

# La Société Française de Génie des Procédés et le **cnam**

organisent une journée

## **Big data et génie des procédés : apports et limites**

**mercredi 5 septembre 2018 au Cnam**  
(amphithéâtre Laussédat, 2 rue Conté, Paris 3<sup>e</sup>)  
accueil à partir de 9h30 ; début des conférences à 10h

---

### **CONTEXTE et OBJECTIFS de la JOURNÉE**

De nos jours, dans l'industrie des procédés, les nombreux instruments de mesure installés dans les usines de transformation et de fabrication, ainsi que dans les laboratoires de contrôle, produisent, chaque seconde, chaque minute ou chaque heure, des quantités gigantesques de données d'opération. Ces informations concernent notamment les températures, les débits, les pressions, les compositions et la consommation d'énergie. Or, l'énorme quantité de ces données rend difficile l'extraction de l'information utile à la prise de décision, entre autres pour la résolution d'un problème.

L'objectif de cette journée organisée par le Groupe Thématique « Ingénierie Avancée des Procédés » est de comprendre si les méthodes de calcul utilisées pour la conception, le dimensionnement, la régulation et l'exploitation des procédés et les méthodes d'analyse des données *Big Data* peuvent être complémentaires ou non, tout en précisant les apports et limites de chacune des approches.

---

### **QU'EST-CE que le *BIG-DATA* ?**

À l'ère du *Big Data* (ou « méga-données » en français), les données en provenance de diverses sources sont de plus en plus nombreuses. Certains y voient un casse-tête, tandis que d'autres y voient une occasion de transformer rapidement les données en information décisionnelle. En effet, il semble que l'utilisation et les bénéfices qu'on pourrait récolter des méga-données demeurent encore trop limités en raison d'un manque de connaissance des méthodes de traitement et d'analyse de ces données.

Une difficulté dans l'application de ces méthodes provient notamment de la qualité de la collecte des données qui, dans de nombreux cas, n'est pas rigoureusement contrôlée : pas de plan d'expériences, pas de données soigneusement recoupées et identifiées, pas de stratification précise des échantillons, pas de détection de données aberrantes. Par conséquent, les algorithmes de *Big Data* doivent s'accommoder de données de qualité souvent médiocre. Les données *Big Data* sont souvent incomplètes, entachées d'erreurs, ou de bruit. Avant d'être archivées et analysées ces données doivent être « prétraitées », « filtrées ». Après cette étape vient typiquement l'analyse. Les tâches que l'on peut assigner à un système de traitement *Big Data* se séparent en deux catégories : les tâches « prédictives »

ou « décisionnelles » et les tâches de « découverte » ou de « minage » (*data mining*) des données. Ces deux grandes classes de problèmes ont des objectifs et des contraintes très différents.

Les tâches prédictives recouvrent en particulier la classification et la prédiction, mais ne se limitent pas à ces problèmes. La classification est un ensemble de méthodes qui permettent d'identifier la « classe » d'une donnée, en utilisant une base de données d'apprentissage. La prédiction (ou régression) est un ensemble de méthodes statistiques permettant, en utilisant un ensemble d'apprentissage, une réponse en fonction d'un ensemble de variables explicatives. Cette réponse peut être un label catégoriel, attribut numérique, voire une fonction (pour la régression fonctionnelle) etc. Les tâches prédictives ont dans certains cas une dimension « temps réel ». Par exemple, pourquoi attendre qu'une défaillance se produise et subir des coûts de réparation onéreux ? La réponse est, normalement, parce que nous ne pouvons pas prédire l'avenir. Pourtant, les informations provenant des machines peuvent nous renseigner sur la façon dont des instruments, ou équipements fonctionneront à long terme. La technologie fondée sur les mêmes principes, à savoir détecter à quel endroit les erreurs se sont produites dans le passé et rechercher les situations futures où les mêmes conditions sont susceptibles de réapparaître, est utilisée également pour déceler des problèmes. La prochaine étape sera celle de la maintenance automatisée : lorsque la défaillance d'une machine est prévue, une nouvelle peut être envoyée à l'atelier pour que le remplacement se fasse avant que la panne ne survienne.

Les tâches de « découverte » ou de « minage » consistent à extraire des « structures » de grands ensembles de données. Ces structures peuvent être des clusters (des groupes d'observations ayant des comportements homogènes), des règles d'association (pour détecter des relations entre les ensembles de variables), de l'analyse de réseaux (identifier les relations entre les différents nœuds d'un réseau). Ces tâches de *data mining* font plutôt l'objet d'analyses « hors-ligne ».

---

## PROGRAMME détaillé

---

10h00-10h15 : Philippe ARPENTINIER : **Introduction et présentation de la journée**

10h15-10h45 : Jean-Claude CHARPENTIER (Directeur de Recherche, LRGP, Nancy) : **Les grandes tendances du Génie des Procédés au XXI<sup>e</sup> siècle.**

10h45-11h15 : Athanassios KONTOPOULOS (Air Liquide, Computational & Data Science Scientific Director, R&D) : **Big Data, théorie et applications en entreprise.**

11h15-11h45 : Mounir BAKKALI (CEA Tech - LIST - DM2I, responsable des partenariats industriels) : **Présentation des activités de CEA Tech sur l'analyse de données.**

11h45-12h15 : Amine BENSY (EDF R&D) : **Suivi des performances des procédés et identification des sources de défaillance.**

**12h30-13h30 : Déjeuner**

13h45-14h15 : Romain ROUX (Axens) : **Digitalisation de l'expérience client dans le raffinage.**

14h15-14h45 : Patrice KIENER (InModelia) : **Le langage open source R pour le traitement des données en génie chimique, des plans d'expérience aux données massives.**

14h45-15h45: Nick ZIOGAS (CERN, Knowledge Transfer) - Matthias BRAEGER (CERN, C2MON data acquisition & control framework) - Lorenzo MONETA (CERN, ROOT & TMVA big data analysis framework and ML library) : **ROOT & TMVA for data analysis and machine learning and C2MON for data acquisition and control.**

15h45-16h15: Ewrim ÖRS (Air Liquide, Frankfurt Research and Technical Center) : **Data science applied to the monitoring of catalyst deactivation.**

16h15-16h30: **Conclusion collective**