

Projets du département IMI

2022

1 Réseaux de neurones récurrents pour la prédiction d'un rayonnement thermique

Proposé par : Air Liquide

Contact : Benjamin Le Creurer (benjamin.le-creurer@airliquide.com)

Descriptif : Le transfert de chaleur dans des fours industriels haute température du type vapo-reformage est dominé par le rayonnement thermique. Pour être prédictif, les codes de calculs doivent donc utiliser un modèle de rayonnement le plus précis possible. Cependant, les modèles de rayonnement, même moins précis, sont très coûteux en termes de temps de calcul et d'espace mémoire.

Une approche du type réseau de neurones a été utilisée à partir des résultats d'un code 2D de rayonnement, et des premiers résultats ont été obtenus. L'objectif étant de remplacer le modèle physique des codes de R&D par un modèle appris, plus rapide.

Le principal objectif de ce projet serait de partir d'une base de données existante, et de construire un modèle de réseau de neurones récurrents, qui pourrait s'avérer plus adapté au problème initial (pour cela, on utilisera au choix une des bibliothèques ML de Python). Les résultats obtenus seront à comparer à ceux du réseau de neurones standard.

Selon l'avancée des étudiants, une deuxième étape serait (au choix) :

- d'implémenter un modèle de réseaux de neurones convolutif (CNN) et/ou graphique (GNN), en l'entraînant sur la même base de données, puis comparer les résultats ;
- de construire un modèle de prédiction, sans avoir recours aux bibliothèques ML de Python (l'idée étant de faciliter la réécriture du programme dans un langage autre que Python).

2 Modélisation multi-échelle pour la propagation électro-magnétique

Proposé par : Airbus

Contacts : Pierre Benjamin (pierre.benjamin@airbus.com) et Eric Duceau (eric.duceau@airbus.com)

Descriptif : L'objectif de ce projet est de préparer une modélisation multiéchelle dans le domaine de la propagation électromagnétique (équation des ondes). On considère un petit détail (vis-à-vis de la longueur d'onde) localisé sur une structure de grande taille. On cherche à effectuer un couplage entre la partie grande traitée par un schéma numérique connu (et qui est déjà disponible) avec un modèle « appris » (qu'il s'agira de produire) et ceci « proprement » (donc en comprenant les maths qui font marcher l'histoire !).

Il s'agit donc d'éviter des surmaillages pénalisants à cause de l'importance des calculs, et surtout délicats du point de vue des stabilités et erreurs numériques. La grande difficulté est de travailler sur une bonne formulation de couplage. Le projet de département commencera par la partie « facile » : la production du modèle réduit/appris et ses caractéristiques. Ensuite, on entre dans la recherche et il faudra réfléchir avec les chercheurs d'Airbus (même si les idées sont déjà bien en place).

Un déroulement éventuel du projet serait :

- Modéliser le problème par des éléments finis de frontière (une modélisation similaire est déjà disponible et peut donner les références nécessaires) ;
- Construire le « modèle appris » à partir des données synthétiques fournies par le code aux éléments finis ;
- Concevoir une campagne de tests sur des structures connues (par exemple des petites cavités) ;
- Analyser mathématiquement le couplage et les tests sur la maquette fournie.

3 Utilisation de techniques de Machine Learning pour le calcul d’empreinte carbone des entreprises

Proposé par : BNP Paribas

Contacts : Thibaut Heurtebize (thibaut.heurtebize@bnpparibas.com)

François Soupé (francois.soupe@bnpparibas.com)

Remarque : Projet commun avec le département SEGF.

Descriptif : Les entreprises sont responsables d’une grande partie des émissions de gaz à effet de serre. Selon leurs activités, ces émissions sont segmentées en trois « scopes » :

- Scope 1 : ce sont toutes les émissions directes de gaz à effet de serre. Cela inclut la combustion de carburants, les émissions de véhicules détenus par l’entreprise, et les émissions fugitives. Les émetteurs les plus importants sont ceux des secteurs de l’énergie, des services, des matériaux, et du transport.
- Scope 2 : il s’agit des émissions indirectes dues à la consommation d’électricité, de chauffage, de machines à vapeur et de systèmes de refroidissement. Les émetteurs les plus importants sont dans les secteurs des matériaux, de l’agro-alimentaire, des télé-communications, et de l’énergie.
- Scope 3 : cette dernière scope englobe toutes les autres émissions indirectes. Les secteurs les plus émetteurs sont par exemple ceux de l’énergie, de l’automobile, des matériaux, des services, et des biens d’équipement.

La BNP a développé plusieurs modèles permettant d’estimer l’empreinte carbone d’entreprises (Scope 1 et Scope 2). Ceci leur permet de ne pas dépendre des fournisseurs de données.

Le projet consiste en une étude approfondie de deux modèles. L’un est linéaire (Elastic Net) et l’autre est non-linéaire (X-Trees).

Il s’agira notamment d’étudier la sensibilité des résultats aux données, de comprendre comment les résultats peuvent être présentés, puis de les expliquer (notamment dans le cas du modèle non-linéaire).

4 Imagerie médicale par EIT et résolution du problème inverse par D-Bar

Proposé par : Gense Technologies

Contact : Adrien Touboul (adrien@gense.tech)

Remarque : Décalage horaire à prendre en compte pour les discussions à distance avec le porteur de projet (Hong Kong)

Descriptif : Nous vous proposons de travailler en collaboration avec une start-up Hong-Kongaise sur une méthode d’imagerie médicale de tomographie par impédance électrique [1]. La technique se base sur l’injection de courants électriques et de mesures de potentiels à la frontière du domaine Ω dont on veut acquérir une image. Les équations de Maxwell nous permettent d’affirmer que la distribution de potentiel électrique V à l’intérieur du corps suit une équation aux dérivées partielles elliptique du type

$$\operatorname{div}(\sigma(x)\nabla V(x)) = 0, \quad \forall x \in \Omega,$$

où σ représente la conductivité de l’organe traversé. La reconstruction de σ à partir de la mesure de V à la frontière $\partial\Omega$ pour différentes injections de courants permet d’obtenir une image de chaque organe en fonction

de leur conductivité. Ce problème rentre dans la catégories des problèmes inverses car il est mal posé : la fonction

$$H : \Lambda_{V,\sigma} \mapsto \sigma$$

qui associe $\Lambda_{V,\sigma}$, l'ensemble des mesures potentiels/courants sur la frontière à la conductivité du domaine σ , n'est pas continue. Elle est donc très sensible aux incertitudes de mesure. L'unicité reste tout de même vérifiée car sous certaines hypothèses, H est injective [2].

Nous nous intéressons dans ce projet à la méthode numérique de reconstruction dite 'D-Bar', basée sur une preuve constructive d'unicité du problème en 2D [2, 3]. Vos objectifs seront les suivants :

- Comprendre le problème et la démonstration d'unicité (au moins dans les grandes lignes).
- Implémenter la méthode en Python dans le projet opensource pyEIT (une implémentation Matlab existe déjà et pourra inspirer la version Python)
- Tester la reconstruction sur des conductivités simulées et comparer avec la référence.
- Tester sur des données réelles, acquises par la start-up.

References

- [1] David S. Holder. *Electrical Impedance Tomography: Methods, History and Applications*. CRC Press, 2004.
- [2] Adrian I. Nachman. Global uniqueness for a two-dimensional inverse boundary value problem. *Annals of Mathematics*, 1996.
- [3] Samuli Siltanen, Jennifer Mueller and David Isaacson. An implementation of the reconstruction algorithm of A. Nachman for the 2D inverse conductivity problem. *Inverse Problems*, 16(3):681, 2000.

5 Imagerie médicale par EIT et Deep Learning pour problèmes inverses

Proposé par : Gense Technologies

Contact : Adrien Touboul (adrien@gense.tech)

Remarque : Décalage horaire à prendre en compte pour les discussions à distance avec le porteur de projet (Hong Kong)

Descriptif : Nous vous proposons de travailler en collaboration avec une start-up Hong-Kongaise sur une méthode d'imagerie de médicale de tomographie par impédance électrique [1]. La technique se base sur l'injection de courants électriques et de mesures de potentiels à la frontière du domaine Ω dont on veut acquérir une image. Les équations de Maxwell nous permettent d'affirmer que la distribution de potentiel électrique V à l'intérieur du corps suit une équation aux dérivées partielles elliptique du type

$$\text{div}(\sigma(x)\nabla V(x)) = 0, \forall x \in \Omega,$$

où σ représente la conductivité de l'organe traversé. La reconstruction de σ à partir de la mesure de V à la frontière $\partial\Omega$ pour différentes injections de courants permet d'obtenir une image de chaque organe en fonction de leur conductivité. Ce problème rentre dans la catégories des problèmes inverses car il est mal posé : la fonction

$$H : \Lambda_{V,\sigma} \mapsto \sigma$$

qui associe $\Lambda_{V,\sigma}$, l'ensemble des mesures potentiels/courants, sur la frontière à la conductivité du domaine σ , n'est pas continue. Elle est donc très sensible aux incertitudes de mesure. L'unicité reste tout de même vérifiée car sous certaines hypothèses, H est injective [2].

Des travaux récents utilisent des méthodes issues du Deep Learning avec des données d'apprentissage issues de la simulation pour résoudre ce problème inverse [3]. Ces méthodes semblent prometteuses, pourvu qu'on les couple à du *transfer learning* pour les adapter aux données réelles, moins nombreuses et plus coûteuses à obtenir. Vos objectifs seront donc les suivants :

1. Reproduire les résultats de l'article *A learning-based method for solving ill-posed nonlinear inverse problems: a simulation study of lung EIT* [3].
 - Choisir la forme des données d'apprentissage et les générer via des simulations par éléments finis. La bibliothèque opensource pyEIT vous proposera un solveur facile à utiliser.
 - Mettre en place la pipeline et des réseaux de neurones décrits dans l'article et arriver à un niveau d'erreur comparable sur l'ensemble de test.
2. Identifier la façon dont l'architecture du réseau de neurone peut être adaptée pour le transfer learning [5, 4]. L'entraîner sur les données réelles mesurées par le dispositif développé par la start-up.

References

- [1] David S. Holder. *Electrical Impedance Tomography: Methods, History and Applications*. CRC Press, 2004.
- [2] Adrian I. Nachman. Global uniqueness for a two-dimensional inverse boundary value problem. *Annals of Mathematics*, 1996.
- [3] Jin Keun Seo, Kang Cheol Kim, Ariungerel Jargal, Kyoungun Lee, Bastian Harrach. A learning-based method for solving ill-posed nonlinear inverse problems: a simulation study of lung EIT. *SIAM journal on Imaging Sciences*, 12(3):1275–1295, 2019.
- [4] Jason Yosinski, Jeff Clune, Yoshua Bengio, Hod Lipson. How transferable are features in deep neural networks? Preprint, *arXiv:1411.1792*, 2014.
- [5] Chuanqi Tan, Fuchun Sun, Tao Kong, Wenchang Zhang, Chao Yang, and Chunfang Liu. A survey on deep transfer learning. In *International conference on artificial neural networks*, pp 270–279. Springer, 2018.

6 Optimisation de la recommandation de régimes alimentaires

Proposé par : GoodSesame

Contact : Matthieu Corcin (matthieu@goodsesame.com)

Descriptif : GoodSesame est une entreprise innovante visant à accompagner ses utilisateurs vers des régimes alimentaires sains, éthiques et éco-friendly, tout en respectant leur contraintes personnelles (budget, temps disponible, allergies et préférences alimentaires...) grâce à l'utilisation de techniques de Data Science. L'objet du projet consiste à proposer, implémenter (en Python) et tester des solutions permettant d'améliorer les performances de l'algorithme de recommandation en introduisant de nouvelles dimensions d'optimisation, tout en maîtrisant la perte de performance selon les autres dimensions.

Exemple de pistes de recherche :

- Utilisation optimale des ingrédients et minimisation du gaspillage alimentaire

- Prise en compte des préférences de l'utilisateur, déterminées de manière implicite au travers de ses interactions avec le service
- Minimisation de l'empreinte carbone des repas...

7 Development of a Numerical Solver for High-Fidelity Simulations of Wind Flow in Urban Areas

Proposé par : GRASP

Contact : Konstantin Kuznetsov (konstantin.kuznetsov@enpc.fr)

Remarque : Les échanges avec le porteur de projet se feront en anglais.

Descriptif : GRASP has a wide expertise in satellite remote sensing. The company is interested in the development of its own tools for the local scale urban modelling. One of the most demanding tool is a Navier-Stokes equations solver for the modelling of velocity fields in urban areas. We are proposing to develop a SIMPLEC algorithm to solve Navier-Stokes equation in the Boussinesq approximation. During this work we are going to use a relatively new framework called `ParallelStencil.jl`. This framework provides a unique interface to a distributed memory computer architecture, using either CPU or GPU. This approach promises lower computational costs (due to GPU calculations), and lower development cost (due to the use of the language of high abstraction JuliaLang). Built-in language tools allows to have an almost free access to automatic differentiation, which makes it possible to use advanced methods of Data Assimilation.

8 Représentation et visualisation de données de microscopie 3D en réalité augmentée

Proposé par : INRAE

Contacts : François Deslandes (francois.deslandes@inrae.fr)
Alain Trubuil (alain.trubuil@inrae.fr)

Description : Il est aujourd'hui possible d'acquérir de nombreuses données en microscopie 3D, 3D+Temps pour l'observation et l'analyse de processus biologiques, aux échelles intracellulaires et cellulaires. La réalité augmentée permet de visualiser avec moins de contraintes, dans l'environnement de l'utilisateur, et avec des grossissements importants des données généralement associées à des maillages. Elle permet également d'interagir efficacement avec ces données pour les explorer ou effectuer des sélections. Dans le cadre ce projet, nous souhaitons d'abord étudier des représentations parcimonieuses d'images 3D (*e.g.* [1]) et ensuite utiliser ces représentations pour la visualisation en réalité augmentée. Une seconde partie du projet sera consacrée à la visualisation fluide de données 3D+T. Il s'agira alors de proposer et d'évaluer des interpolations temporelles qui puissent être mises en œuvre en visualisation en réalité augmentée.

Applications potentielles : visualisation d'embryons en biologie végétale. Utilisation du casque de réalité augmentée Hololens II et du moteur Unity.

References

- [1] Bevan L Cheeseman, Ulrik Günther, Krzysztof Gonciarz, Mateusz Susik, and Ivo F Sbalzarini. Adaptive particle representation of fluorescence microscopy images. *Nature communications*, 9(1):1–13, 2018.

9 Approches de type Proxy et Machine Learning pour le calcul du SCR

Proposé par : Milliman

Contact : Adel Cherchali (adel.cherchali@milliman.com)

Descriptif : La directive Solvabilité II qui régit l'activité de l'assurance et de la réassurance en Europe a introduit de nouveaux défis computationnels. Tout d'abord, le pilier I de cette réglementation impose le calcul d'une Value at Risk (VaR) sur les pertes du portefeuille à horizon d'un an pour calculer l'exigence en capital requis (SCR) que devront immobiliser les assureurs. Par ailleurs, l'ORSA (Own Risk and Solvency Assessment) nécessite le calcul de fonds propres afin d'éviter la ruine d'une compagnie d'assurance sur tout un horizon de projection.

L'objectif de ce projet est d'investiguer les méthodes numériques d'approximation du SCR.

Dans un cadre plus général, le problème est de calculer la probabilité de grandes pertes d'un portefeuille financier. Ce calcul est particulièrement complexe car la valorisation du portefeuille d'une compagnie d'assurance vie n'admet pas de formules fermées et requiert un grand nombre de simulations Monte-Carlo. Ce type de problème nécessitant la simulation de facteur de risques (« simulations primaires ») jusqu'à une certaine date puis un deuxième jeu de simulations dites « secondaires » pour valoriser le portefeuille conditionnellement à chaque scénario primaire est appelé « Simulation dans les Simulations » [1], et reste actuellement trop coûteux en temps de calcul pour être utilisable en pratique.

Dans un premier temps, nous étudierons différentes méthodes de calibrages, en particulier *Curve Fitting* [2, 3], *Least-Square Monte-Carlo* (LSMC) [2, 5, 7] et *Replicating Portfolio* [4]. Pour cela, nous considérerons tout d'abord un portefeuille de type Black-Scholes avec taux d'intérêt constant (un seul facteur de risque). Puis, nous augmenterons progressivement le nombre de facteurs de risques dans le portefeuille qui comprendra plusieurs sous-jacents et un taux d'intérêt stochastique.

Dans un second temps, nous étudierons les approches de type Machine Learning [9] (réseaux de neurones [6, 8] et forêts aléatoires) sur un exemple en grande dimension.

Enfin, si le temps le permet, nous appliquerons l'ensemble des méthodes développées dans les modèles précédents à un cas industriel.

References

- [1] L. Devineau et S. Loisel. Construction d'un algorithme d'accélération de la méthode des « simulations dans les simulations » pour le calcul du capital économique Solvabilité II. *Bulletin Français d'Actuariat*, vol. 10, n. 17, pp. 188-221, 2009.
- [2] A. Pelsser and J. J. Schweizer, The Difference between LSMC and Replicating Portfolio in Insurance Liability Modeling, in *European Actuarial Journal*, vol. 6, n. 2, pp. 441-494, 2016.
- [3] A. Koursaris. A Comparison Between Curve Fitting and Least Squares Monte Carlo Techniques. Rapport technique, Barrie & Hibbert, 2011.
- [4] M. Cambou, D. Filipovic. Replicating portfolio approach to capital calculation. *Finance stoch.*, 22, 181-203, 2018.
- [5] A. S. Krah, Z. Nikolic, R. Korn. A least-squares Monte Carlo framework in proxy modeling of life insurance companies. *Risk*, 6.2:62, 2018.
- [6] M. Nilsson, E. Sandberg. *Application and Evaluation of Artificial Neural Networks in Solvency Capital Requirement Estimations for Insurance Products*, 2018.

- [7] T. O. Nteukam, F. Planchet, J. Ren. Internal Model in Life insurance: Application of Least Square Monte-Carlo in Risk Assessment. *Les cahiers de la recherche de l'ISFA*, n. 12, 2014.
- [8] A. S. Krah, Z. Nikolic, R. Korn. Least-Squares Monte Carlo for Proxy Modeling in Life Insurance: Neural Networks. *Risks*, vol. 8, n. 4, p. 116, 2020.
- [9] R. Gauville. *Projection du ratio de solvabilité : des méthodes de Machine Learning pour contourner les contraintes opérationnelles de la méthode des SdS*. Mémoire d'actuaire, EURIA, 2017.

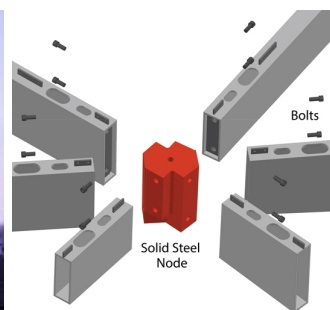
10 Des pavages de la sphère aux structures à double courbure à nœuds identiques

Proposé par : Laboratoire Navier (École des Ponts)

Contacts : Xavier Tellier (xavier.tellier@enpc.fr) et Cyril Douthe (cyril.douthe@enpc.fr)

Descriptif : Les structures à double courbure permettent d'allier performance mécanique et expressivité architecturale. Le principal obstacle à leur utilisation est la complexité qu'elles engendrent du point de vue de la fabrication des éléments constitutifs. Pour les structures de type Gridshell, cette complexité est pour beaucoup concentrée au niveau des assemblages, qui sont souvent les éléments les plus coûteux d'une telle structure. Dans ce projet, on cherchera à simplifier la fabrication des nœuds en les rendant identiques, ce qui permettrait de standardiser leur production.

On sait déjà obtenir une répétition des assemblages sur des objets simples, tels que les surfaces de révolution ou les polyèdres d'Archimède. Cependant, pour des formes plus générales, le problème s'avère très contraint, si bien que tenter d'optimiser une forme arbitraire sous contrainte « rendre tous les nœuds identiques » est en général infructueux. On propose donc une génération *bottom-up*, qui utilisera une nouvelle forme de dualité. Cette dualité permet de ramener le problème de conception au problème suivant : comment recouvrir une portion de sphère avec des quadrangles tous identiques ? Ce problème de géométrie sphérique n'a été étudié dans la littérature scientifique que dans certains cas particuliers. Il fait en particulier apparaître un problème combinatoire, qui consiste à construire un pavage de la sphère à partir de tuiles identiques, dont les dimensions vérifient certaines règles géométriques.



G1



G3



K13

Le projet portera sur la formulation et la résolution de ce problème. Les résultats pourront être traduits par la réalisation de prototypes/maquettes, en utilisant des méthodes de prototypage rapides tel que l'impression 3D.

11 Courbes de fragilité pour réacteur nucléaire

Proposé par : Phimeca Engineering

Contacts : Gaëtan Blondet (blondet@phimeca.com) et Sylvain Girard (girard@phimeca.com)

Descriptif : Phimeca travaille régulièrement avec l'IRSN (Institut de Radioprotection et de Sûreté Nucléaire) pour l'estimation de probabilités de défaillance des composants des centrales nucléaires françaises (parc EDF). Ces estimations sont réalisées dans le cadre des Études Probabilistes de Sûreté (EPS)¹, dans le cas où un séisme se produit dans les environs de la centrale. L'enjeu de cette démarche est de garantir un niveau élevé de sûreté des centrales, tout en sachant que ces dernières sont âgées et que les exigences de sûreté ont été renforcées, suite aux différents accidents nucléaires (Three Mile Island, 1979 ; Tchernobyl, 1986 ; Fukushima, 2011).

Un des résultats de ces études est un ensemble de courbes de fragilité, exprimant la probabilité de défaillance d'un composant du réacteur en fonction d'une mesure de nocivité du séisme.

L'estimation de ces courbes repose sur deux éléments principaux :

- Un modèle physique, rendant compte de la réponse de l'équipement étudié à un séisme (par exemple, un élément de tuyauterie). Ce modèle peut être assez complexe et présente des temps de calcul de 10 heures à 100 heures par simulation.
- Un modèle d'incertitudes, rendant compte des manques de connaissance et des variabilités des paramètres du modèle physique. Cela concerne, par exemple, les sollicitations sismiques, les propriétés des bâtiments de la centrale et celle de l'équipement étudié. Ce modèle contient plusieurs dizaines de variables aléatoires associées à des lois de probabilités spécifiques.

Ces deux éléments définis, une propagation d'incertitudes est réalisée :

- 1) En définissant un plan d'expérience à partir des distributions des variables aléatoires.
- 2) En réalisant les expériences lors d'une campagne de calcul. Le modèle physique voit ainsi les valeurs de ses paramètres être modifiées pour chaque expérience, et les réponses de l'équipement étudié sont sauvegardées. Cette campagne de calcul engage des ressources de calcul importantes (cluster HPC) pour réaliser plusieurs centaines de simulations.
- 3) En analysant les résultats de la campagne. On cherche à évaluer la probabilité qu'un critère de défaillance (par exemple, un effort mécanique dans un ancrage de tuyauterie) dépasse une valeur seuil au-delà de laquelle l'équipement est défaillant (par exemple, la rupture de l'ancrage).

On présente en première approximation l'évolution de la probabilité de défaillance en fonction de la nocivité du séisme, ce qui donne des courbes de fragilités sismiques.

La campagne de calculs est longue et demande beaucoup de ressources de calcul. Une première piste d'amélioration consiste à optimiser l'échantillonnage pour réduire le nombre de simulations nécessaires.

La phase d'analyse peut être améliorée pour donner de meilleures estimations des probabilités de défaillances.

¹<https://www.irsnn.fr/FR/Larecherche/Thematiques-recherche/Pages/Etudes-probabilistes-de-surete-1569.aspx>

12 Monitoring and managing the lifespan of batteries of a fleet of electric vehicles

Proposé par : Phimeca Engineering

Remarque : Les échanges avec le porteur de projet se feront en anglais.

Contacts : Francesco Cancelliere (cancelliere@phimeca.com) and Sylvain Girard (girard@phimeca.com)

Descriptif : Batteries are responsible for a predominant share of the price of electric vehicles. Their longevity strongly impacts the cost of possession and is of prime interest to fleet operators. We aim to provide operator of free floating electric vehicle fleet (Velib', Jump, Bolt...) with a service for

- planning replacement and maintenance,
- cutting associated cost,
- and generally optimizing the longevity of their batteries, either by technological innovation or incentives to alter user behaviour.

Battery longevity estimation and forecast is currently tackled either from a data-centric (for instance by time series statistical modelling, mostly for forecasting), or a physics-based (at various scale, from the battery pack to the microscopic phenomena, mostly for diagnosis) viewpoint. We will compare both approaches, and possibly hybridise them.

The vocable “Digital Twin” has recently become almost as fashionable in industrial circle as Big Data some while ago, or Deep Learning and IoT. This project is an occasion to put some meaning behind the buzzword.

13 Évaluer le domaine de fonctionnement d'un modèle numérique

Proposé par : Phimeca Engineering

Contacts : Claire Eleuthériane-Gerrer (gerrer@phimeca.com) et Sylvain Girard (girard@phimeca.com)

Descriptif : Phimeca intervient souvent sur des questions de fiabilité et robustesse de grosses installations industrielles. Le déroulé des études est similaire : le client fournit un modèle physique (de réacteur de centrale nucléaire, de moteur de voiture, autres...), auquel Phimeca applique des techniques probabilistes et/ou statistiques afin de répondre à une question de sûreté (probabilité de défaillance de l'installation par exemple).

N'étant pas spécialistes des physiques de nos clients, nous traitons les modèles comme des boîtes noires : on en connaît les entrées et sorties, mais pas le contenu.

Un problème que nous rencontrons régulièrement est que le modèle ne fonctionne pas correctement sur toute la zone prévue pour l'analyse : soit du fait d'un crash de la simulation, soit du fait de résultats non valides.

Exemple. J'ai un modèle boîte noire. Si on pouvait regarder dans la boîte noire, ce modèle n'est autre qu'une fonction :

$$f(a, b) = \ln(ab).$$

Les paramètres a et b doivent être étudiés sur la zone $[-1, 1]$. Vu la fonction, on comprend bien que des combinaisons de valeurs de a et b sont interdites (il faut que a et b aient le même signe et un produit non nul).

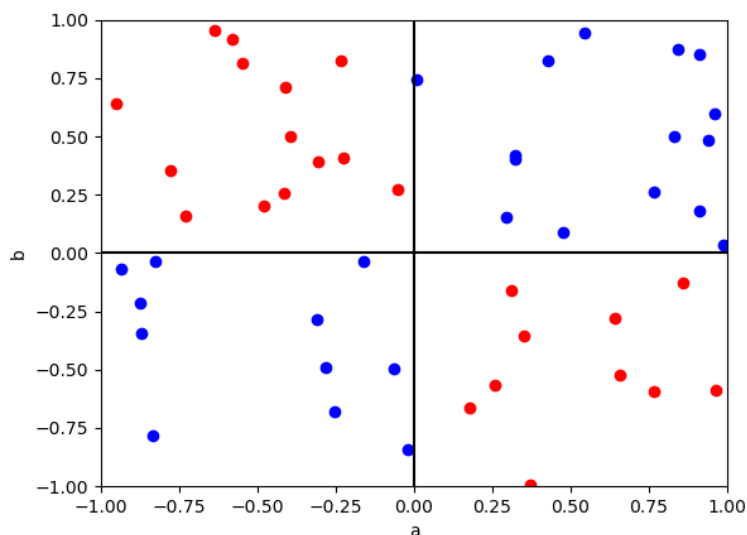


Figure 1: Simulations ratées (rouge) et réussies (bleu) selon les valeurs de a et b

Reprenons le point de vue «boîte noire». Comme on ignore le contenu du modèle, on tire quelques points (par Monte-Carlo, ou plan d'expérience plus finolé) pour vérifier que tout l'analyse statistique sera possible pour $a, b \in [-1, 1]$. Et on peut voir que les parties du graphe ($a \geq 0, b \leq 0$) et ($a \leq 0, b \geq 0$) causent toutes des ratés de simulation (Figure 1).

Dans cet exemple, on s'en sort car l'espace des entrées est de dimension 2, et donc tout à fait visualisable : mais quid quand l'espace des entrées est bien plus grand ?

Admettons qu'on ait pour modèle $f(a, b, c, d, e) = \ln(abcde)$: ça complique... Et a fortiori si les zones de non-fonctionnement du modèle sont de formes plus fantaisistes.

Comment :

- Estimer le volume respectif des deux zones ?
- Identifier les zones de fonctionnement ou non-fonctionnement du modèle ?

La question est difficile, et des pistes de réflexion, de réponse seront appréciées !

14 Modèles prédictifs pour le rendement d'actifs

Proposé par : Qube Research & Technologies

Contacts : Karl Bertoli (karl.bertoli@qube-rt.com)

Pierre-Alain Reigner (pierre-alain.reigner@qube-rt.com)

Descriptif : Qube Research & Technologies (QRT) est un groupe de gestion d'actifs spécialisé dans la gestion dite « systématique » et « quantitative ». Ses décisions d'investissement se basent sur une approche statistique et diversifiée. Son équipe de recherche construit des modèles prédictifs sur un vaste ensemble d'actifs (notamment actions, instruments dérivés) à partir de multiples jeux de données. Ces prédictions sont individuellement incertaines mais cette incertitude est réduite par la diversification des actifs en portefeuille et des modèles de prédiction.

Le projet proposé est l'étude de deux problèmes de prédiction typiques de l'activité de QRT, résultant en la soumission de solutions à ces problèmes.

1. Le premier problème sera traité à partir de février 2022. Il s'appuie sur un Challenge Data² dont les résultats seront alors déjà publiés : les élèves seront invités à dialoguer avec des chercheurs quantitatifs de QRT qui les guideront dans la proposition d'une solution. Ces allers-retours seront l'occasion pour les élèves de découvrir plus largement l'activité de QRT.
2. Le deuxième problème sera traité sans supervision, car il s'agira de note Challenge Data actif. Il s'agira pour le groupe d'élèves de s'approprier le problème de prédiction et de proposer une solution qui sera évaluée et classée parmi celles des autres participants du Challenge. Pour cette partie les élèves travailleront en autonomie car QRT ne peut pas favoriser certains participants au Challenge Data.

Les élèves devront :

1. Travailler en groupe pour proposer une solution à chaque problème de prédiction.
2. Soumettre des solutions aux deux problèmes de prédiction sur la plateforme du Challenge Data.
3. Restituer leur démarche et leurs résultats dans un court rapport et une présentation orale.

15 Tracking en natation

Proposé par : Laboratoire d'hydraulique Saint-Venant (École des Ponts)

Contact : Rémi Carmigniani (remi.carmigniani@enpc.fr)

Descriptif : Le bassin de l'INSEP est équipé de dix caméras (cinq caméras sous-marines et cinq caméras aériennes) permettant de suivre la trajectoire des nageurs lors de départ (voir la figure 2 avec quatre desdites caméras). Nous avons mis en place un système de tracking à partir d'une intelligence artificielle permettant de suivre 14 marqueurs et ainsi de remonter à la trajectoire du centre de masse des nageurs. L'objet de ce projet est d'améliorer le système existant.

Notamment, la calibration n'est faite que pour le plan passant par le centre de la ligne d'eau. Or, les nageurs ne nagent pas toujours au centre. À partir des données marquées, il nous faut repérer les phases importantes : amplitudes des oscillations en ondulations, coups de bras, battements. . .

Les élèves se concentreront sur ces deux piliers : amélioration du suivi du plan du nageur, détection des phases à partir des résultats obtenus sur les marqueurs et reconstruction de la trajectoire de la main.

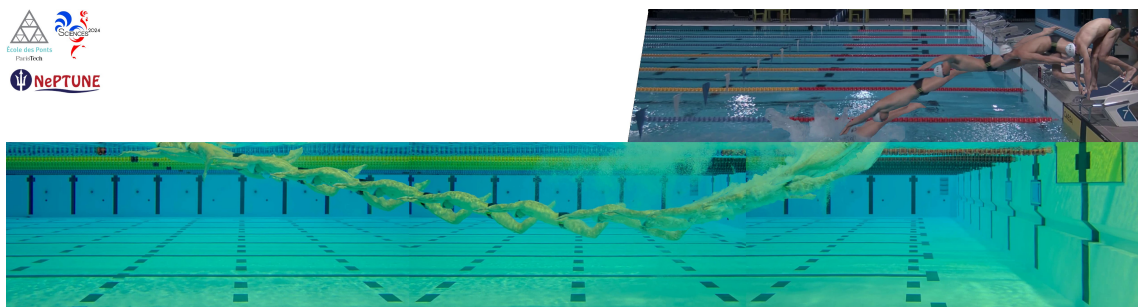


Figure 2: Tracking d'un nageur lors d'un départ

Vous trouverez à ce lien une vidéo d'un départ de natation avec le tracking du squelette à 14 marqueurs : <https://drive.google.com/drive/folders/1qu22N450-ZFo7bKZrZJwhhh3WpVBKdRr?usp=sharing>.

²<https://challengedata.ens.fr/participants/challenges/44/>

16 Optimisation des flux de passagers à l'interface quai-train

Proposé par : SNCF et IRT SystemX

Contacts : Mostepha Khaoudjia (mostepha.khouadjia@irt-systemx.fr)

Elise Grison (elise.grison@sncf.fr)

Éric Duceau (eric.duceau@enpc.com)

Descriptif : L'objectif de ce projet est d'optimiser les flux de passagers à la montée et la descente des trains, sur les quais et dans les voitures. Il s'agira de proposer des solutions optimales robustes et validées, compatibles avec les infrastructures existantes (quais) ou sur le point d'être acquises (nouvelles générations de rames).

La valeur ajoutée est liée à la fois à l'amélioration (hausse) des flux d'échange à infrastructure équivalente et à la preuve que les solutions optimales sont robustes vis-à-vis des variabilités de comportement des personnes ou des groupes.

Méthode suggérée :

- développer (ou enrichir) un logiciel de simulation (base d'agent possible, le 2D sera suffisant) ;
- le tester sur des configurations représentatives pertinentes par rapport au problème envisagé ; il ne s'agit pas de modéliser les situations à faible densité mais celles où il y a des « pertes de temps » dues au comportement insuffisamment coopératif des individus ayant des objectifs contradictoires (monter et descendre) mais pas incompatibles (laisser descendre avant de monter) ;
- coupler le simulateur avec un algorithme d'optimisation : la difficulté est de gagner 5 à 10% d'un temps de simulation qui, en situation réelle, avoisine 10 à 15 secondes.

Les entrées (données disponibles) sont les distributions des gens sur le quai et dans les voitures. Les sorties (quantités d'intérêt) sont les temps de chargement / déchargement et tout autre indicateur local (densité, émergence de comportements) ou global (moyennes) sur les flux.

Les sources d'incertitudes sont justement les variabilités de comportements des usagers ou des groupes conduisant à des stratégies de positionnement et de déplacements individuels (doubler, pousser, se mettre à contre-sens) ou collectifs (se mettre en ligne, marcher à droite) favorisant l'émergence de situations favorables ou défavorables.

Éléments pour définir une démarche méthodologique :

- Une modélisation par agents est plausible (mais ce n'est pas la seule) : on pourrait imaginer « apprendre » à l'agent certains comportements. Pour cela, les observations à échelle 1 sont nécessaires.
- L'émergence des comportements collectifs peut être simulée ; une technique relevant de l'IA est classique pour piloter des paramètres de simulation afin de favoriser certains scénarios.
- La simulation peut théoriquement permettre de proposer des solutions d'optimisation ; leur validation est le vrai test d'intérêt vu les nombreuses sources d'incertitudes des modèles.

17 Optimisation de puissance électrique générée par une éolienne

Proposé par : WIND my ROOF

Contact : Yanis Maacha (yanis.maacha@windmyroof.com)

Descriptif : WIND my ROOF est une start-up qui conçoit et commercialise des éoliennes de toiture. Pour introduire le contexte du projet, lors de la phase de vie de l'éolienne, nous pouvons piloter le courant injecté sur le réseau. La valeur du courant à injecter n'est pas à laisser au hasard, elle dépend de la vitesse de vent

instantanée et des caractéristiques de l'éolienne. Plus il y a de vent, plus il faut injecter de courant pour avoir une production de puissance optimale et inversement. La figure 3 illustre bien la situation.

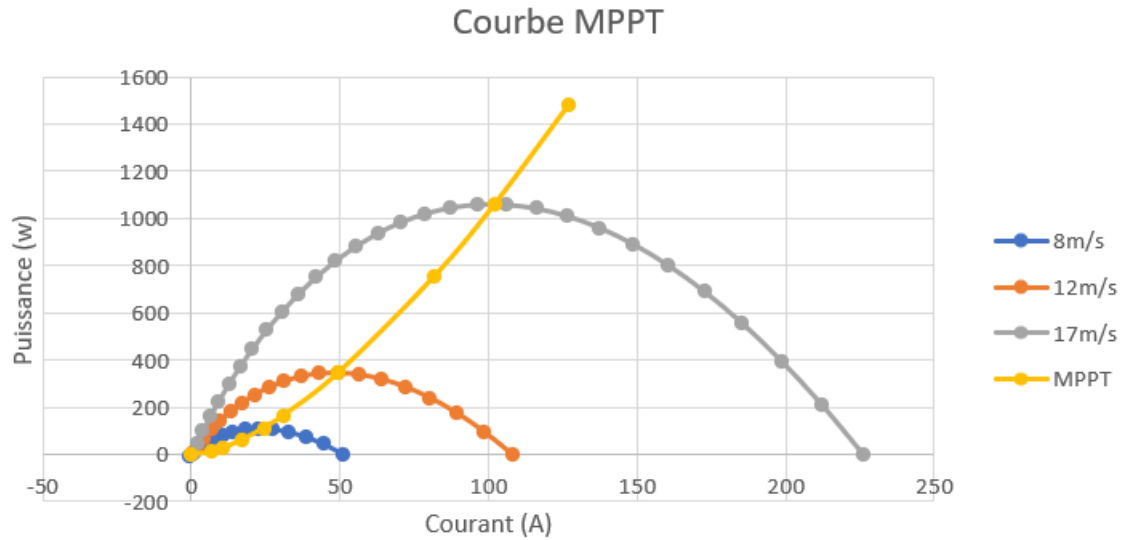


Figure 3: Puissance produite en fonction du courant injecté sur le réseau

Ce graphique est issu de mesures effectuées en soufflerie, il trace la puissance produite en fonction du courant injecté sur le réseau et ce pour plusieurs vitesses de vent. Les différentes courbes indiquent les puissances que l'on peut produire avec un vent de 8 m/s (courbe bleue), 12 m/s (courbe orange) etc... On voit que pour chaque vitesse de vent, il existe un courant à injecter qui maximise la puissance produite, c'est le tracé de la courbe jaune.

Le cœur de notre problème, c'est de déterminer à chaque instant ce courant optimal, sans connaître précisément les caractéristiques de notre éolienne ni la vitesse du vent. On dispose cependant des données instantanées (courant, tension, etc...) et de leur historique. L'objectif est de mettre à profit ces mesures pour maximiser l'énergie produite sur toute la durée de vie de l'éolienne.