

Sujet de stage de fin d'études / M2 – Année universitaire 2021/2022

Analyse de sensibilité en présence de variables "gâchettes"

MOTS-CLÉS :

Variables gâchettes – Analyse de sensibilité – Apprentissage statistique – Méthodes à noyaux – Mesures d'importance

CONTEXTE GÉNÉRAL :

Au sein d'EDF R&D, le **département PRISME** (Performance, Risque Industriel et Surveillance pour la Maintenance et l'Exploitation) a pour mission de proposer des solutions innovantes pour une exploitation plus performante des différents moyens de production du groupe EDF : maîtrise des risques, amélioration de la sûreté, optimisation des performances technico-économiques, maîtrise de la durée de vie des installations et des composants. Au sein de PRISME, le groupe « Gestion d'Actifs, Incertitudes et Apprentissage statistique », dans lequel se déroulera le stage, contribue aux développements et à la diffusion de méthodes de traitement des incertitudes dans les outils de calcul scientifique.

La **thématique du "traitement des incertitudes"**¹ ("*uncertainty quantification*"²) désigne l'ensemble des méthodes probabilistes et statistiques dédiées à la modélisation des incertitudes, à leur quantification et à leur propagation dans les codes de simulation numérique. Portée par le **projet VIGIE** (Validation et Gestion des Incertitudes des Expériences numériques) et diffusée à travers le développement d'une plateforme dédiée (**logiciel OpenTURNS**³), cette thématique fait l'objet de travaux de recherche au sein du département PRISME et joue un rôle central et transversal au sein de nombreuses activités du groupe EDF.

PROBLÉMATIQUE INDUSTRIELLE :

Les codes de calcul utilisés pour la modélisation et la prévision des phénomènes physiques, ainsi que pour la simulation du comportement de systèmes complexes, dépendent typiquement de plusieurs variables (ou paramètres) en entrée, que l'on peut considérer de nature aléatoire (e.g., variables météorologiques dans le contexte de la prévision de la production d'une centrale photovoltaïque). Dans ce contexte, l'analyse de sensibilité⁴ vise à étudier comment les incertitudes en entrée d'un code de calcul influencent la variabilité de la quantité d'intérêt en sortie. Les objectifs principaux de cette analyse sont de deux sortes : d'une part, identifier, parmi toutes les variables, celles qui sont significativement influentes sur la sortie du code de calcul (tri qualitatif) ; d'autre part, classer les variables en fonction de leur influence (classement quantitatif). Cela permet, entre autres, de simplifier les codes considérés en ne prenant en compte que les variables les plus influentes tout en raffinant les modèles probabilistes associés aux variables identifiées comme potentiellement critiques.

Parmi les différents types de variables que les codes de calcul peuvent prendre en entrée, on se concentrera lors de ce stage sur les variables dites "gâchettes". Ces variables sont typiquement des variables discrètes ayant la particularité de présenter des dépendances hiérarchiques avec d'autres paramètres en entrée. En pratique, en fonction de la valeur (possiblement non numérique) prise par ces variables gâchettes, les paramètres influents sur la sortie du code de calcul seront différents. Par exemple, dans le cadre de l'analyse de la déformation d'assemblages de combustible nucléaire, la présence d'une grappe dans l'architecture du réacteur peut être vue comme une variable gâchette. En effet, en fonction de la valeur (ici, binaire) prise par cette variable (grappe présente ou non), d'autres paramètres tels que le type de grappe ou sa masse seront "enclenchés" et auront, de facto, une influence à quantifier sur la sortie du code de calcul.

Dans le contexte de l'analyse de sensibilité, la présence de variables gâchettes présente un verrou méthodologique fort, et il n'existe pas dans la littérature scientifique, à l'heure actuelle, de consensus sur des méthodes susceptibles de quantifier leur influence sur la quantité d'intérêt de façon précise et en tenant compte des dépendances hiérarchiques avec d'autres paramètres d'entrée. On pourra également noter que l'on se place dans un contexte industriel dans lequel les codes de calcul utilisés sont souvent coûteux en termes de temps d'évaluation (de plusieurs heures à plusieurs jours). Il sera, par conséquent, nécessaire d'identifier des solutions parcimonieuses en nombre d'appels au code. Identifier et proposer une solution à la problématique des variables gâchettes permettra d'améliorer les processus d'analyse de sensibilité sur un large éventail d'applications industrielles liées à la production d'électricité.

OBJECTIFS DU STAGE & PLANNING PRÉVISIONNEL :

L'objectif de ce stage est d'identifier et de proposer des solutions permettant d'effectuer une analyse de sensibilité sur des codes de calcul en présence de variables gâchettes (ainsi que de paramètres continus plus communément traités). Ce stage s'articulera sur plusieurs phases, définies ci-dessous :

1. Définition et formalisation analytique du problème d'analyse de sensibilité en présence de variables gâchette.
2. Étude bibliographique permettant d'identifier les pistes et les méthodes d'analyse de sensibilité prometteuses dans ce contexte.
3. Implémentation numérique des solutions identifiées et/ou proposées (langages envisagés : Python ou R).
4. Analyse des performances des solutions implémentées sur des cas jouets.
5. Application des méthodologies développées sur des jeux de données industriels et analyse des résultats.

¹ De Rocquigny E., Devictor N., Tarantola S. (Eds.). "Uncertainty in Industrial Practice: A Guide to Quantitative Uncertainty Management", Wiley, 2008.

² Ghanem R., Higdon D., Owhadi H. (Eds.). "Handbook of Uncertainty Quantification", Springer, 2017.

³ OpenTURNS: An Open source Initiative for the Treatment of Uncertainties, Risks'N Statistics (www.openturns.org).

⁴ looss, B. "Revue sur l'analyse de sensibilité globale de modèles numériques." *Journal de la Société Française de Statistique* 152.1 (2011): 3-25.

La possibilité d'utiliser les indices HSIC^{5,6} (Hilbert-Schmidt Independence Criterion) issus du domaine du machine learning à la présence de variables gâchettes constitue une première piste de travail identifiée comme prometteuse. En fonction de l'état d'avancement des travaux, d'autres pistes pourront être explorées.

PROFILS :

Étudiant.e de M2 (mathématiques appliquées / probabilités-statistiques / modélisation & simulation numérique) ou d'écoles d'ingénieur.e.s (généralistes avec majeure en mathématiques appliquées / probabilités-statistiques).

COMPÉTENCES REQUISES :

- Solides compétences en statistiques, analyse numérique et probabilités.
- Maîtrise des langages Python ou R. A minima, maîtrise d'un langage de calcul scientifique proche (e.g., Matlab) et forte motivation pour monter en compétence sur un de ces deux langages.
- Aisance dans la communication, orale et écrite, en Français et en Anglais.

APTITUDES PERSONNELLES SOUHAITÉES :

- Goût pour la recherche (méthodologies, concepts mathématiques, recherche bibliographique et applications industrielles).
- Ouverture d'esprit, curiosité et autonomie.

CONTACTS (ENCADRANTS INDUSTRIELS) :

Julien Pelamatti (julien.pelamatti@edf.fr), Vincent Chabridon (vincent.chabridon@edf.fr), Bertrand Iooss (bertrand.iooss@edf.fr)

DURÉE ENVISAGÉE :

6 mois à compter de février, mars ou avril 2022

NOTES ADDITIONNELLES :

Les travaux associés à ce stage s'inscrivent également dans le cadre d'un groupement d'intérêt scientifique (GIS) baptisé LaRTISSTE pour "Laboratoire de Recherche en Traitement des Incertitudes par apprentissage Statistique pour la Sécurité, la conception et la Transition Énergétique". Des échanges techniques et scientifiques avec d'autres membres des laboratoires académiques et industriels concernés sont envisagés.

LIEU :

EDF R&D – EDF Lab Chatou
Département PRISME (Performance, Risque Industriel et Surveillance pour la Maintenance et l'Exploitation)
Groupe P17 GAIA (Gestion d'Actifs, Incertitudes et Apprentissage statistique)
6 Quai Watier, 78 401, Chatou, France

⁵ Gretton, A. et al. "Measuring statistical dependence with Hilbert-Schmidt norms." *International conference on algorithmic learning theory*. Springer, Berlin, Heidelberg, 2005.

⁶ Da Veiga, S. "Global sensitivity analysis with dependence measures." *Journal of Statistical Computation and Simulation* 85.7 (2015): 1283-1305.