al. [2013]), pourrait judicieusement être remplacé par un critère du type *MaxPro*, garantissant de meilleures propriétés pour les projections sur l'ensemble des sous-ensembles de facteurs.
- ⇒ **Approche basée sur une « méthode de répulsion pour construire des plans space-filling optimaux sous contraintes »** : Perrin et Cannamela [2017] ont très récemment proposé une méthode permettant de construire des plans space-filling optimaux (maximisant un critère d'occupation de l'espace) grâce à une méthode basée sur des répulsions. Cette approche peut être appliquée pour la construction de plans LHS space-filling et étendue aux domaines convexes bornés, non hypercubiques, en utilisant une fonction de régularisation.

L'objectif du stage est donc d'étudier ces différentes pistes, puis d'implémenter et tester les plus pertinentes sur des exemples analytiques (compromis entre réponse adaptée à la problématique, implémentation simple et extension facile au cas avec plusieurs dizaines de paramètres d'entrée). A plus long terme, les outils développés pourront être intégré dans la plateforme URANIE développée par le CEA pour le traitement des incertitudes.

Damblin, G., Couplet, M., and Iooss, B. [2013]. *Numerical studies of space filling designs: optimization of Latin hypercube samples and subprojection properties*. Journal of Simulation, 7:276–289.

Johnson, M. E., Moore, L. M., and Ylvisaker, D. [1990]. *Minimax and maximin distance designs*. Journal of Statistical Planning and Inference, 26:131–148.

Joseph, V. R., Gul, E., and Ba, S. [2015]. *Maximum projection designs for computer experiments*. Biometrika, 102(2):371–380.

Lekivetz, R. and Jones, B. [2015]. *Fast Flexible Space-Filling Designs for Nonrectangular Regions*. Quality and Reliability Engineering International, 31(5):829–837.

McKay, M., Beckman, R., and Conover, W. [1979]. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21:239–245.

Petelet, M., Iooss, B., Asserin, O. and Loredo, A. [2010]. *Latin hypercube sampling with inequality constraints*. *AStA Advances in Statistical Analysis*, 94: 325–339.

Perrin, G. and Cannamela, C. [2017]. *A repulsion-based method for the definition and the enrichment of optimized space filling designs in constrained input spaces*. *Journal de la Société Française de Statistique*, 58: 37-67.

Pronzato, L. and Müller, W.G. [2012]. *Design of computer experiments: space filling and beyond*. *Statistics and Computing*, 22: 681-701.

Discipline : Statistiques et probabilités, Informatique

Formation souhaitée : BAC +5 (Stage pour dernière année d'Ecole d'ingénieur ou Master 2)

Localisation : Laboratoire d'Etudes et Modélisations des Systèmes (DER/SESI/LEMS), CEA de Cadarache

Durée : 6 mois

Encadrant CEA: Amandine MARREL (DER/SESI/LEMS) – [amandine.marrel\[at\]cea.fr](mailto:amandine.marrel@cea.fr)