



Colloque IMT
**« L'IA au cœur des mutations
industrielles »**

04 avril 2019

Posters

Colloque IMT - Jeudi 4 avril 2019

L'IA au cœur des mutations industrielles

Apprentissage

1	Learning Ensembles of Monotone Boolean Functions For Interpretable Classification	EURECOM
2	De l'extraction automatique de modèles dans le cadre de problèmes d'optimisation.	IMT ATLANTIQUE
3	Réseaux de Neurones Récurrents Multi-tâches pour l'Analyse Automatique d'Arguments	IMT MINES ALES
4	Traitement de données incertaines : Application au tri des matières plastiques	IMT MINES ALES
5	High-reliability renewable production forecast for ancillary services'	MINES PARISTECH
6	Dealing with Topological Information within a Fully Convolutional Neural Network	MINES PARISTECH
7	Approches complémentaires d'apprentissage automatique : cas d'application dans l'analyse sensorielle	MINES ST ETIENNE
8	Adaptive collaborative topic modeling for online recommendation	TELECOM PARISTECH
9	Learning-based software metrics suggestion approach	TELECOM SUDPARIS
10	Real-time anomaly detection with in-flight data	TELECOM SUDPARIS

Optimisation

11	Algorithmes évolutionnaires et Contraintes : comment converger au plus vite vers les solutions Pareto-optimales en Aide à la Décision.	IMT MINES ALBI
12	Définition de l'assortiment idéal dans un réseau de magasin	IMT MINES ALES
13	Towards Distributed Real-Time Coordination of Shoppers' Routes in Smart Hypermarkets	IMT LILLE DOUAI

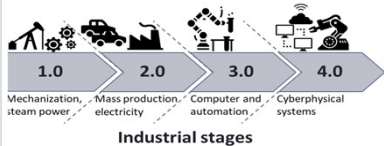
Robotique, Perception, Interaction

14	Coordination of Distributed Decision Processes and Applications	IMT LILLE DOUAI
15	Computer vision for model order reduction in mechanics	MINES PARISTECH
16	Coordination spontanée et résiliente d'objets intelligents	MINES ST ETIENNE

Données et Connaissances

17	Configuration/customisation de produit, un terrain d'application "successful" de nombreuses techniques d'IA : approches par contraintes, quelques problèmes et travaux.	IMT MINES ALBI
18	Modèles d'ensemble pour la prévision hydrologique par réseaux de neurones	IMT MINES ALES
19	IA et deep learning au service de l'environnement	IMT MINES ALES
20	Etude d'une approche de Retour d'Expérience pour la découverte d'enseignements génériques	IMT MINES ALES
21	Embedding de données d'interaction sociale	IMT ATLANTIQUE
22	STEM and Teens: Algorithm Bias in Social Media	IMT_BS
23	Apports et défis de la simulation multi-agent pour le jumeau numérique d'une usine	MINES ST ETIENNE

Contribution and challenges of multiagent simulation for factory digital twin



Digital twin

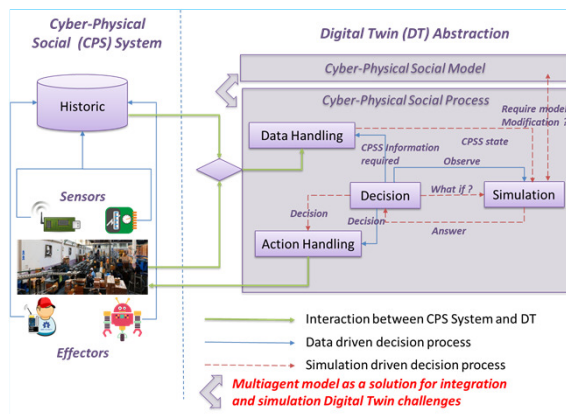
A key technology for the Industry of the Future

- One of top 10 **strategic technology** trends for 2019 (cf. Gartner)
- A factory digital twin serves as a **virtual replica** of what is actually happening on the factory floor in near-real time

SOLYSTIC
SOSi™



- **SOLYSTIC** is a provider of solutions for mail and parcel industry
- **SOSi™** is an inhouse developed factory/supply chain digital twin; simulates **1 production year in 10 minutes**
- Enables **digital VSM** and **optimization of processes**



Multiagent approach

A bottom-up modeling approach for modeling a CPS System

- **Agents** for modeling goal-driven decision processes
- **Environment** for modeling information perception and actions
- **Interaction** for modeling influences between components
- **Organization** for modeling formalized and/or implicit rules

Multiagent model of a CPS System

A solution for making easier the transition

- **FROM Simulation as a tool for Observation**: to understand the behavior of the reference system thanks to a model that is considered as a miniature reproduction of the reference system
- **TO Simulation as a tool for Validation**: to test an hypothesis of the reference system, to validate or to certify the underlying theory.

Why?

A CPS System is often

- **Complex**: the global behavior of the system is hard to model and any modifications is difficult
 - **Multiagent solution**: The global system is not explicitly designed
 - **Multiagent solution**: The multiagent concepts can be understood by non-experts
- **Open**: new component may be added to the system leading to a modification of the initial model
 - **Multiagent solution**: only the new components and its interaction with other components must be designed
- **Heterogeneous**: different decision models, data models, ...
 - **Multiagent solution**: the resulting model is independent of the domain and can integrate several point of views.
- Decentralized: many local decisions without a centralized control
 - **Multiagent solution**: decentralization management is the core of a multiagent system.

Contact : flavien.balbo@mines-stetienne.fr

Parties prenantes



Auteurs

Flavien Balbo
Sophie Peillon
Benjamin Serra

Partenaires

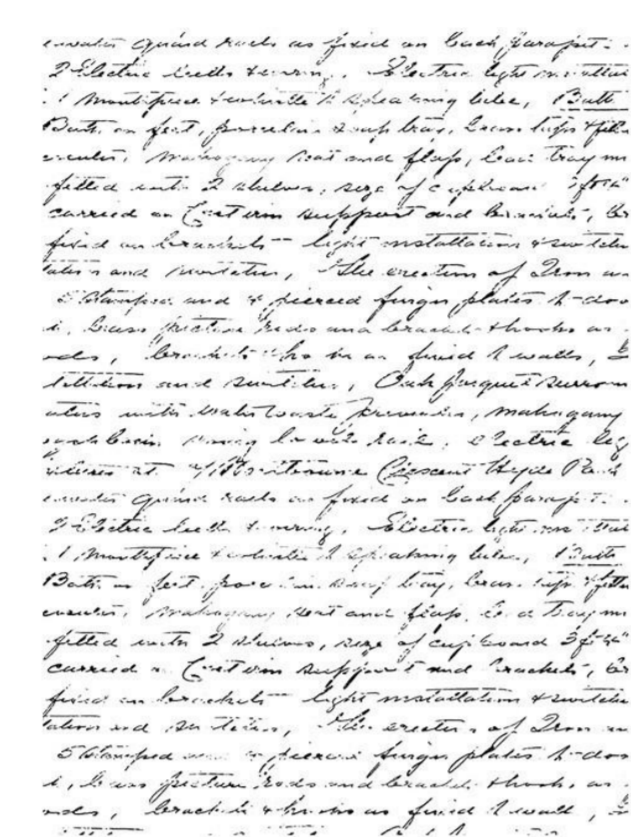


Embedding de données d'interaction sociale

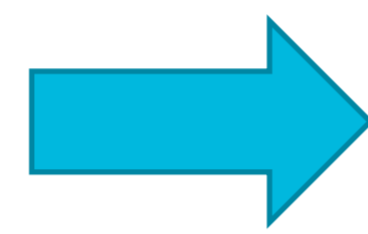
Une nouvelle approche dynamique

L'embedding de données : Une approche interdisciplinaire pour l'apprentissage de la représentation de données brutes

- **Méthode de prétraitement issue du traitement de texte** – Plongement vectoriel des mots, conservant la proximité sémantique : Représentation + réduction de la dimension
- **Méthode adaptée à d'autres types de données** – Images, séries temporelles, graphes.

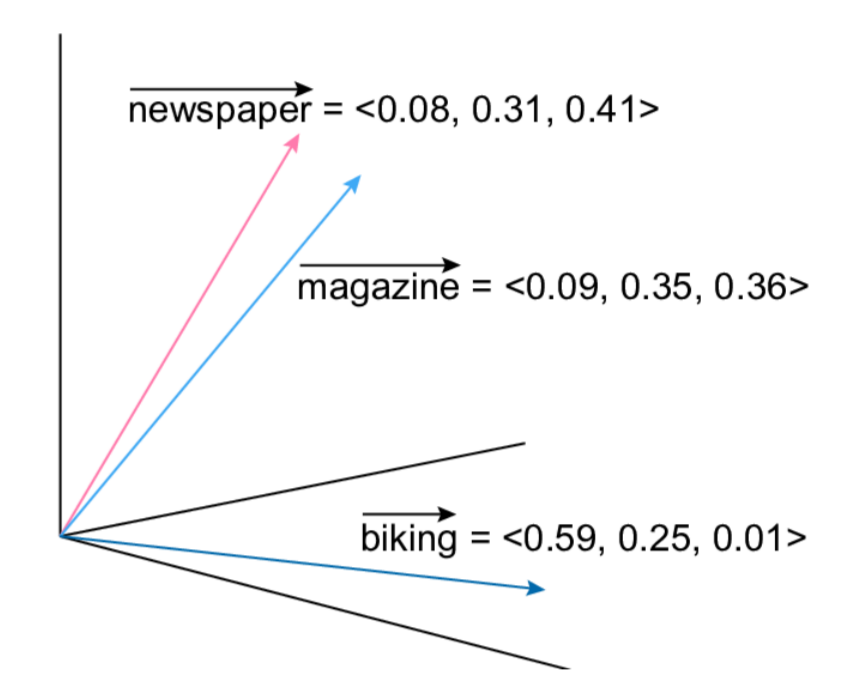


Texte brut



1	5	5	3	9	7	8	4	0	3	8
5	4	7	9	2	0	5	8	3	9	4
4	1	5	2	8	0	5	8	9	4	3
7	9	4	7	0	8	3	6	1	1	8
0	1	7	6	3	8	7	9	9	0	1
9	4	3	7	9	3	0	7	2	8	1
9	6	9	2	7	0	1	6	8	3	4
4	5	8	9	3	3	0	8	4	1	7
3	4	1	4	8	0	1	4	3	4	1
8	0	4	3	9	7	6	9	4	2	1
0	1	5	7	9	1	5	6	8	3	1

Matrice de co-occurrence des mots

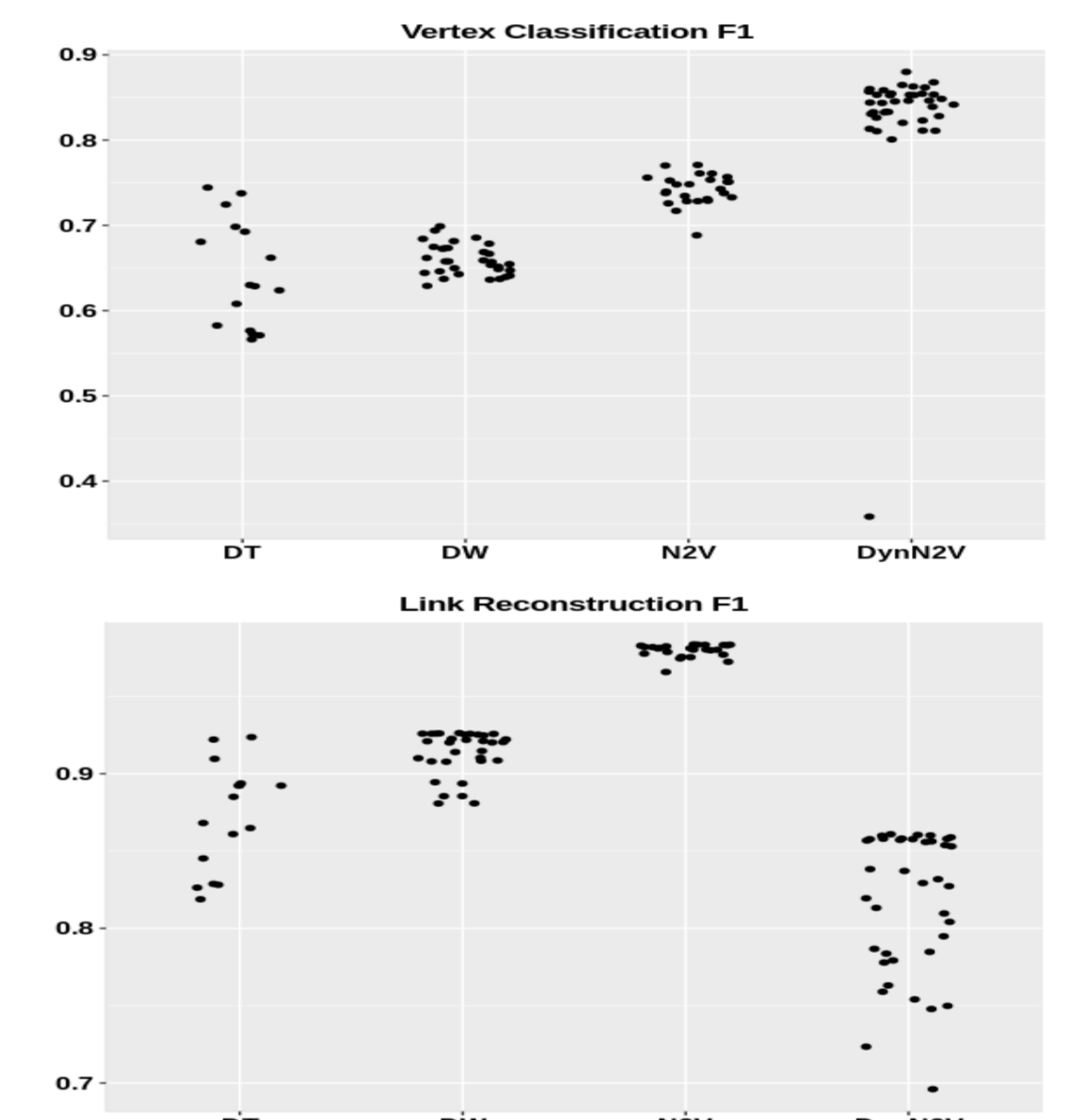


Cas des données d'interaction : Etat de l'art et limites

- **Objectif** – Embedding en vue de la détection et caractérisation de communautés, prédiction de liens...
- **DeepWalk [1] & Node2Vec [2] : des approches disruptives** – Ces approches se sont imposées comme références en embedding de graphes sociaux.
- **Limite : dimension temporelle** – Le temps demeure fondamental pour la compréhension des phénomènes d'interaction. Des approches d'embedding récentes intègrent le temps (DynamicTriad [3])

DynamicNode2Vec :

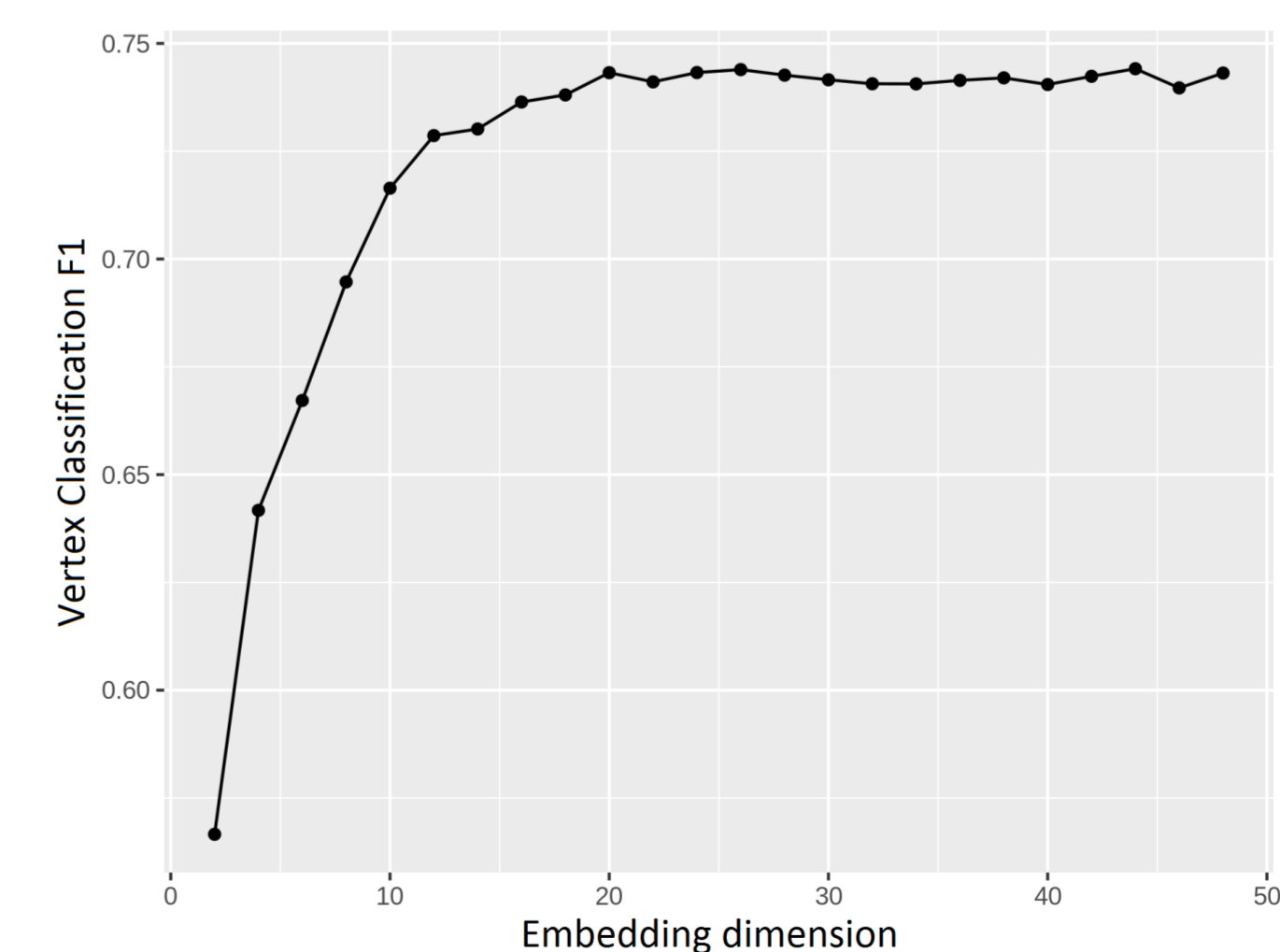
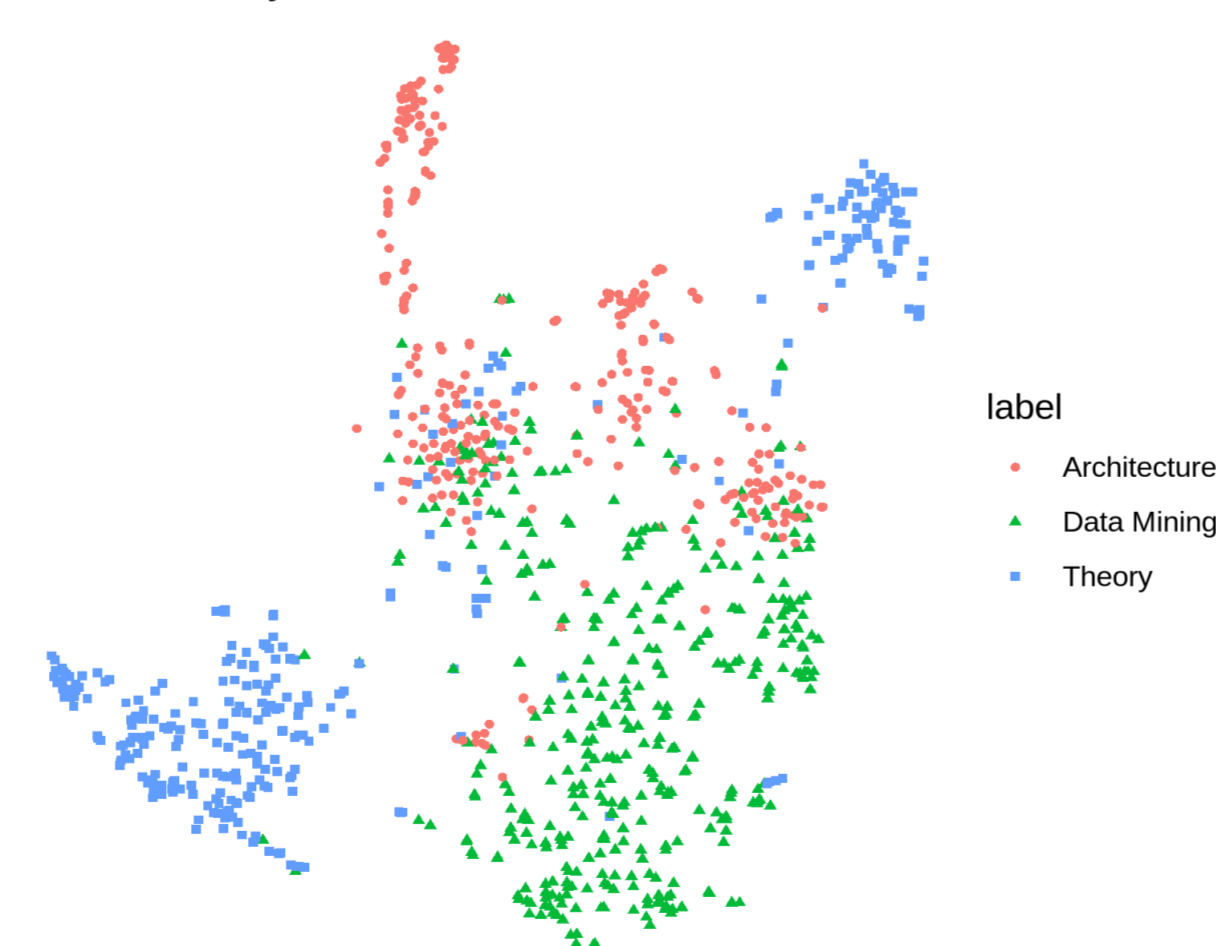
- **Principale contribution : prise en compte de la dimension temporelle** – Lissage temporel pour l'apprentissage des embeddings des différents pas de temps.
- **Intérêt** – Embeddings continus dans le temps (alignement temporel), meilleure appréhension de la dynamique des nœuds et des communautés.
- **Résultats d'inférence probants** – La classification des nœuds s'en trouve améliorée, la reconstruction de liens est moins performante.



Expérimentations :

- **Jeu de données** – Co-publications scientifiques [4]
 - 51 060 nœuds (chercheurs), 794 552 interactions (co-publications)
 - 17 pas de temps de 4 ans (avec chevauchement de 2 ans)
 - 6 communautés (Architecture, Réseaux, Sécurité, Data Mining, Théorie, Graphisme)
 - Un chercheur appartient à une communauté si plus de la moitié de ses articles sont publiés dans une conférence s'y rapportant
- **Projection 2-D des embeddings** – A l'aide de t-SNE [5] (technique de réduction de la dimension)
- **Evolution de la qualité de l'embedding en fonction de la dimension** – 12 à 20 dimensions suffisent

DynN2V 2-D visualisation



Perspectives :

- **Tests sur d'autres jeux de données**
- **Prise en compte d'éventuelles métadonnées** – Approche supervisée, adaptation de la fonction coût
- **Ajout de la cohérence temporelle dans la construction des marches aléatoires** – Adaptation de l'algorithme CTDNE [6]
- **Application d'algorithmes de vision par ordinateur sur les embeddings** – Essentiellement des CNN

Parties prenantes



Auteur

Mounir HADDAD

Encadrants

Cécile BOTHOREL
Philippe LENCA
Dominique BEDART
Mikaël CAMPION

Références

[1] DeepWalk : Perozzi, Bryan, Rami Al-Rfou, and Steven Skiena. "Deepwalk: Online learning of social representations." *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2014.

[2] Node2Vec : Grover, Aditya, and Jure Leskovec. "node2vec: Scalable feature learning for networks." *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2016.

[3] DynamicTriad : Zhou, L.; Yang, Y.; Ren, X.; Wu, F. and Zhuang, Y., 2018, Dynamic Network Embedding by Modelling Triadic Closure Process, In AAAI, 2018

[4] Amner : Tang, Jie, et al. "Armetminer: extraction and mining of academic social networks." *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2008.

[5] T-SNE : Maaten, Laurens van der, and Geoffrey Hinton. "Visualizing data using t-SNE." *Journal of machine learning research* 9.Nov (2008): 2579-2605.

[6] CTDNE : Nguyen, Giang Hoang, et al. "Continuous-time dynamic network embeddings." *Companion of the The Web Conference 2018 on The Web Conference 2018*. International World Wide Web Conferences Steering Committee, 2018.

ÉTUDE D'UNE APPROCHE DE RETOUR D'EXPÉRIENCE POUR LA DÉCOUVERTE D'ENSEIGNEMENTS GÉNÉRIQUES : Application au domaine Humanitaire

Parties prenantes



Auteurs

Cécile L'HÉRITIER
Sébastien HARISPE
Abdelhak IMOUSSATEN
Gilles DUSSERRE
Benoît ROIG

Partenaire



CONTEXTE GÉNÉRAL ET ENJEUX

Processus décisionnels faisant intervenir de la connaissance experte

- ▶ Patrimoine de connaissances construit par l'expérience terrain
- ▶ Tirer parti d'expériences similaires passées, erreurs ou succès
- ▶ Découverte de connaissances *via* des interactions Homme-Machine

Spécificités liées au cadre humanitaire

- ▶ Témoignage oral, *turnover*
- ▶ Contexte d'urgence
- ▶ Décisions complexes, critiques
- ▶ Instinct

Missions

Spécificité du contexte



Similarité de la stratégie logistique ?



Succès ?



RetEx : leçons ?

OBJECTIFS

Contribuer à l'amélioration du processus de Retour d'Expérience (RetEx) :

- ▶ Analyser les expériences passées pour extraire des enseignements génériques d'intérêt pour le domaine humanitaire.
- ▶ Identifier les divergences ou similitudes expliquant leur succès ou échec.

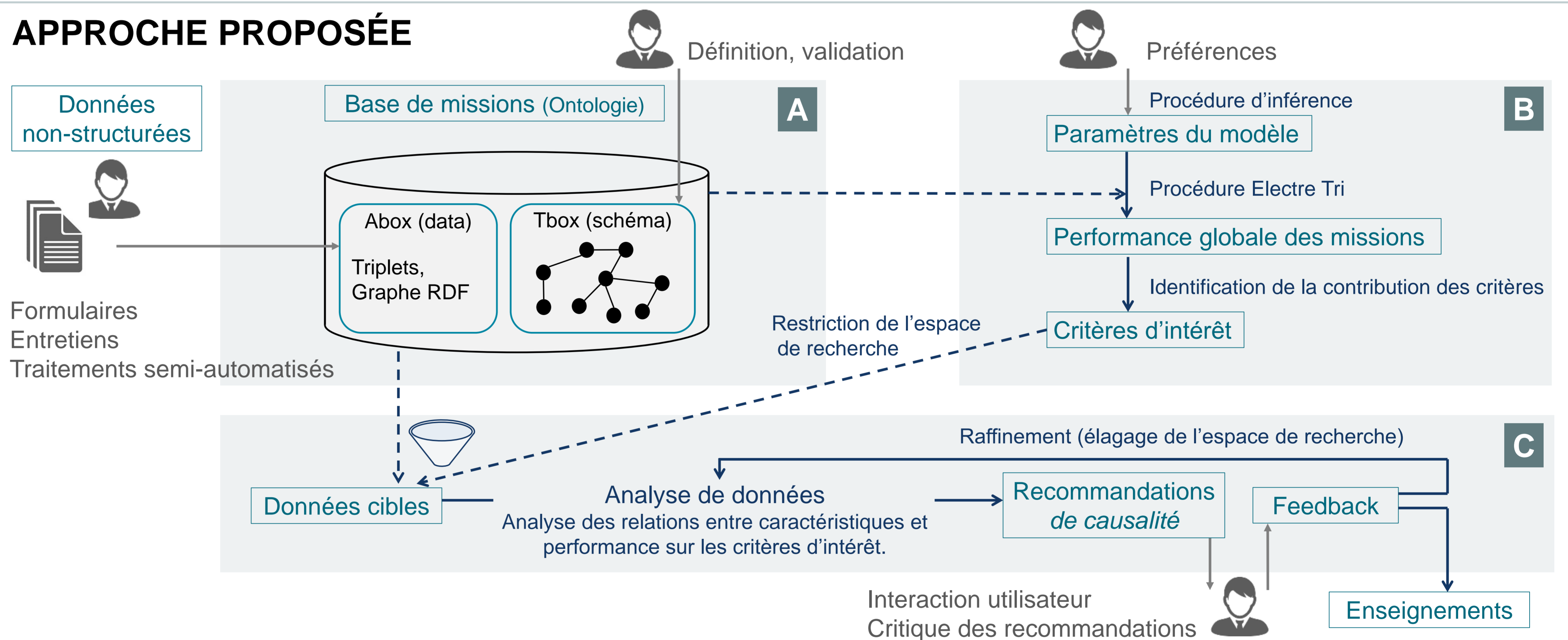
Une approche au carrefour de plusieurs domaines :

la Représentation des Connaissances, l'Analyse Multicritère et l'Extraction de Connaissances

Limitations théoriques et pratiques pour la mise en œuvre d'une approche de RetEx :

1. Acquisition de l'information et des connaissances nécessaires à l'analyse
 - ▶ Représentation : compromis entre compréhension et exploitation
 - ▶ Chronophage, effort cognitif
2. Analyse et exploitation
 - ▶ Données imparfaites
 - ▶ Faible nombre d'observations

APPROCHE PROPOSÉE



A Représentation et formalisation

Définir un cadre de formalisation pour valoriser les données, et constituer une base de missions exploitable.

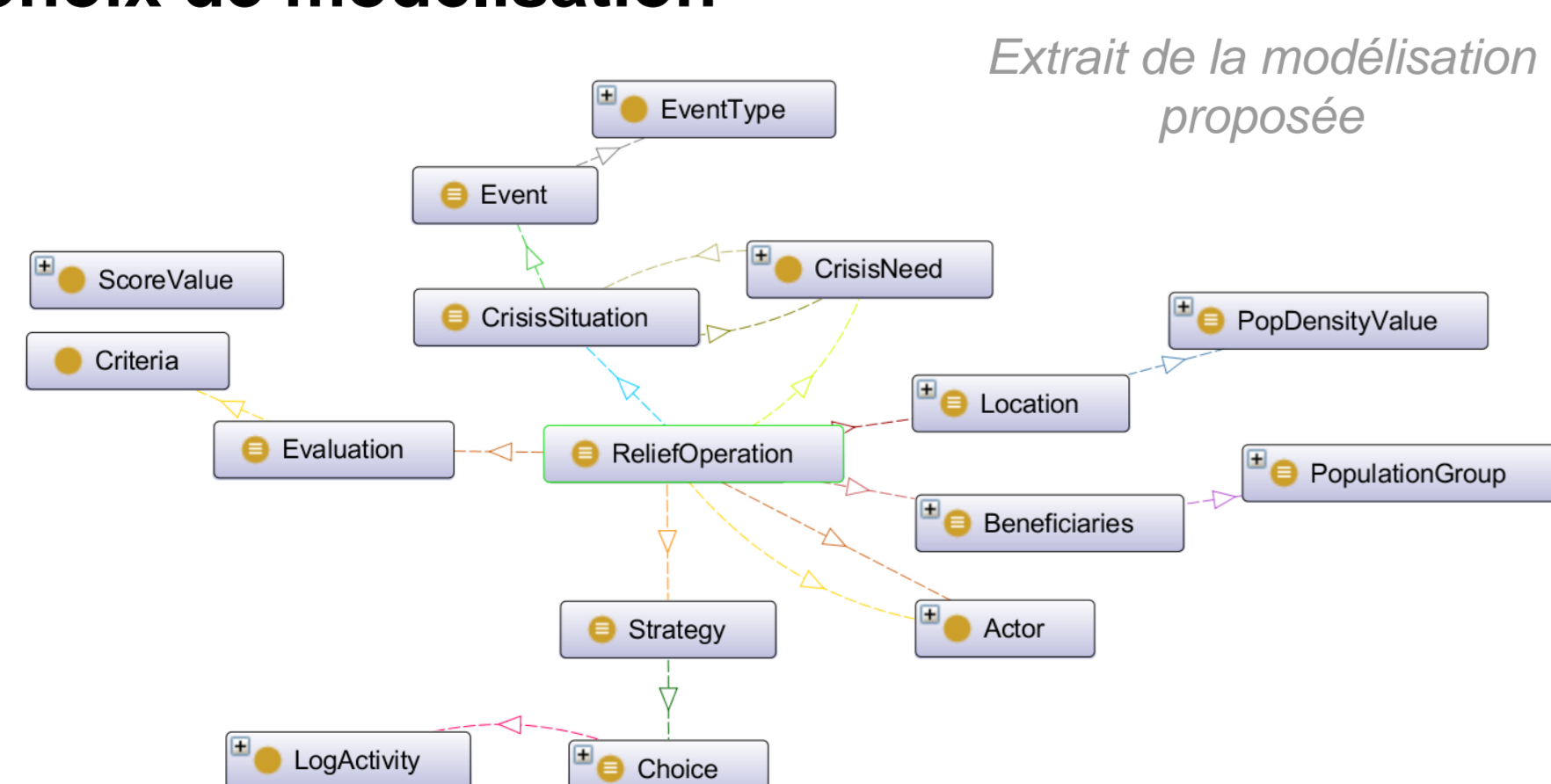
Choix d'une représentation formelle

- ▶ Connaissance *a priori*
- ▶ Réutilisation d'ontologies existantes
- ▶ Inférence
- ▶ Consistance

Standards W3C : spécifications RDF(S), OWL, SWRL

- ▶ Compromis complexité de raisonnement / expressivité

Choix de modélisation

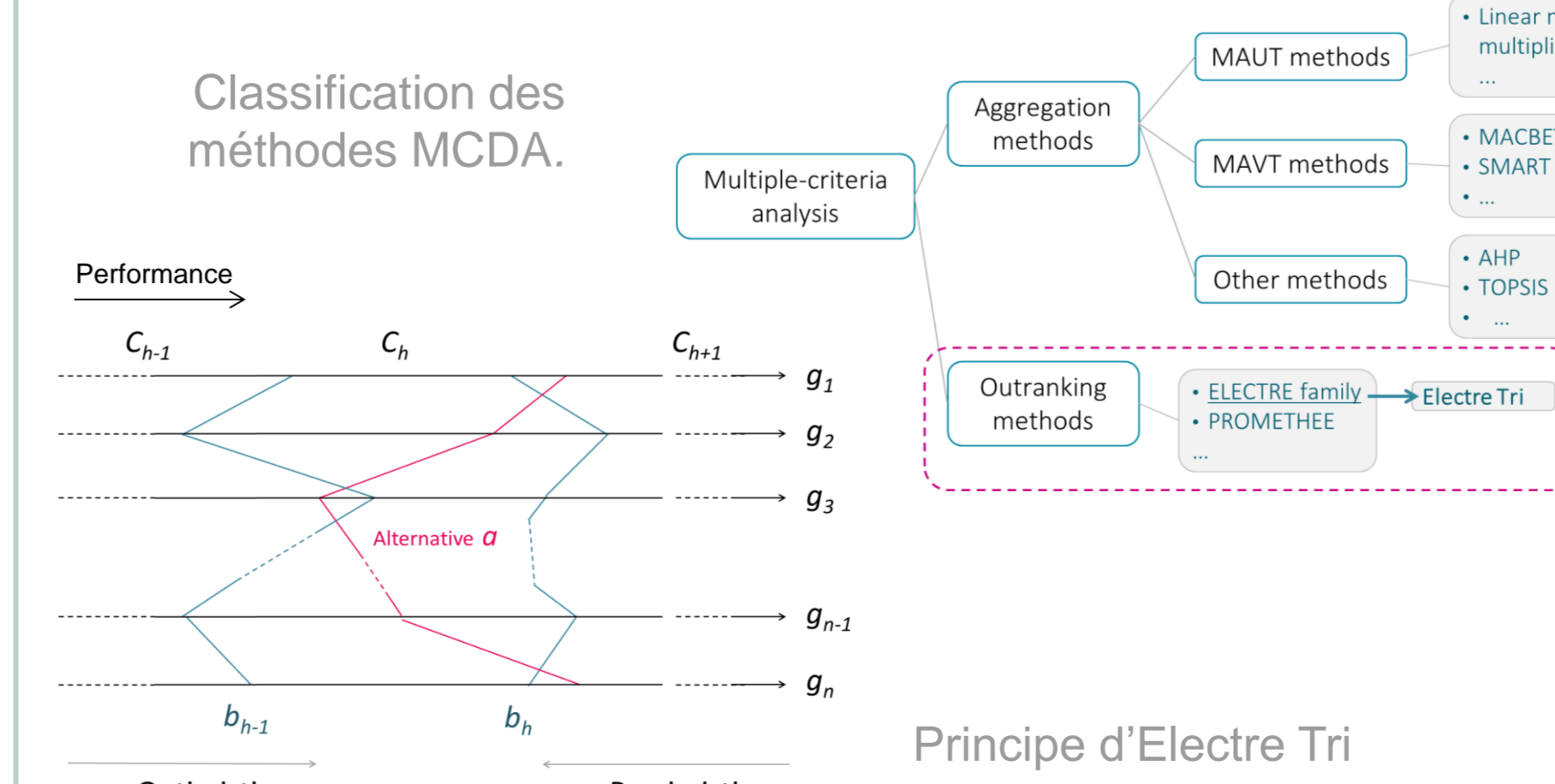


B Evaluer et déterminer des critères d'intérêt

Réduire l'espace de recherche et cibler le RetEx pour en tirer des enseignements pertinents.

1. Evaluation du succès / échec : Electre Tri

- ▶ Non-compensatoire
- ▶ Prise en compte de l'imprécision
- ▶ Identification indirecte des paramètres



2. Identification des critères d'intérêt

Définir la contribution des critères à la performance [1]

C Extraction de connaissances

Recherche de corrélations et caractéristiques discriminantes expliquant la performance sur les critères d'intérêt.

Cadre évidentiel (Théorie de Dempster-Shafer)

- ▶ Gestion de l'imprécision
- ▶ Prise en compte de la structuration des données (connaissance *a priori*)

Identification et sélection de recommandations : évaluation objective

- ▶ Force de la relation
- ▶ Représentativité
- ▶ Robustesse de la mesure
- ▶ Compromis entre généralité et spécificité des recommandations (contenu informationnel)
- ▶ Caractère discriminant (pertinence)

Interaction utilisateur : évaluation subjective

Critique de la pertinence des recommandations

[1] C. L'Héritier, A. Imoussaten, S. Harispe, G. Dusserre, B. Roig, "Identifying Criteria Most Influencing Strategy Performance: Application to Humanitarian Logistical Strategy Planning", In *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems IPMU*, pp. 111-123, 2018.

[2] C. L'Héritier, S. Harispe, A. Imoussaten, G. Dusserre, B. Roig, "Etude d'une approche de Retour d'Expérience pour la découverte d'enseignements génériques dans le domaine humanitaire", In *29es Journées Francophones d'Ingénierie des Connaissances, IC 2018*, pp. 87-94, 2018.



IMT Mines Alès
École Mines-Télécom

IA et *deep learning* au service de l'environnement

Prévision opérationnelle des variables hydro(géo)logiques

Parties prenantes



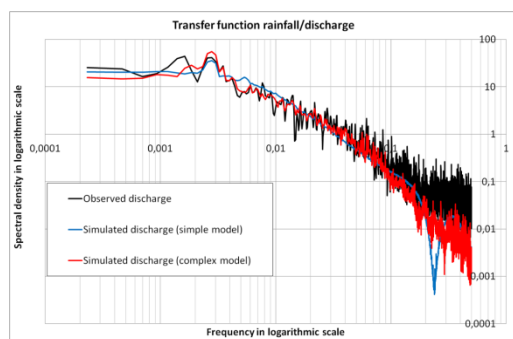
Auteurs

Guillaume Artigue
Anne Johannet
Marc Vinches

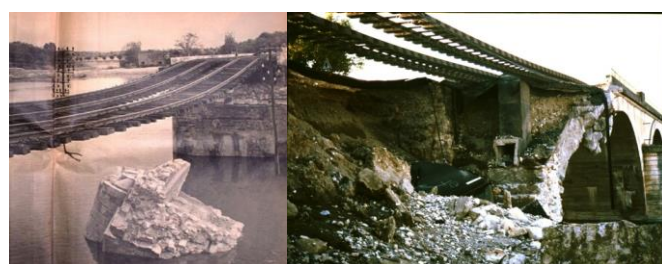
Partenaires



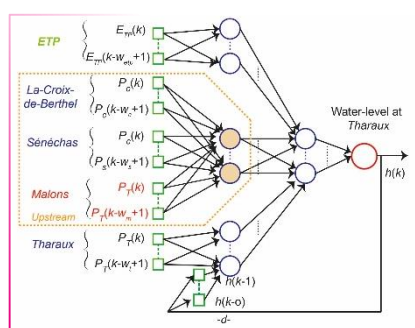
Mines Alès
est certifiée par



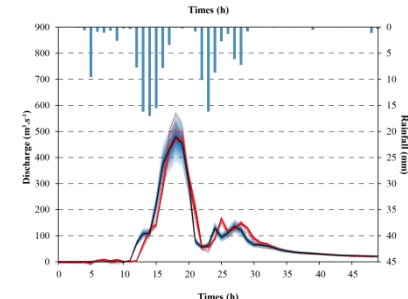
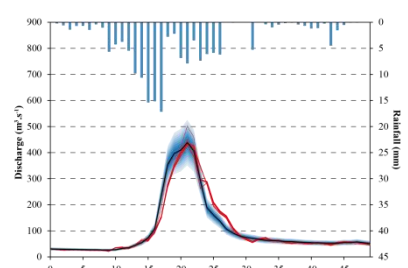
Analyse spectrale - Thèse V. Taver



Les mêmes causes produisent les mêmes effets :
Pont SNCF de Ners : 1958 (gauche), 2002 (droite)



Deep-learning : Etude du rôle de l'évapotranspiration sur le bassin de la Cèze



Prévision des crues du Lez à Lavalette (entrée de Montpellier).
Traits : noir (observé), rouge (simulé)

Si les gens ne pensent pas que les mathématiques sont simples, c'est qu'ils ne réalisent pas à quel point le monde est complexe.

J. Von Neumann, cité by F. L. Alt, 1972

Le monde des incertitudes

L'environnement est très mal connu

- ▶ Les processus environnementaux sont en général complexes :
 - Non linéaires
 - Hétérogènes et anisotropes
- ▶ Les variables environnementales sont entachées de bruits et d'incertitudes considérables

Apprendre par l'exemple

Sans connaissance sur la physique en jeu

- ▶ Il est possible de concevoir des modèles de simulation et des modèles prédictifs :
 - Grâce à une base de données entrées-sorties
 - En utilisant les réseaux de neurones artificiels
- ▶ Ces modèles identifient la relation sous-jacente dans les données
- ▶ Ils sont validés de manière statistique

Modélisation hydro(géo)logiques

12 ans de prévision opérationnelle des crues/étiages

- ▶ SCHAPI (Service Central d'Hydrométéorologie et d'Appui à la Prévision des Inondations)
 - 4 thèses pour la prévision des crues rapides par réseaux de neurones
 - Des prototypes opérationnels pour trois Services de Prévision des Crues (Artois-Picardie, Grand Delta, Meuse-Moselle)
- ▶ Projet ANR FLASH : intégrer l'assimilation de données
- ▶ Agence de l'eau : prévision des débit d'étiages de la Cèze

Relations avec les entreprises

- ▶ Sur 7 thèses : 6 doctorants ont été embauchés par des PME ou ont créé leur entreprise
- ▶ En cours avec AQUASYS : thèse CIFRE sur la modélisation hydrogéologique

Pour en savoir plus :

- ▶ D. Bertin et al. Neural networks-based operational prototype for flash flood forecasting: application to *Liane* flash floods (Fr); FloodRisks 2016.
- ▶ F.L. Alt. Archaeology of computers: Reminiscences, 1945—1947. In Communications of the ACM, special issue: Twenty-fifth anniversary of the Association for Computing Machinery, vol. 15 7, pp. 693-694 (1972)
- ▶ Guillaume Artigue, Prévision des Crues Éclair par Réseaux de Neurones : Généralisation aux Bassins non Jaugés. Thèse de l'Université Montpellier 2, spécialité Eaux Continentales et Sociétés. Ecole Doctorale SIBAGHE. Soutenue le 3 décembre 2012
- ▶ Thomas Darras, Prévision des crues rapides par apprentissage statistique. Thèse de l'Université Montpellier 2, spécialité Eaux Continentales et Sociétés. Ecole Doctorale SIBAGHE. Soutenue le 2 novembre 2014.
- ▶ M. Virgile Taver, Modélisation hydrodynamique des aquifères karstiques par réseaux de neurones. Thèse de l'Université Montpellier 2, spécialité Eaux Continentales et Sociétés. Ecole Doctorale SIBAGHE. Soutenue le 16 décembre 2012

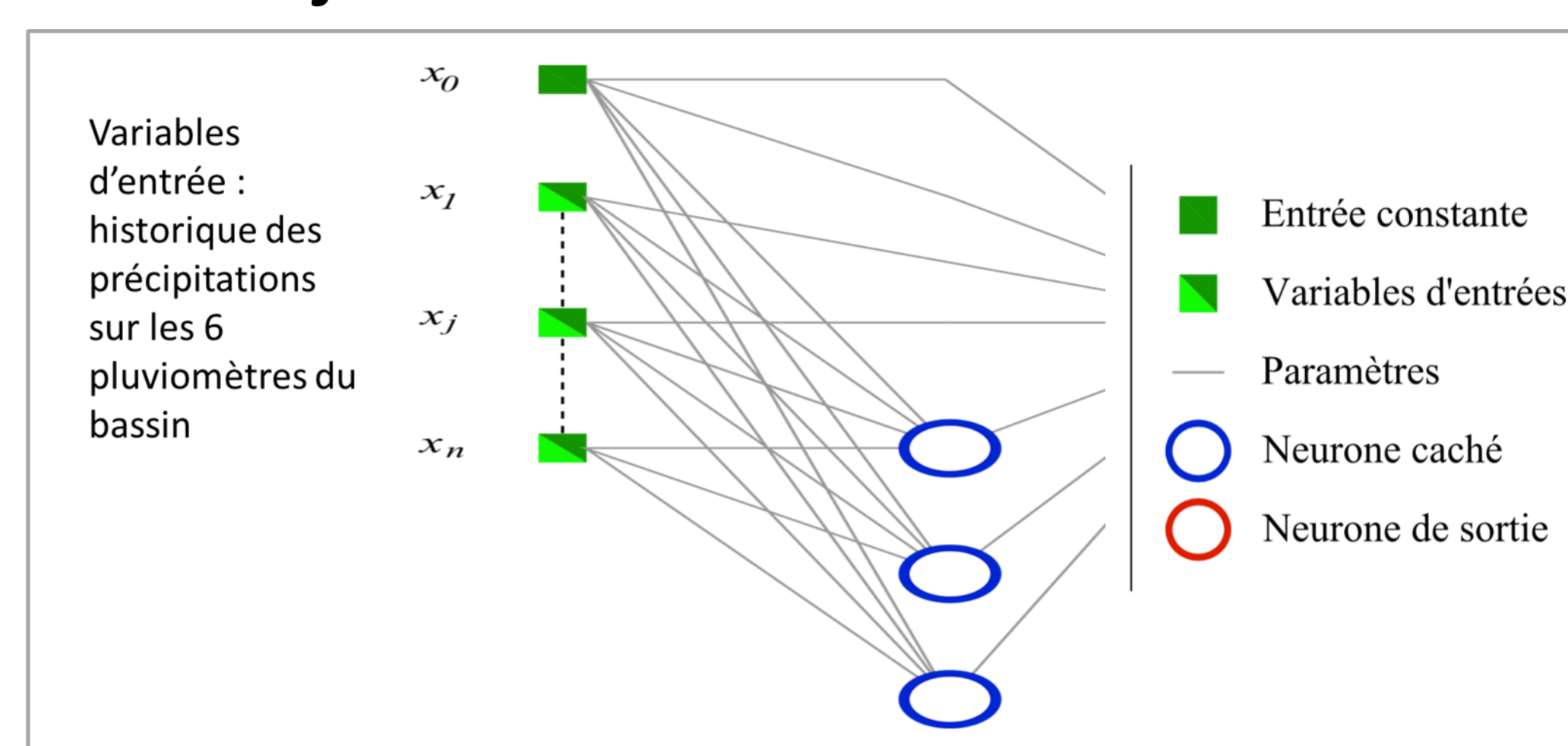
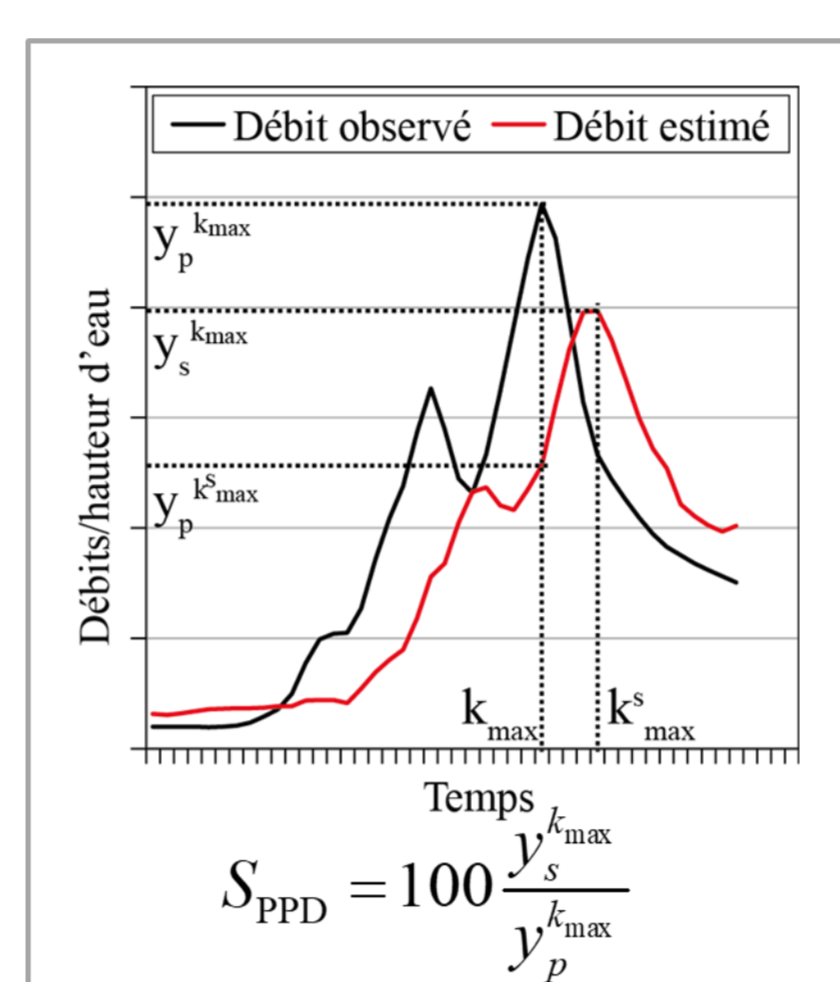
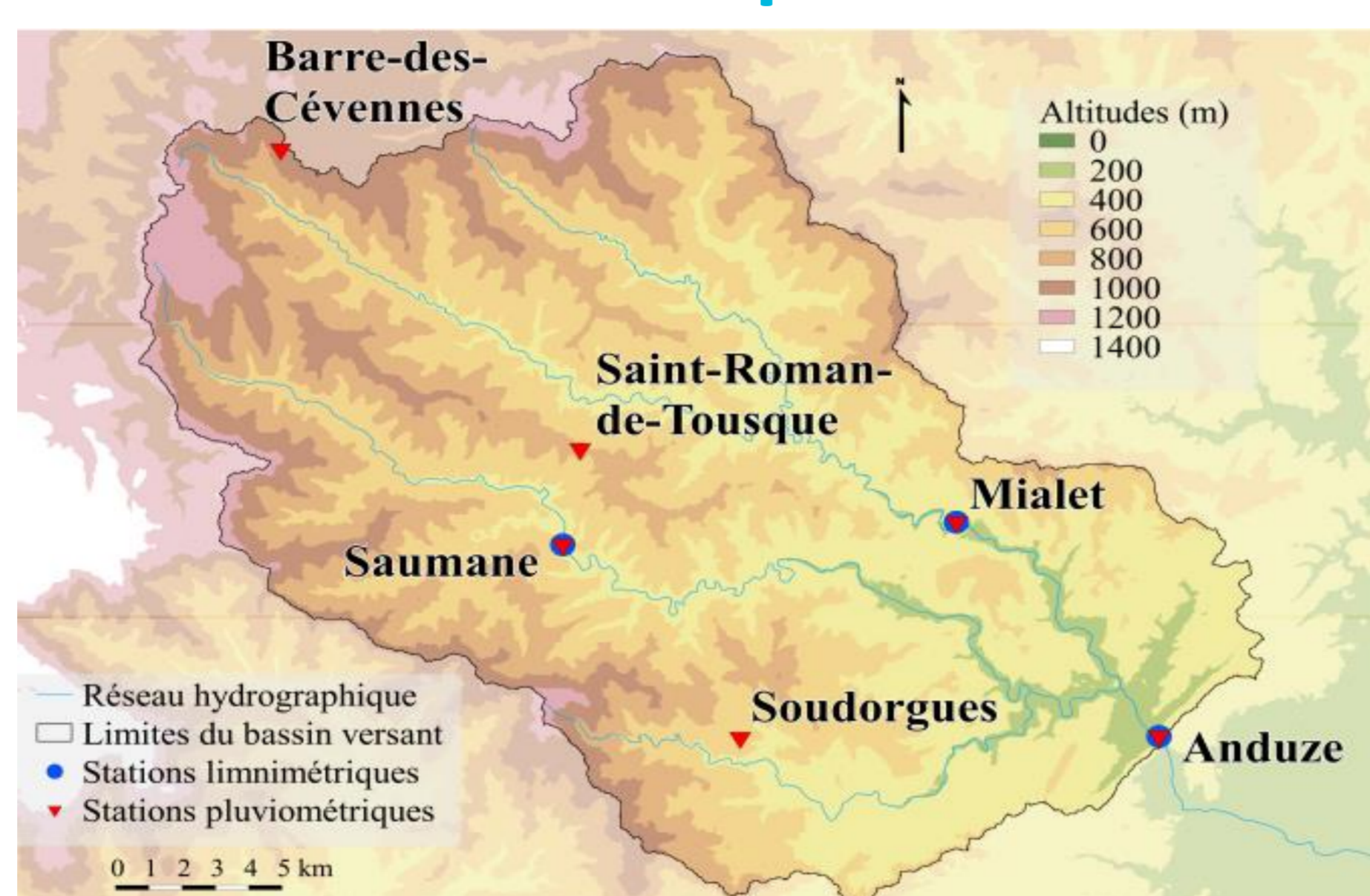


Contacts : anne.johannet@mines-ales.fr, guillaume.artigue@mines-ales.fr, marc.vinches@mines-ales.fr

Définition de modèles de prévision, temps réel, à très courte échéance (2h) :

Les réseaux de neurones artificiels sont utilisés depuis plus de 25 ans pour la modélisation hydrologique. Dans le cadre d'une collaboration avec le SCHAPI, ils ont été appliqués à la prévision en temps réel des hauteurs d'eau à Anduze.

Prévoir les crues rapides à Anduze : test sur l'événement majeur de 2002

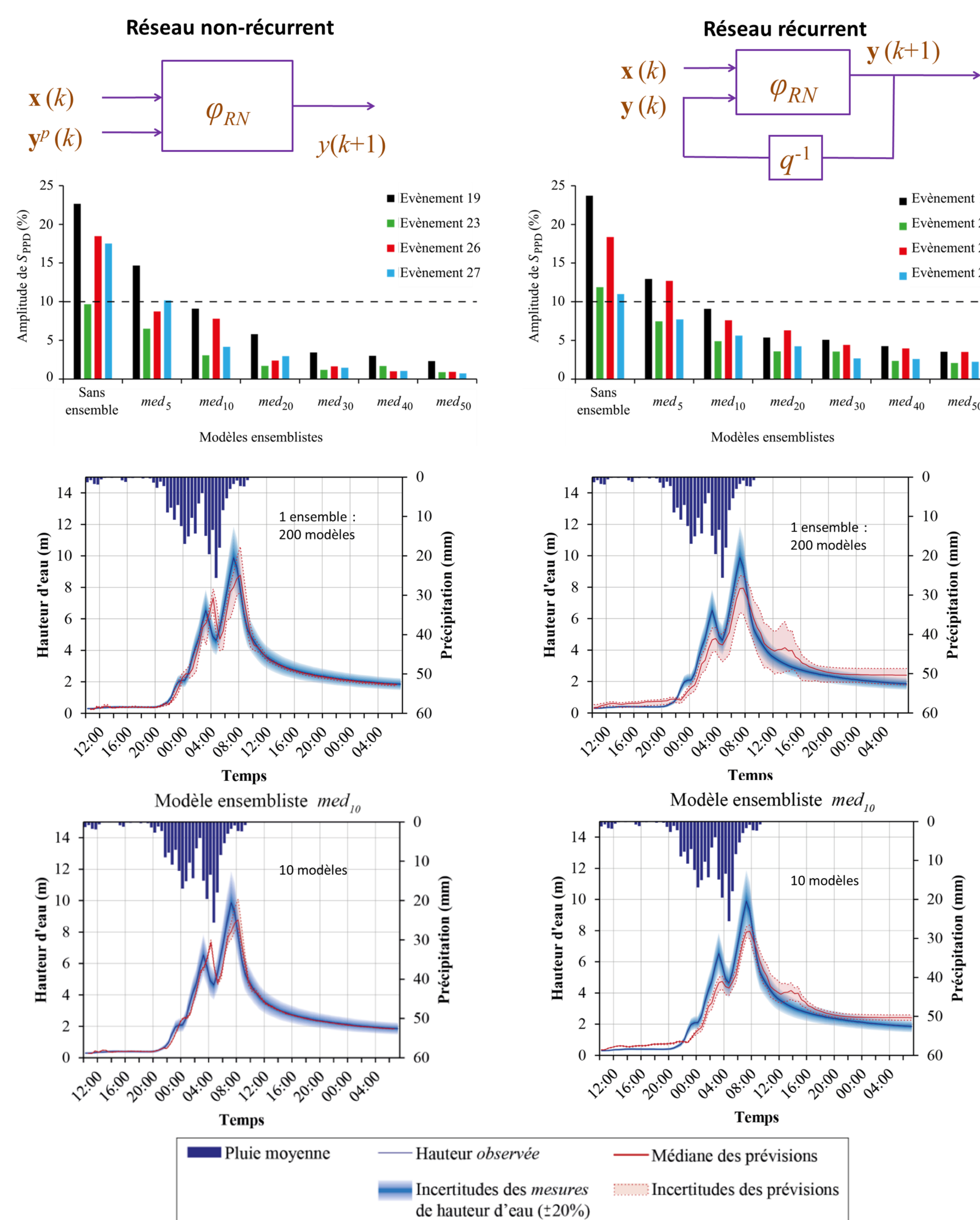


51 événements dont 19 intenses, 9m88 de hauteur maximum, soit 2965 m³/s

Pallier l'incertitude liée aux paramètres : améliorer la robustesse

Avant l'apprentissage, le modèle doit être initialisé avec des paramètres aléatoires : il est montré que les performances du modèle peuvent énormément dépendre de cette initialisation. Un modèle d'ensemble est donc conçu :

- M modèles sont créés et appris
- La sortie à chaque pas de temps est la médiane des M sorties
- Que vaut le M optimal ?



Prévision à 1 h à Anduze – événement de septembre 2002, Darras 2015

Contact : guillaume.artigue@mines-ales.fr

Parties prenantes



Auteurs

Thomas Darras
Guillaume Artigue
Bernard Vayssade
Anne Johannet
Line Kong-A-Siou
Séverin Pistre
Dominique Bertin

Partenaires



Mines Alès est certifiée par



Pour en savoir plus :

Guillaume Artigue

Prévision des crues éclair par réseaux de neurones : généralisation aux bassins non jaugés. Thèse de l'Université Montpellier 2, spécialité Eaux Continentales et Sociétés. Ecole doctorale SIBAGHE. Soutenue le 3 décembre 2012.

Audrey Bornancin-Plantier

Conception de modèles de prévision des crues éclair par apprentissage artificiel. Thèse de l'Université Pierre et Marie Curie, spécialité Informatique. ED EDITE, soutenue le 25 février 2013.

Thomas Darras

Prévision des crues rapides par apprentissage statistique. Université de Montpellier II. Ecole Doctorale SIBAGHE, spécialité Eaux Continentales et Sociétés, soutenue le 3 novembre 2015.

Parties prenantes



Auteurs

Michel Aldanondo
Elise Vareilles
Paul Pitiot
Paul Gaborit

Projets



ATLAS et OPERA
Configuration de produits et services en manufacturier



HELIMAINTENANCE
Configuration de processus maintenance en aéronautique

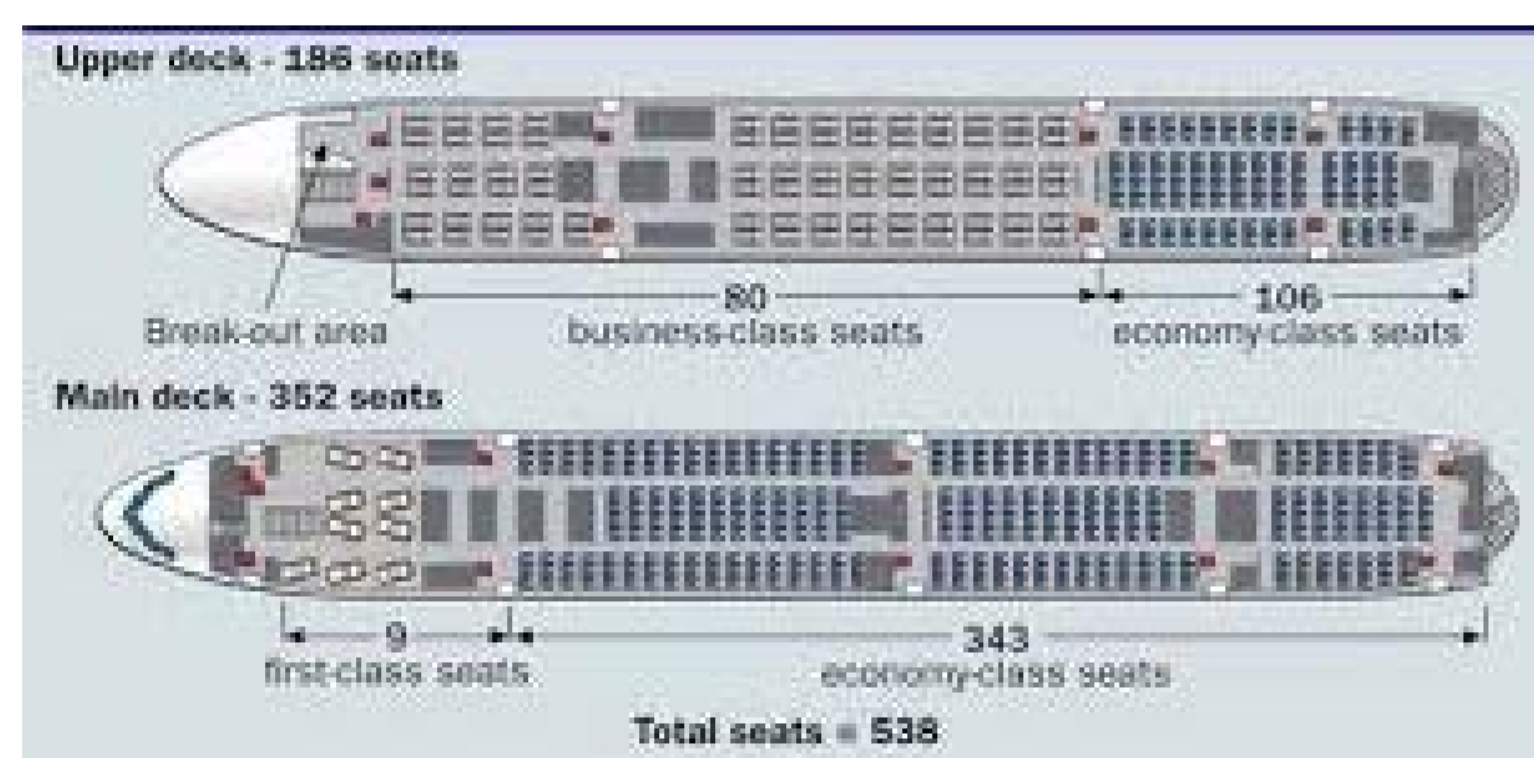


Agence de l'Environnement et de la Maitrise de l'Energie

CRIBA
Configuration de rénovation énergétique en génie civil

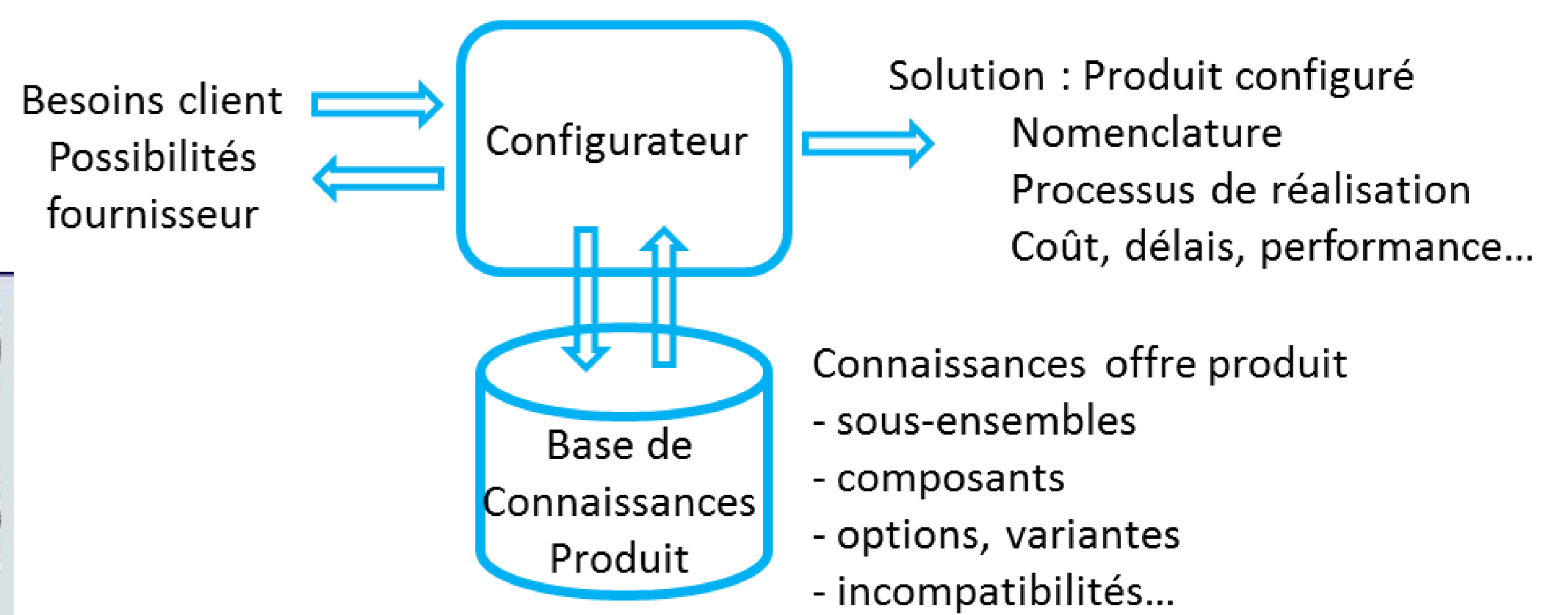
Et les laboratoires et entreprises des consortiums de ces projets

Pôles



Configuration de produit/système/service

- **Éléments de définition** : La configuration est un type de conception où le produit est défini comme un assemblage de composants standards prédéfinis sujet à des contraintes de compatibilité variées.
- **Applications** : Configuration de véhicules, d'ordinateurs, de cuisines, de bibliothèques, de vélos, de chaussures, de cabines d'avion, de grues...
- **Situation** : En relation client fournisseur entre : le commercial (CRM), les études (CAD) et la production (ERP)



- **Les outils d'aide à la configuration** : Raisonnement par analogie, Inférence ontologique, Propagation de contraintes (le plus fréquent)

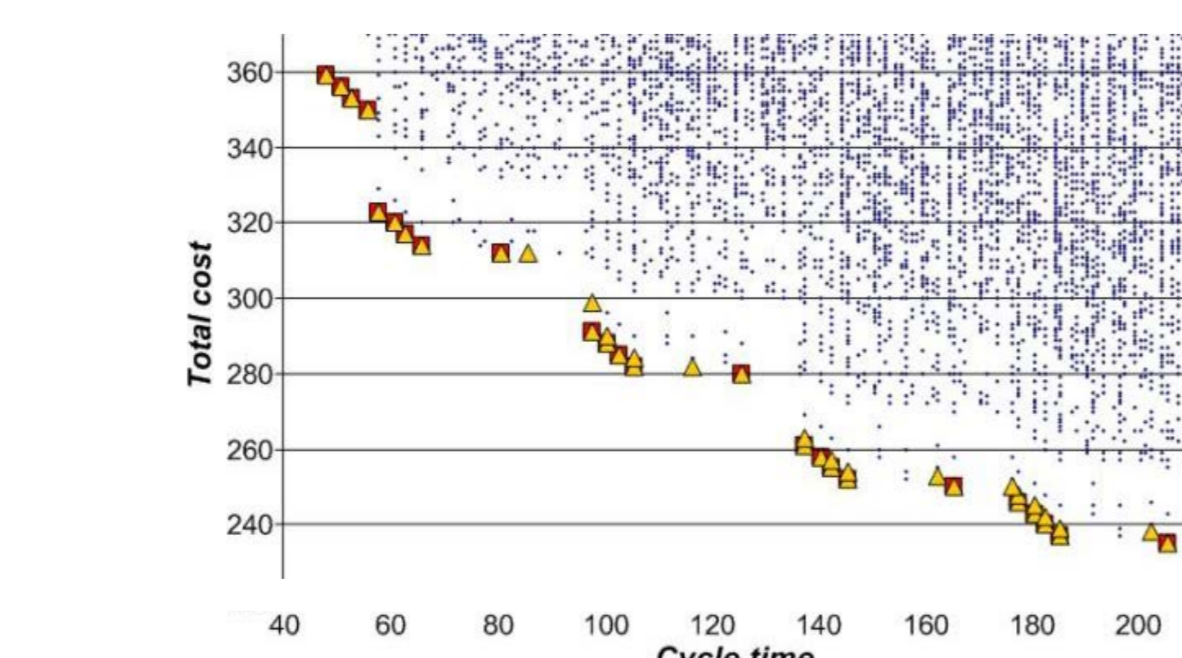
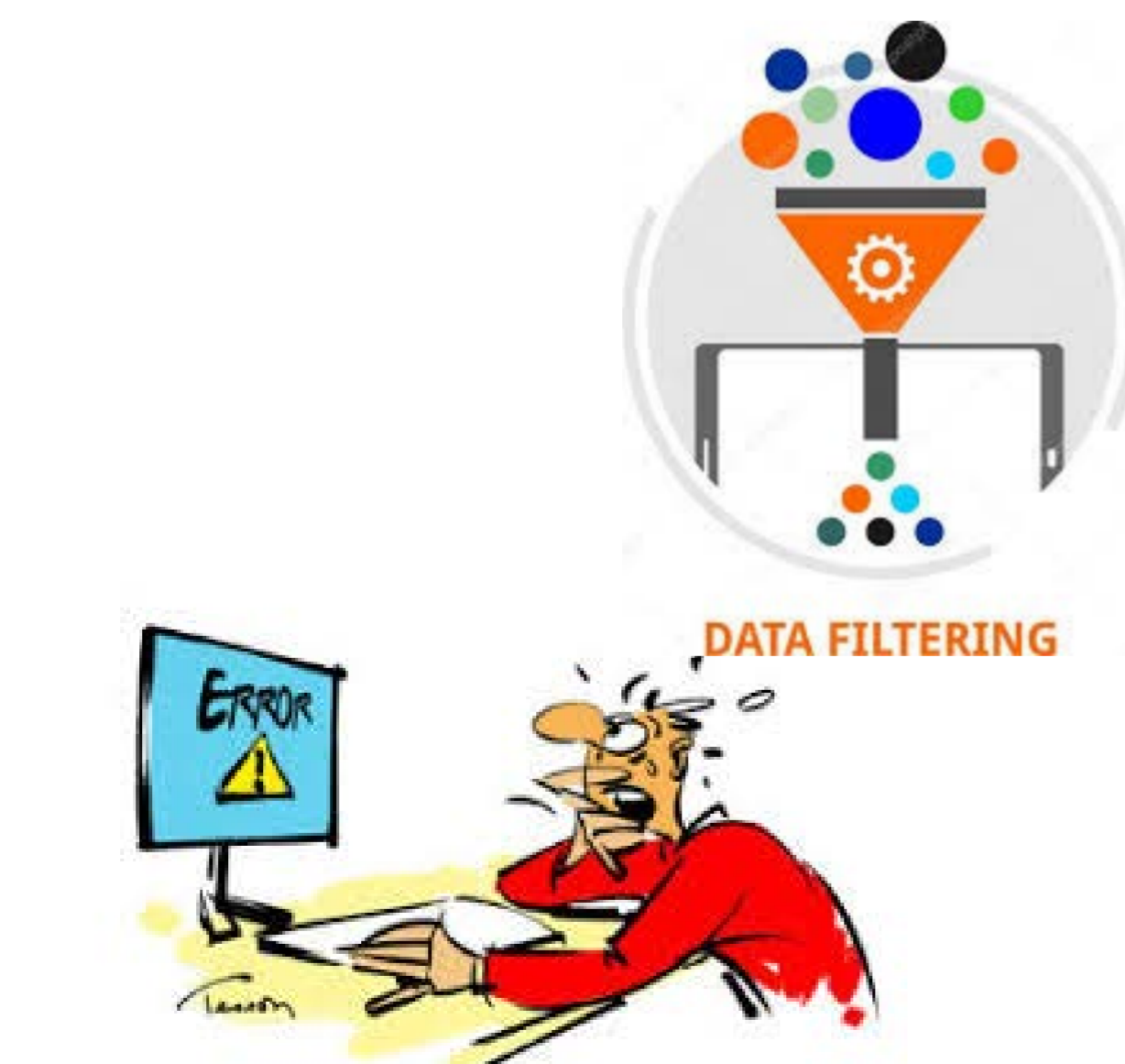
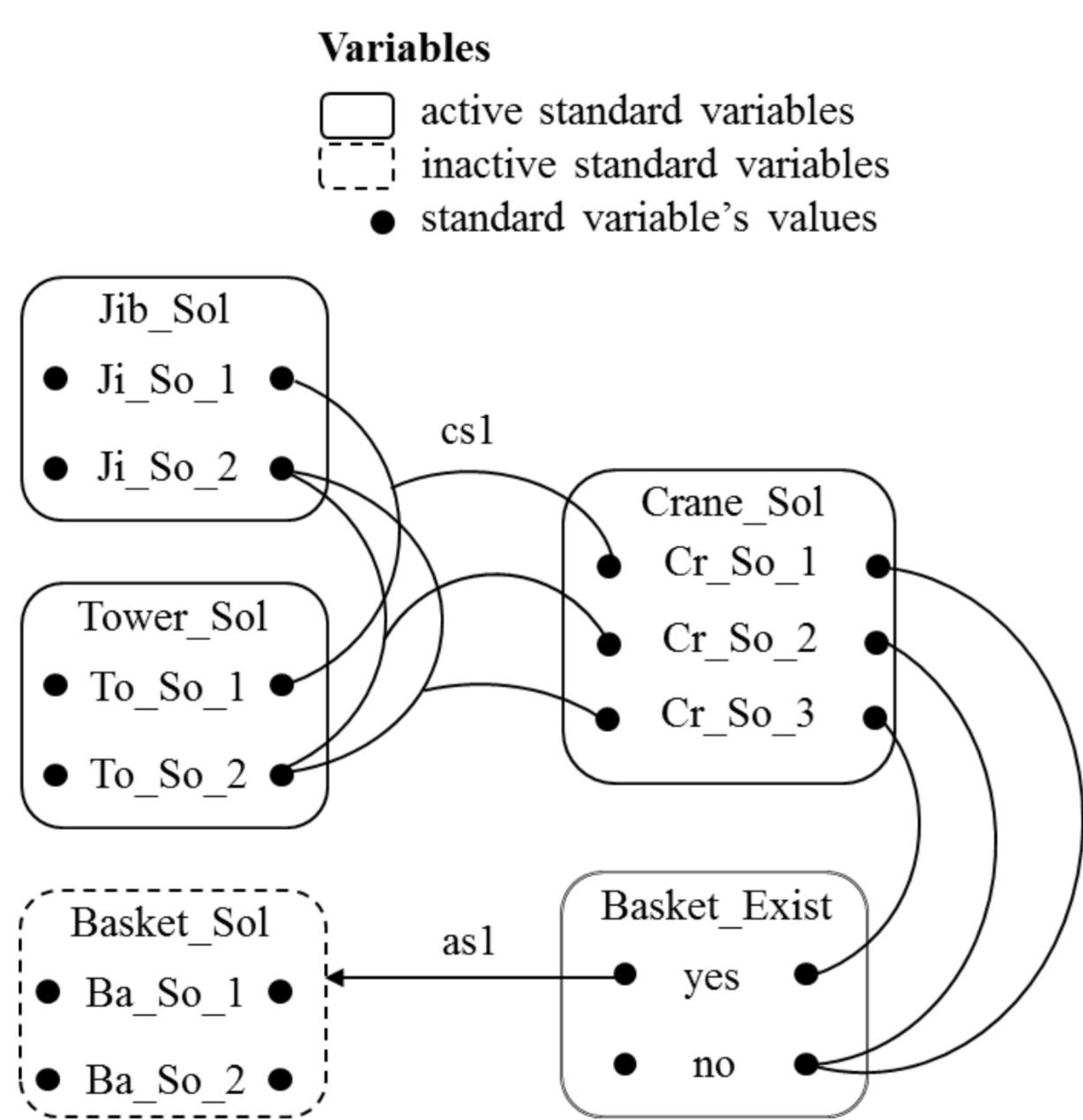
One of the AI technologies that successfully found its way into industrial applications is constraint satisfaction (AI magazine Winter 2016 p 67)

Configuration et approches par contraintes (CSP)

Type de problèmes	CSP discret, Sélection de composant	CSP mixte	CSP numérique
Type d'aide à la configuration	Interactive, propagation Propagation faible		Autonome, résolution Peu de solutions exactes
Traitement incohérence	Lors de la modélisation Problème incohérent		Lors de la configuration Solution incohérente
Type de raisonnement	Sur le problème Approche conventionnelle		Sur l'espace de solutions Diversité compilée
Optimisation	Exacte Difficile		Méta heuristique Le plus souvent

Quelques problème et travaux

- **Sur la modélisation des problèmes de configuration**
Architecture des modèles de configuration
Extension des modèles de produit aux systèmes, procédés et services
Prise en compte de composants et sous-ensembles non standards
Détection d'incohérences lors de la modélisation
Extraction construction automatique de modèle
Cohérence des modèles de configuration avec CRM, CAD et ERP
...
- **Mise au point de moteur d'aide à la configuration**
Association filtrage de contraintes discrètes et numériques
Explication d'incohérences lors de la configuration
Association filtrage conventionnel et diversité compilée
Association configuration et recommandation
Optimisation évolutionnaire multicritères
...



Parties prenantes



Auteurs

Gauthier Picard

Henri Fayol Institute
Laboratoire Hubert Curien
CNRS UMR 5516

picard@emse.fr

Pierre Rust

Orange Labs
Laboratoire Hubert Curien
CNRS UMR 5516

pierre.rust@orange.com

Fano Ramparany

Orange Labs
fano.ramparany@orange.com



Partenaires



Coordination distribuée et autonome pour l'IoT

Motivations

- ▶ 25 milliards d'objets en 2020
- ▶ Appareils bon marché mais contraints
- ▶ Aujourd'hui : coordination centralisée au niveau du cloud

Coordination décentralisée

- ▶ Montée à l'échelle, interactions locales
- ▶ Pas de point d'échec central (SPOF)
- ▶ Pas de goulet d'étranglement pour les communications

DCOP : Problème d'optimisation distribuée sous contraintes

Un tuple $\langle \mathcal{A}, \mathcal{X}, \mathcal{D}, \mathcal{C}, \mu \rangle$

- ▶ $\mathcal{A} = \{a_1, \dots, a_{|\mathcal{A}|}\}$, est un ensemble d'agents
- ▶ $\mathcal{X} = \{x_1, \dots, x_n\}$, est un ensemble de variables
- ▶ $\mathcal{D} = \{\mathcal{D}_{x_1}, \dots, \mathcal{D}_{x_n}\}$, les domaines de variables, $\mathcal{D}_{x_i} = \{v_1, \dots, v_k\}$
- ▶ $\mathcal{C} = \{c_1, \dots, c_m\}$, est un ensemble de contraintes souples, c_i est une fonction de coût $\in \mathbb{R} \cup \{\infty\}$ pour toutes affectations aux variables dans sa portée
- ▶ $\mu : \mathcal{X} \rightarrow \mathcal{A}$, une fonction assignant chaque variable à un calcul

Solution : une affectation à toutes les variables qui minimise la somme des coûts $\sum_i c_i$

Algorithmes par envoi de messages

- ▶ **Complets** ADOPT, DPOP, etc.
- ▶ **Approchés** MGM, DSA, Max-Sum, etc.

pyDCOP une librairie d'étude DCOP

Fonctionnalités

- ▶ Implémentation des algorithmes majeurs : DSA, DPOP, MGM, Max-Sum, etc.
- ▶ Utilitaires pour le développement de nouveaux algorithmes
- ▶ Fournit toute l'infrastructure logicielle pour l'étude des algorithmes DCOP: instanciation des agents, déploiement de l'algorithme, contexte d'exécution, envoi des messages, etc.
- ▶ Générateur de problèmes de benchmark
- ▶ Monitoring et production de métriques: temps de résolution, nombre de cycles, charge réseau, coût et qualité de la solution, etc.

pyDCOP pour l'IoT

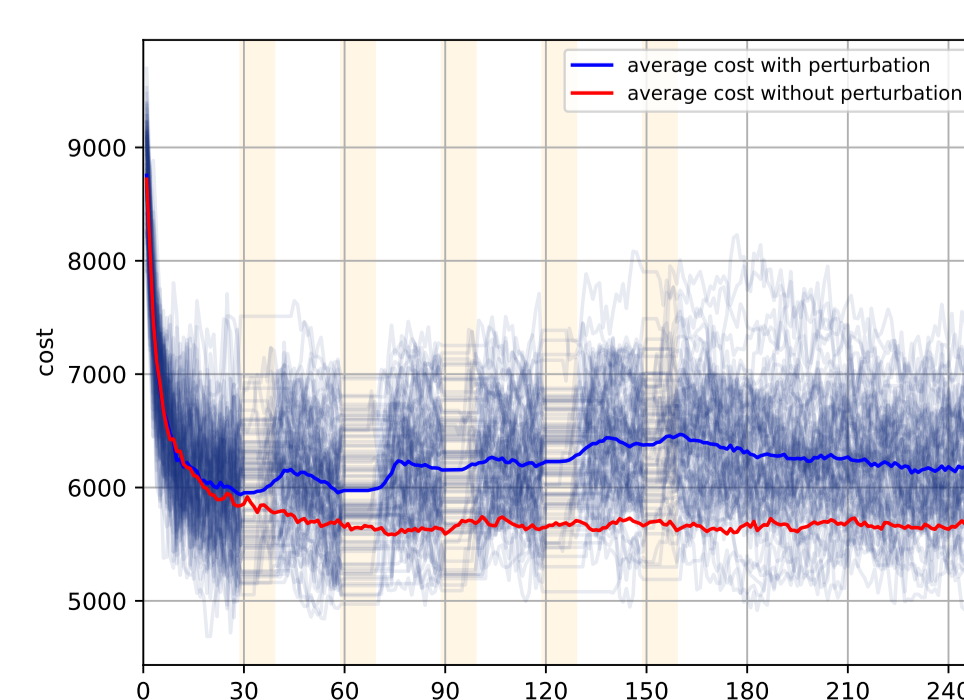
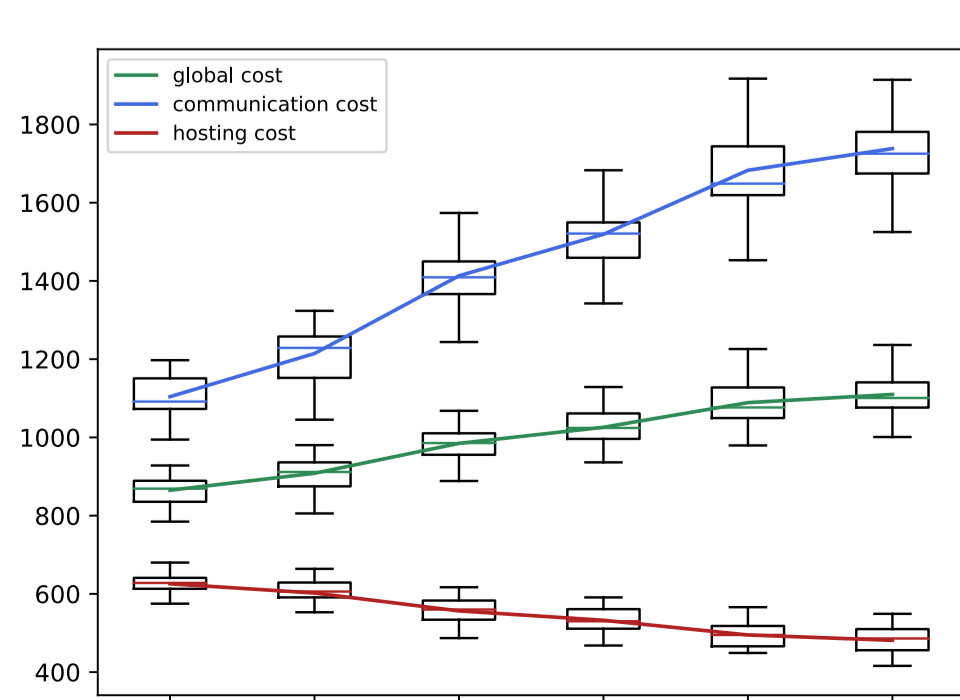
- ▶ Prise en compte de la problématique de la distribution des décisions sur l'infrastructure : optimisation du placement en fonction des coûts de communication, d'hébergement, etc.
- ▶ Résilience du système

Prototypage avec pyDCOP

- ▶ Multi-plateforme
- ▶ Permet un déploiement *réel*: PC, Machine virtuelle, Raspberry Pi
- ▶ Interface graphique de monitoring

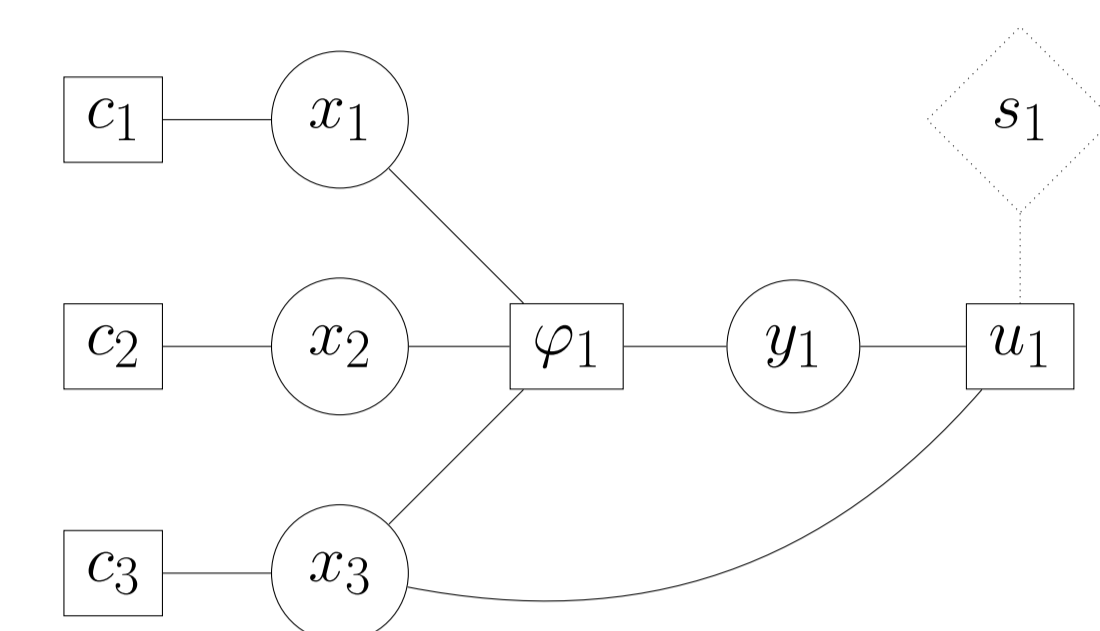
Utilisation

- ▶ Source: <https://github.com/Orange-OpenSource/pyDcop>
- ▶ Documentation: <https://pydcop.readthedocs.io>
- ▶ en python



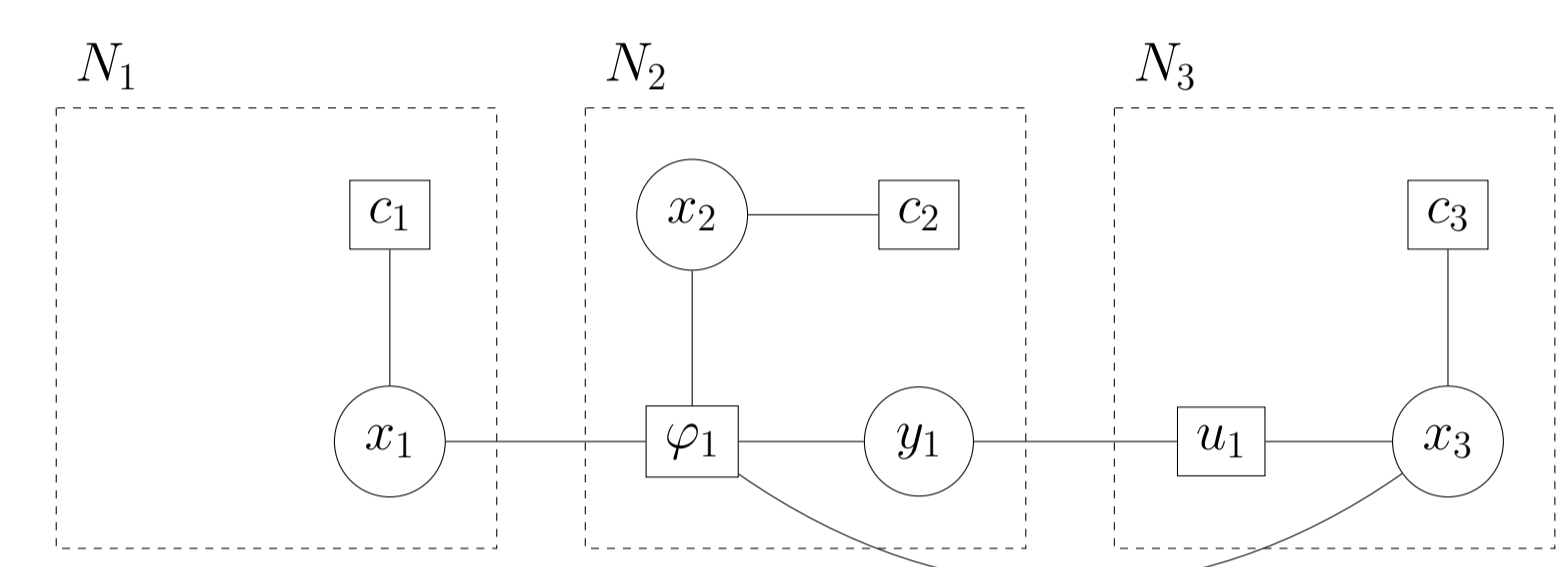
Distribution et résolution d'un DCOP

Représentation en graphe de calculs



- ▶ Graphe de facteurs
- ▶ Graphe de contraintes

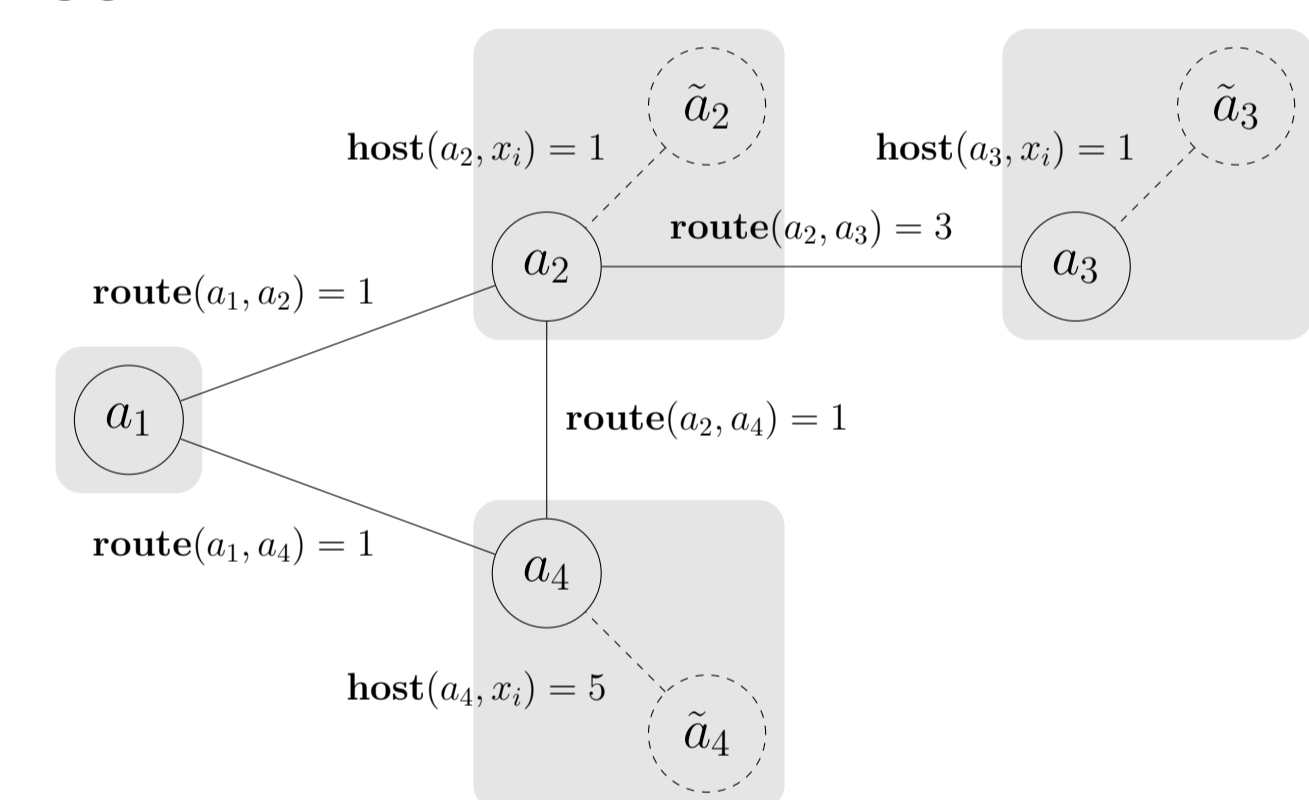
Distribution du graphe de calculs



Un problème de partitionnement de graphe

- ▶ Heuristiques: garder les calculs proches des variables impactées
- ▶ Programmation Linéaire : minimiser les coûts de communication, d'hébergement, en respectant les capacités matérielles des agents
- ▶ etc.

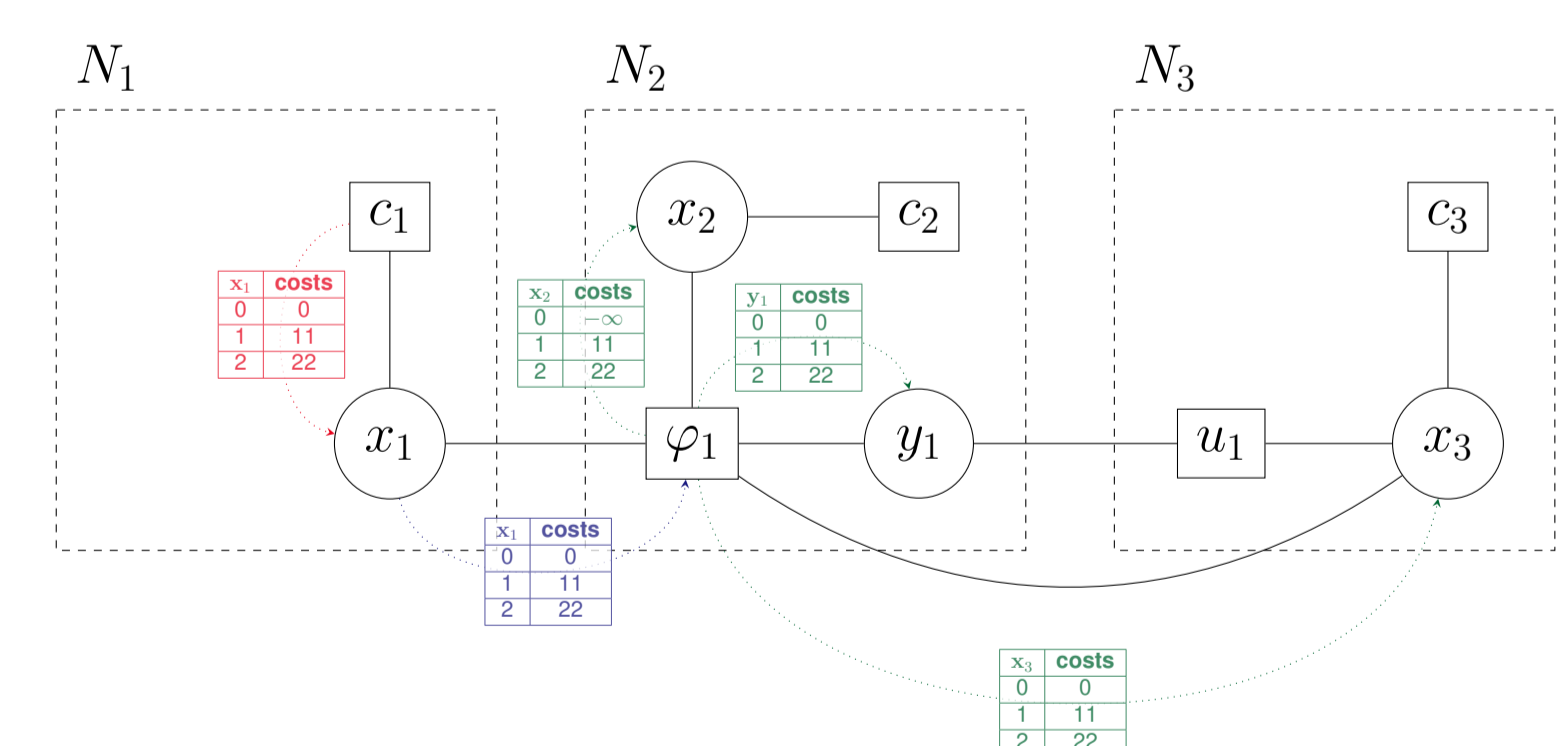
K-Résilience



Survivre à la disparition d'éléments de l'infrastructure

- ▶ Réplication des calculs
- ▶ Réparation en ligne du déploiement

Résolution



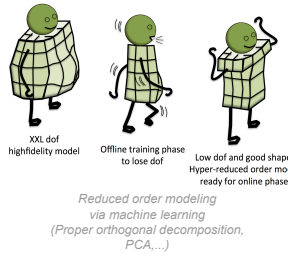
Avec des algorithmes DCOP asynchrones supportant la perte de messages.

- ▶ MaxSum, A-DSA pour le problème initial
- ▶ MGM2 pour la réparation du déploiement

Références

- [1] P. Rust, G. Picard, and F. Ramparany. "Installing Resilience in Distributed Constraint Optimization Operated by Physical Multi-Agent Systems". In: *Autonomous Agents and Multiagent Systems (AA-MAS)*. 2019.
- [2] P. Rust, G. Picard, and F. Ramparany. "On the Deployment of Factor Graph Elements to Operate Max-Sum in Dynamic Ambient Environments". In: *Autonomous Agents and Multiagent Systems – AA-MAS 2017 Workshops, Best Papers, Sao Paulo, Brazil, May 8-12, 2017, Revised Selected Papers*. Vol. 10642. LNAI. Springer, 2017, pp. 116–137.
- [3] P. Rust, G. Picard, and F. Ramparany. "Using Message-passing DCOP Algorithms to Solve Energy-efficient Smart Environment Configuration Problems". In: *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16)*. AAAI Press, 2016, pp. 468–474.

Computer Vision for model order reduction in mechanics of materials with error estimation



Surrogate modeling in continuum mechanics Hybrid Machine-learning/Physics Modeling

- ▶ **Physics informed machine learning** – when data science do not ignore the available knowledge in physics.
- ▶ **Non parametric modeling with computer vision** – 3D tomographic data give access to complex geometries and heterogeneities in mechanics.
- ▶ **Defect tolerance via computer vision** – true geometries have their own mechanical properties (stiffness, life duration, ...).
- ▶ **Model order reduction (ROM)** – Because too long predictions of mechanical properties by solving partial differential equations with physics-based modeling is not consistent with computer vision.
- ▶ **ROM classification by using machine learning** – Usual ROM fail to simplify non parametric models. *Non parametric classification by using Convolutional Neural Network can help.*

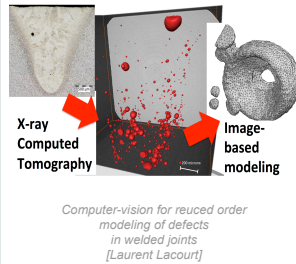
Parties prenantes



Auteurs

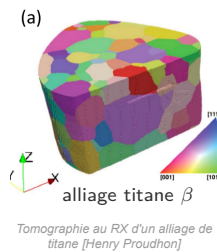
David Ryckelynck
Henry Proudhon
Laurent Lacourt
Franck Nguyen
Basile Marchand
Pierre Kerfriden
Djamel M. Benziane

Partenaires



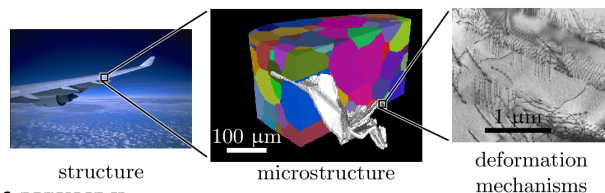
Example of non parametric modeling Modeling the entanglement of scales by machine learning

- ▶ **Scale separation** is assumed during machine learning stage
- ▶ **Data lake** containing individual modeling of each defect
- ▶ **Clustering** of defects according to their morphology and their mechanical properties
- ▶ On the fly assembly of a **component-specific reduced-order-model**, without any scale separation assumption
- ▶ **Fast prediction** of defect nocivity
- ▶ **Time and money savings**, when defects have an admissible nocivity
- ▶ **Cumulative learning** by saving usefull data or by pruning data



Chaire de Mécénat – Safran – Fondation Mines ParisTech Big Méca (2019-2024)

- ▶ Acquisition of large **4D tomographic data** for fracture mechanics
- ▶ **Image-based modeling** in mechanics of materials
- ▶ **Data driven modeling** of fracture for polycrystalline materials
- ▶ **Bayesian neural network** for data driven modeling of fracture. Many parameters interplays in fracture of polycrystalline materials. The data driven approach aims to identify the most influent parameters.
- ▶ Application in aeronautics: **defect inspection** via hybrid ML/Physics modeling



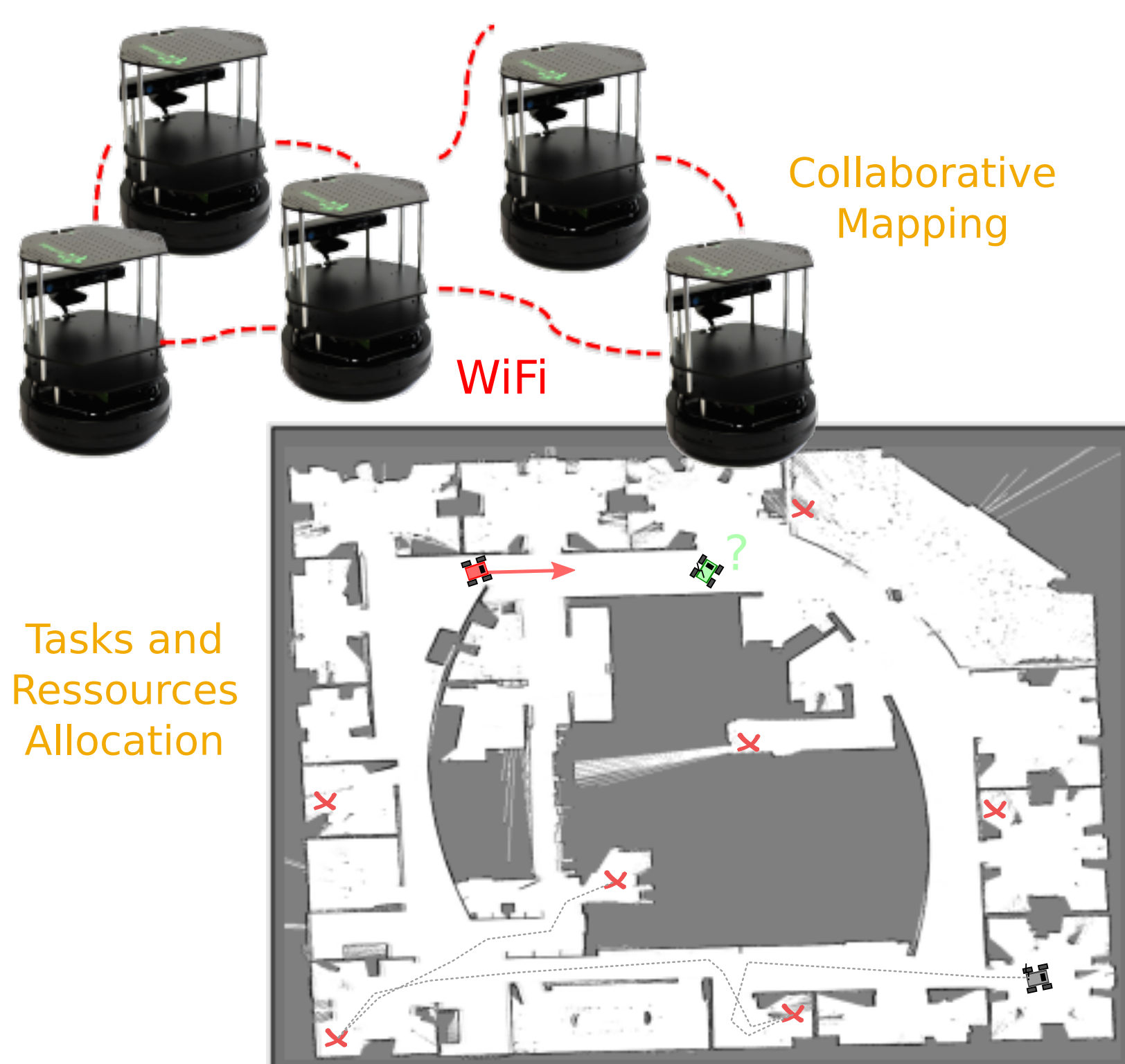
Contact : david.ryckelynck@mines-paristec.fr

Ours application domains

Confront Theory with Real Control Scenarios

Multi-Robot exploration

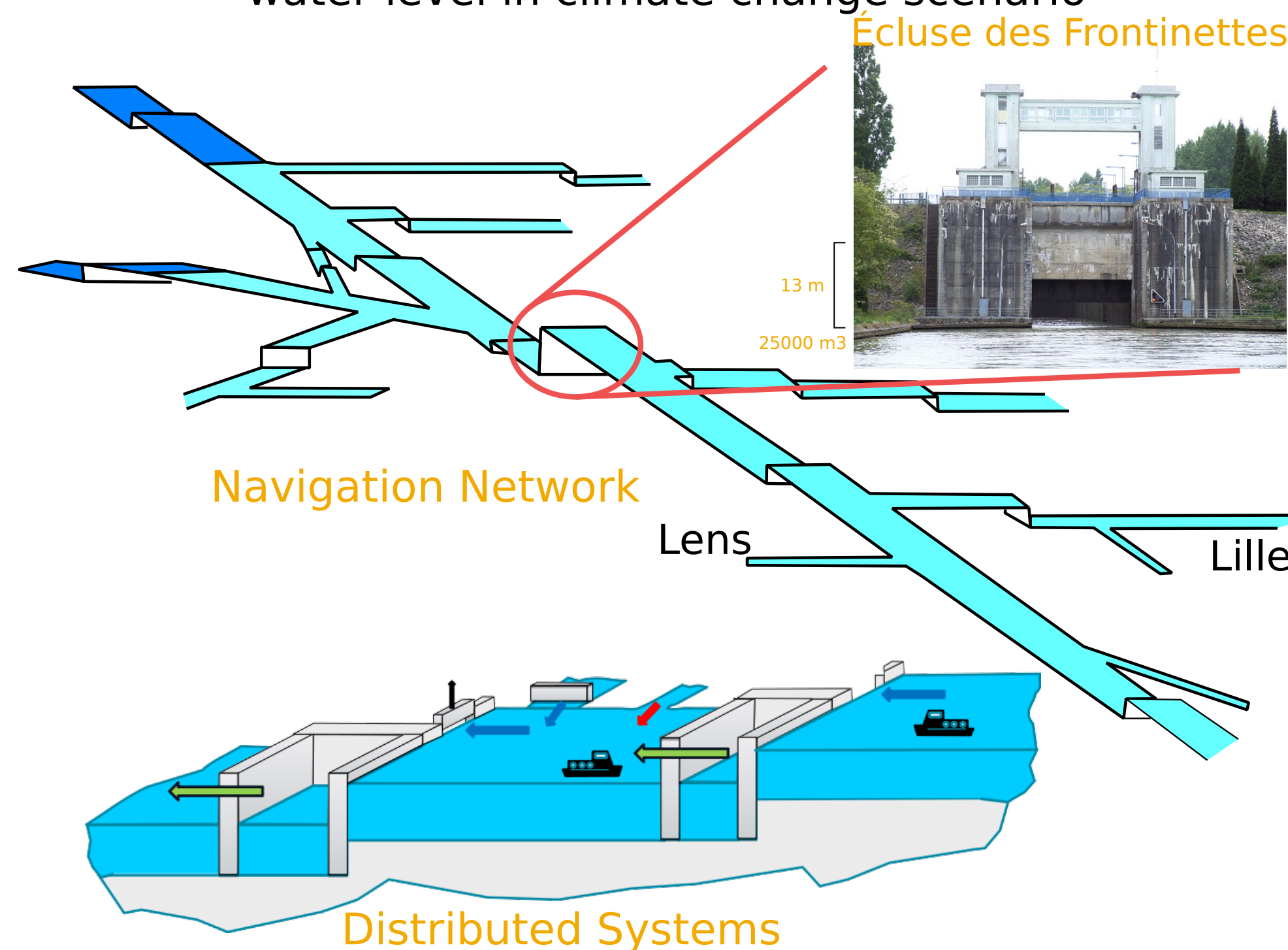
Share robot knowledge, allocate exploration tasks and plan movements



PhD 2019: Johann Ditchl
Post.Doc 2018-2019: Xuan Sang Le

Water Management in Inland navigation network

Coordinate local managers to keep appropriate water level in climate change scenario



PhD 2018: Guillaume Desquesnes
PhD 2022: Debora Alves

Fleet of Intelligent Vehicules

Study new behavior that will appear due to new technology like communication vehicules in actual transportation networks

PhD 2021: Azise Oumar Diallo

Polution control with Aquatic Drones

Decentralised Decision Making in the conception of hot water cylinder

Research center



IMT Lille Douai
École Mines-Télécom
IMT-Université de Lille

Authors

Noury Bouraqadi
Arnaud Doniec
Guillaume Lozenguez
Marin Lujak
Éric Duviella
Luc Fabresse

Partenaires

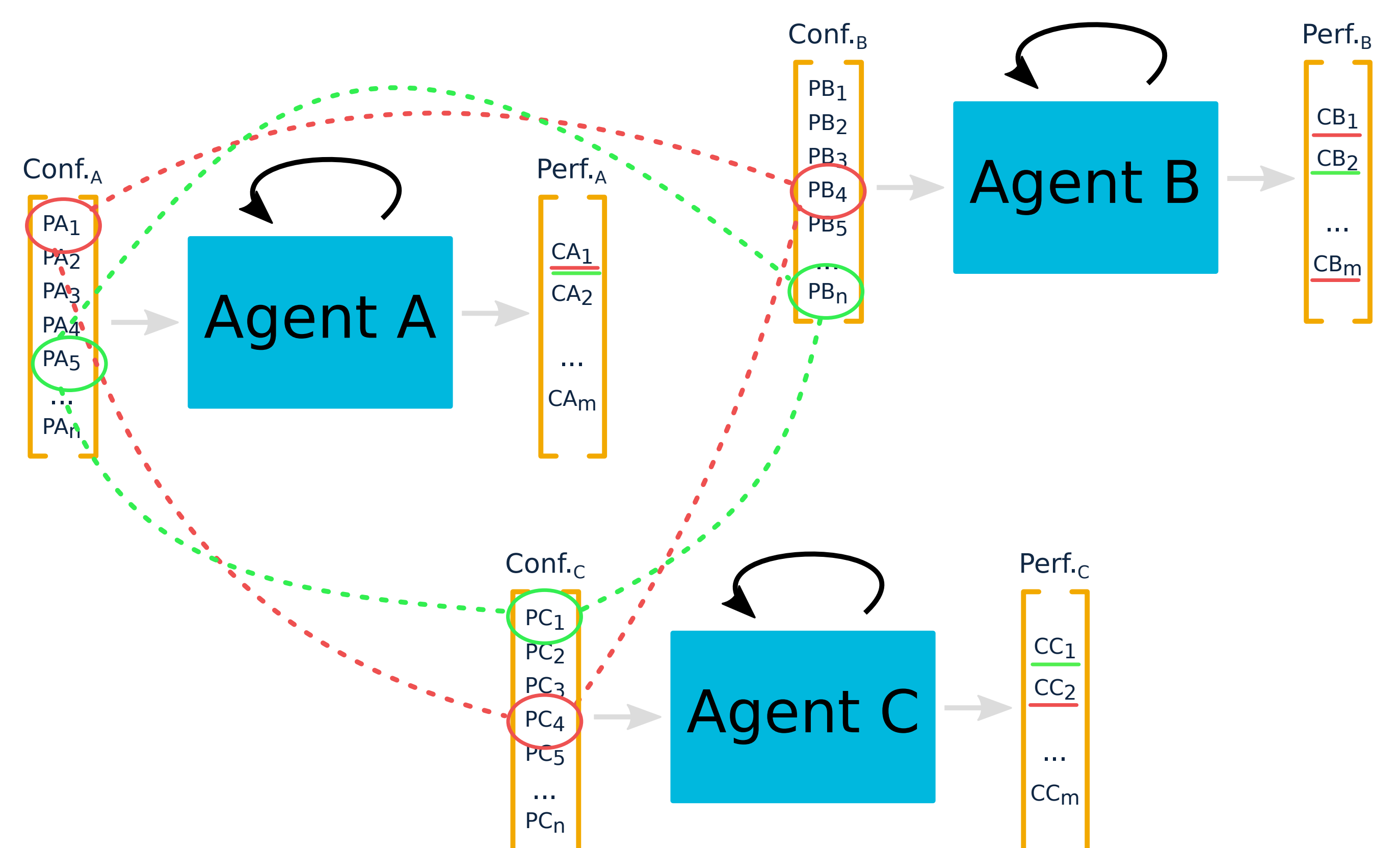


Complex system modeled as interacting components

Coordinating Individual Decision Processes

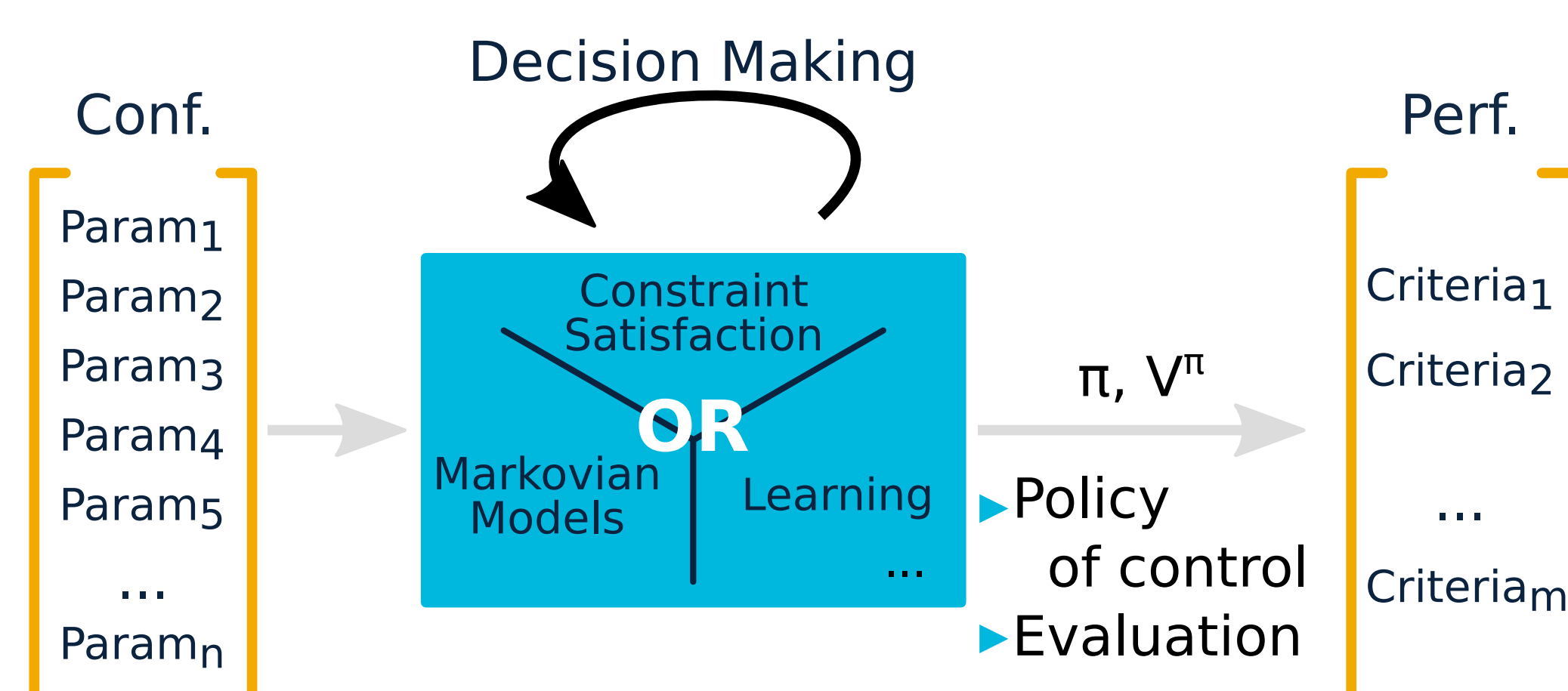
Optimizing configurations:

- Deal with shared parameters with Coordination Protocol (e.g Auction)
- Apprehend Complex systems with Agent Based Modeling
- Handle dynamic environment with Continous and Reactive Coordination



Autonomous and intelligent components

Modeling Individual Decision Making Problems



Propositions:

- Dynamic evolution with Markov Decision Processes
- Multi-variable system with Dynamic Bayesian Network
- Large state space with Decomposition and Monte-Carlo heuristics
- In-line constraint with Learning and Hybrid Architectures

Définition de l'assortiment idéal dans un réseau de magasin

De l'Approche Scientifique à la Solution Industrielle

Contexte et objectif général

La grande distribution



- ▶ **L'assortiment** – Représente l'ensemble des produits présents dans un ou plusieurs magasins
- ▶ **Massification** – Les grands distributeurs sont amenés à gérer un ou plusieurs réseaux de magasins
- ▶ **Assortiment centralisé** – La notion d'assortiment centralisé permet à la grande distribution manager un ensemble d'articles sur un sous-ensemble ou l'ensemble du réseaux de magasins
- ▶ **Objectif de l'assortiment idéal** – Proposer le(s) **BON(s)** produit(s) pour le(s) **BON(s)** client(s)
- ▶ **Contraintes générales** – Répondre aux demandes des clients (Unités de besoins, Tranches de prix ...)
- ▶ **Objectif global** – Améliorer la performance du/des magasin(s) (Chiffre d'affaires, Marges, Stock ...)
- ▶ **Cible de l'assortiment** – Vous, moi ... toute personne faisant ses courses dans un supermarché

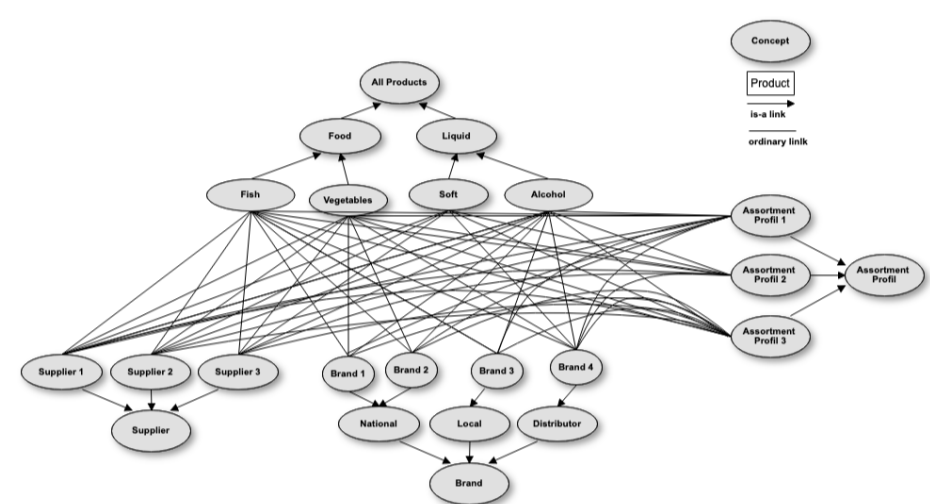
Parties prenantes



IMT Mines Alès
École Mines-Télécom

Auteurs

J. MONTMAIN
J. PONCELET
F. TROUSSET



De l'assortiment OPTIMAL à l'assortiment IDÉAL

Nouveaux contextes et nouvelle approche

- ▶ Sortie d'un contexte purement monétaire, le meilleur assortiment dépend de nombreux facteurs clés :
 - Politique commercial du grand distributeur
 - Localisation géographique
 - Lois locales
 - Concurrence
 - Clientèle cible ...
- ▶ **Ontologie Global et Raisonnement Approché**
 - Exprimer des contraintes *abstraites* (Exemple : - 50% de produits BIO par Rayon, Autant de produits de *Marque Distributeur* que de *Marque National* ...)
 - Habitudes d'achats des consommateurs - Clustering sémantique des clients basé sur la structure marchandise.

Guide méthodologique

L'assortiment idéal dans un réseau de magasins

- ▶ Thématiques de recherche :
 - Analyses Multicritères
 - Optimisation sous-contraintes
 - Optimisation combinatoire
 - Fouille de données (Clustering, Règles d'associations ...)
 - Similarité / Résumé sémantique
- ▶ Optimisation l'assortiment d'un magasin
Localisation, Spécificité, Consommateurs, Cible, Concurrent, Fournisseurs ... du magasin
- ▶ Optimisation l'assortiment centralisé
Détenon des produits, Rupture, Supply chain, Gestion des formats (Hyper, Super, Proxy) pour l'ensemble du réseau de magasin
- ▶ **Le véritable assortiment idéal** c'est l'**Equilibre** entre l'assortiment idéal de chacun des magasins et de l'assortiment centralisé
- ▶ Enrichissement du modèle – **Informations** provenant :
 - d'analyses métiers (Rôles des Catégories, Type de Magasins ...)
 - de données externe (Météo, Données de la concurrence, Open data ...)
 - d'algorithmes scientifiques (Réseaux de neurones, Classification ...)
- ▶ Problématique : Retail - **BIG DATA** (5V)



Modélisation et optimisation de la Configuration de Produit et de la Configuration de Process

Parties prenantes



Auteurs

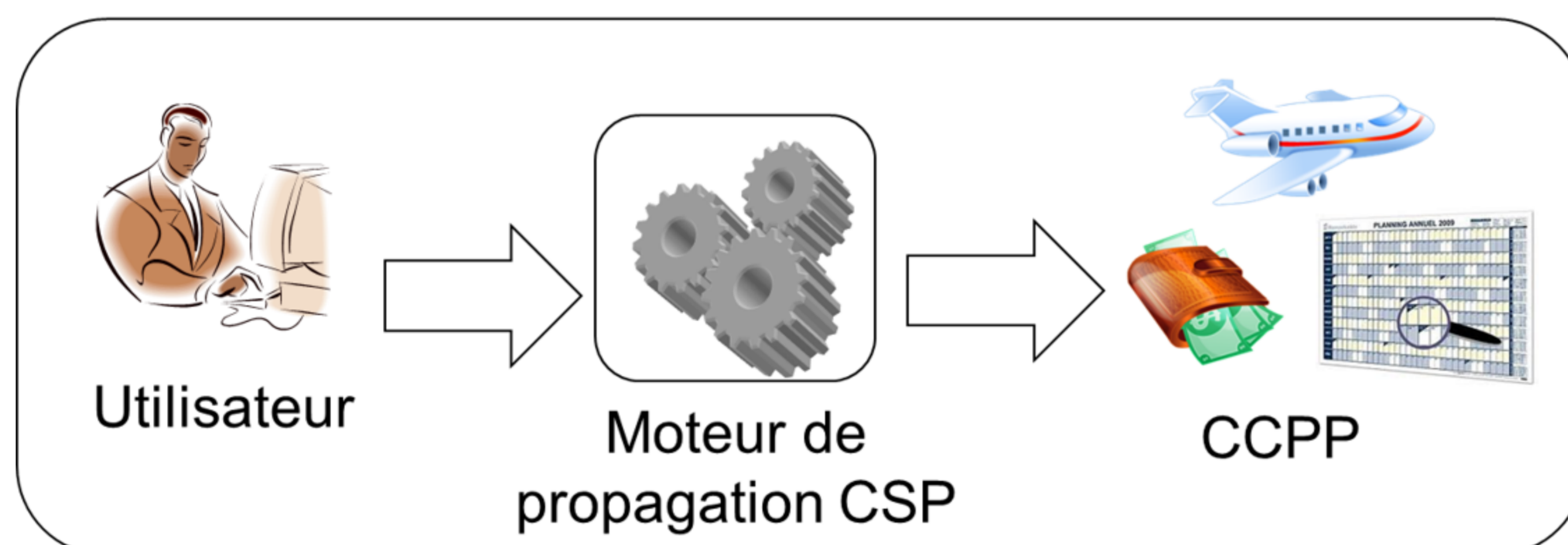
Paul Pitiot
Luis Garcés
Michel Aldanondo
Elise Vareilles
Paul Gaborit

Problématique de modélisation

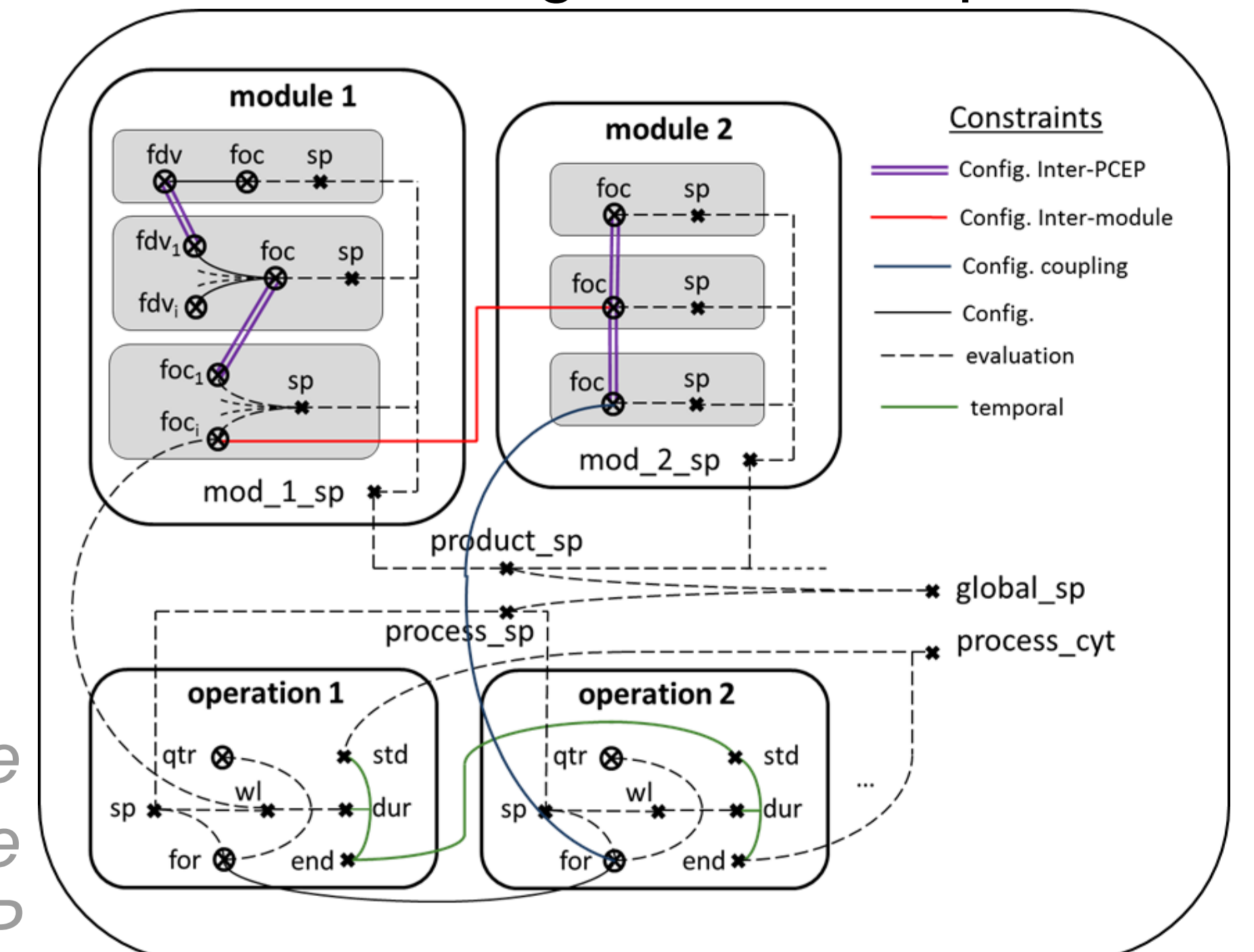
- **Configuration Conjointe Produit / Process** (Concurrent Product Process Configuration CPPC)
- Besoin de considérer **simultanément** les deux problèmes de configuration Produit et Process pour réduire les **incohérences**
- Besoin d'**outil d'aide à la décision** utilisable industriellement
- Besoin d'**optimisation** lors de la recherche de solutions
- Problème **complexe** : Taille extrêmement importante (10^{18} -> 10^{80} solutions), nombreuses **contraintes** à prendre en compte, plusieurs critères considérés (coût, délais, performance, impact environnemental, etc.)

Solution

- Modèle conjoint basé sur le formalisme CSP (Constraint Satisfaction Problem)
- Outil interactif de configuration: Cofiaide
- Proposition d'un modèle générique du CPPC
- Réalisation d'un benchmark (ensemble de modèles) pour évaluer des algorithmes d'optimisation



Outil proposé Cofiaide

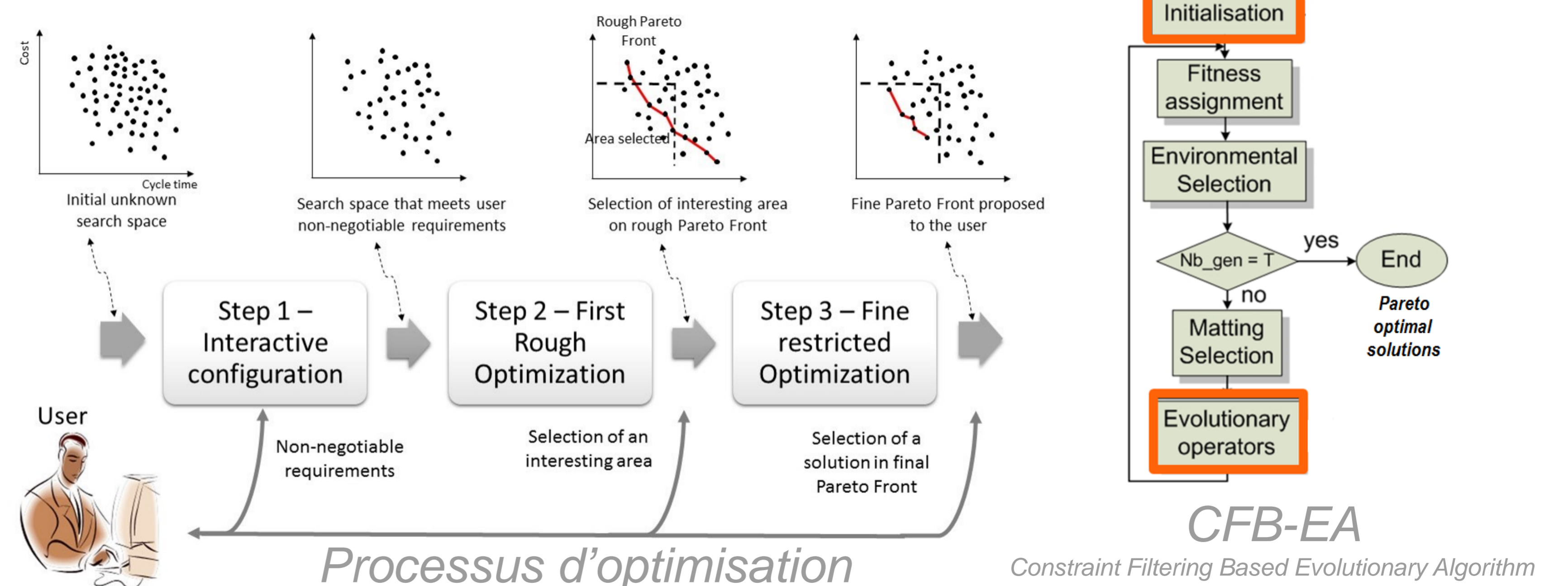


Problématique d'optimisation

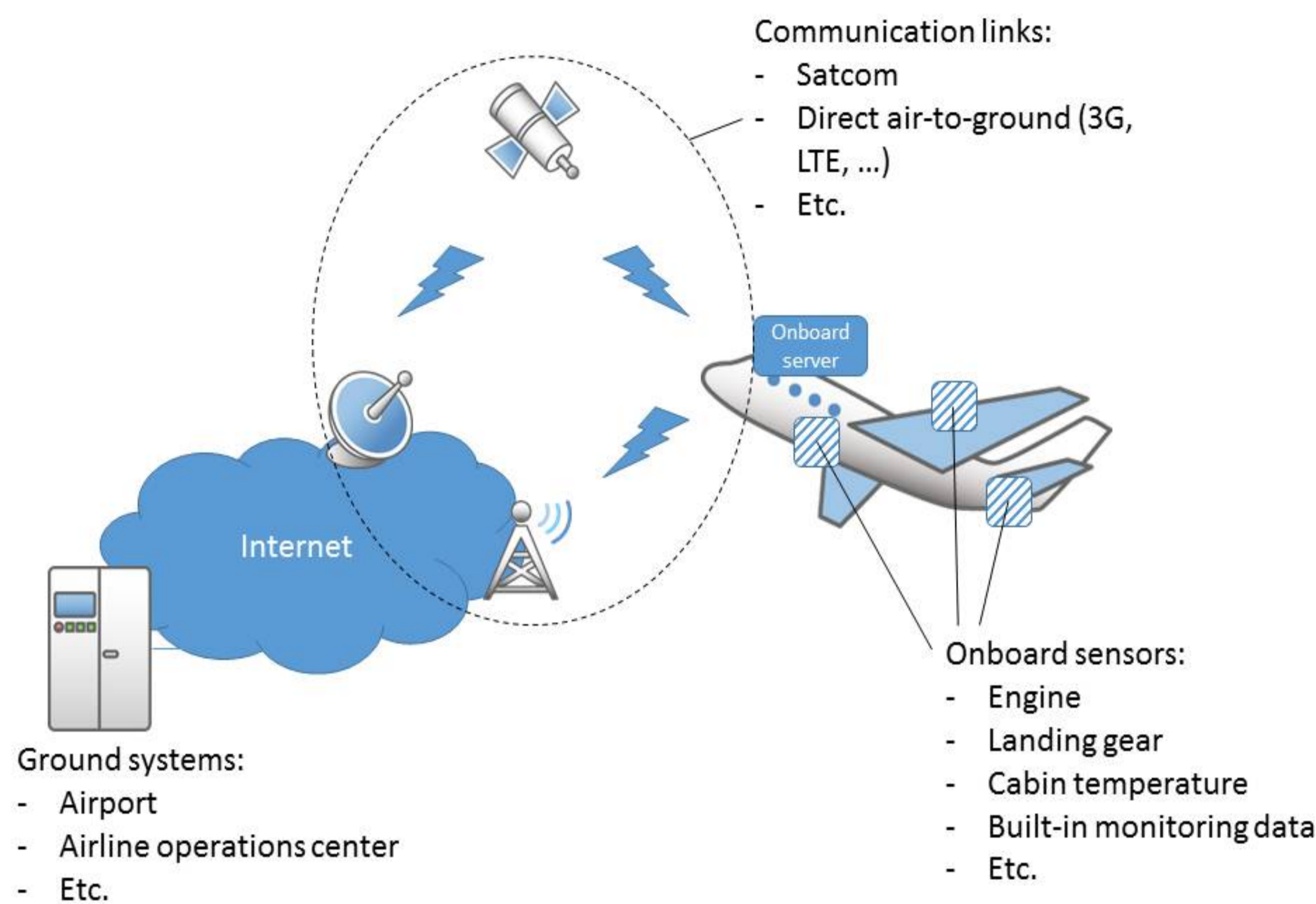
- **Optimisation combinatoire multi-objectif sous contraintes**
- Recherche de solutions optimisées dans des **espaces de recherche très importants**
- Temps de calcul compatible avec un processus d'**aide à la décision** industriel

Solution

- Optimisation par métaheuristique : **algorithmes évolutionnaires**
- Prise en compte des **contraintes** du problème : intégration avec Cofiaide
- **Interaction** avec l'utilisateur pour accélérer la recherche de solutions
- **Evaluation** des méthodes / opérateurs proposés



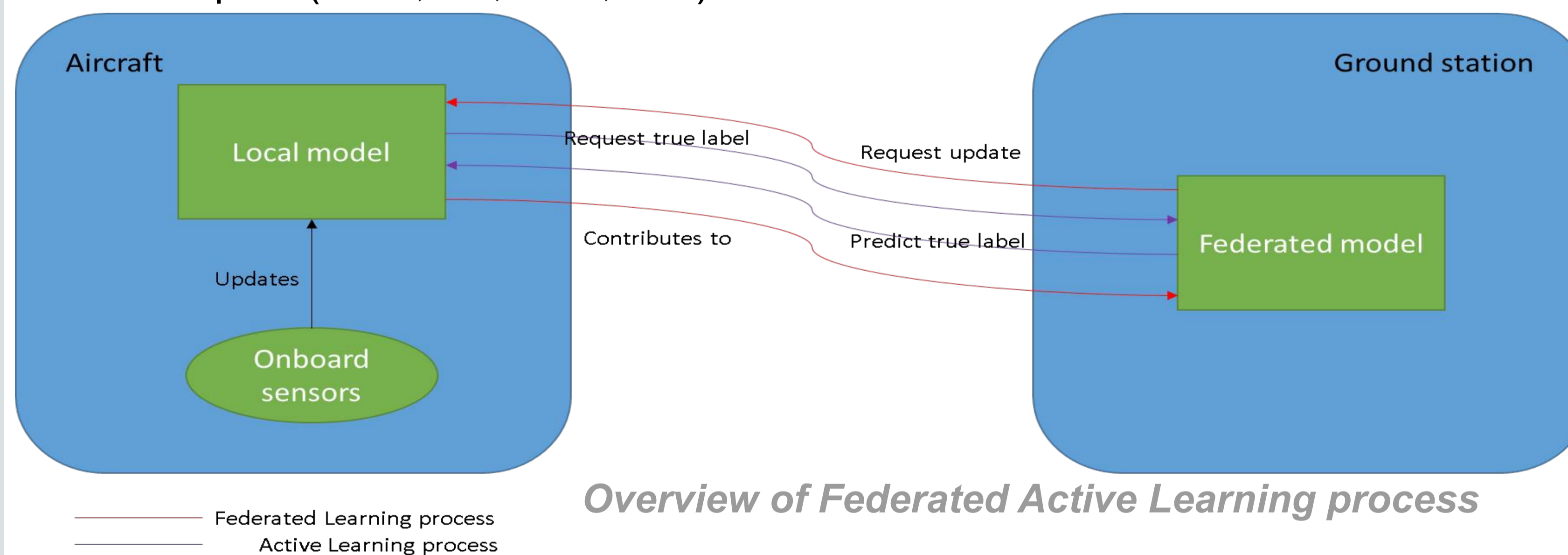
Predictive maintenance in aeronautics: Real-time anomaly detection with in-flight data



Overview of aeronautical components

Specific challenges

- ▶ **Rare events:** imbalance ratio closer to 1:100,000; sampling techniques (s.a. oversampling, undersampling, SMOTE) do not perform well; resilient learning techniques (RF, GBT) work best [1].
- ▶ **Ambiguous data format:** raw input is full text; design **NLP pipeline** to convert to numerical values with tokenizer, stemmer, semantic approach (synonyms, stopwords), vectorizer (bag-of-words, n-gram) and topic classifier (hierarchical k-means, LDA); easy to implement and improve F1 over benchmark from 0.96 to **0.992** [2].
- ▶ **Complex dependencies:** both spatial and temporal dependencies; adapt **spatio-temporal CNN** video classification solution for failure prediction. Baseline performance on problem intractable with standard learning techniques (SVM, RF, GBT, etc.).



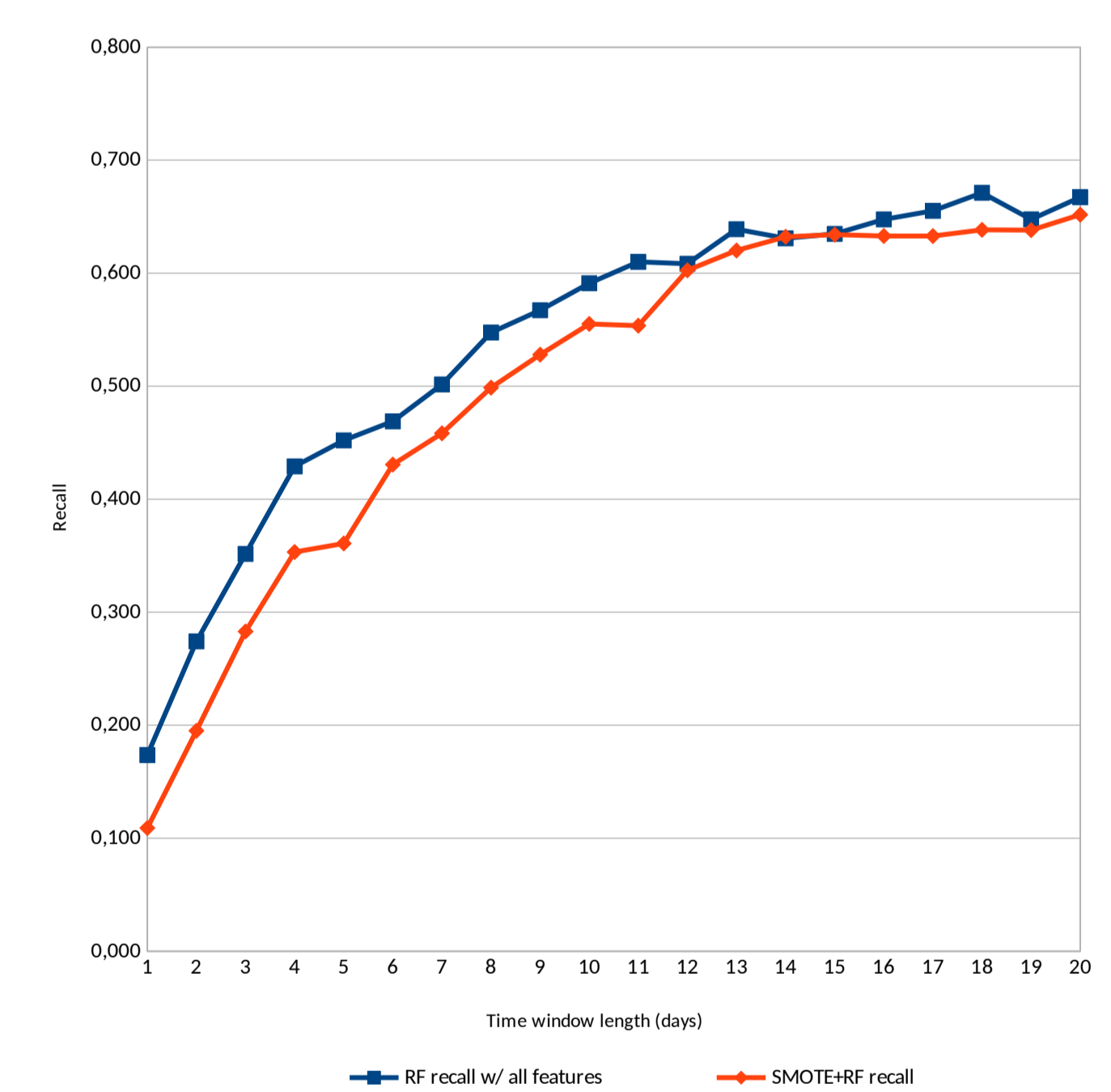
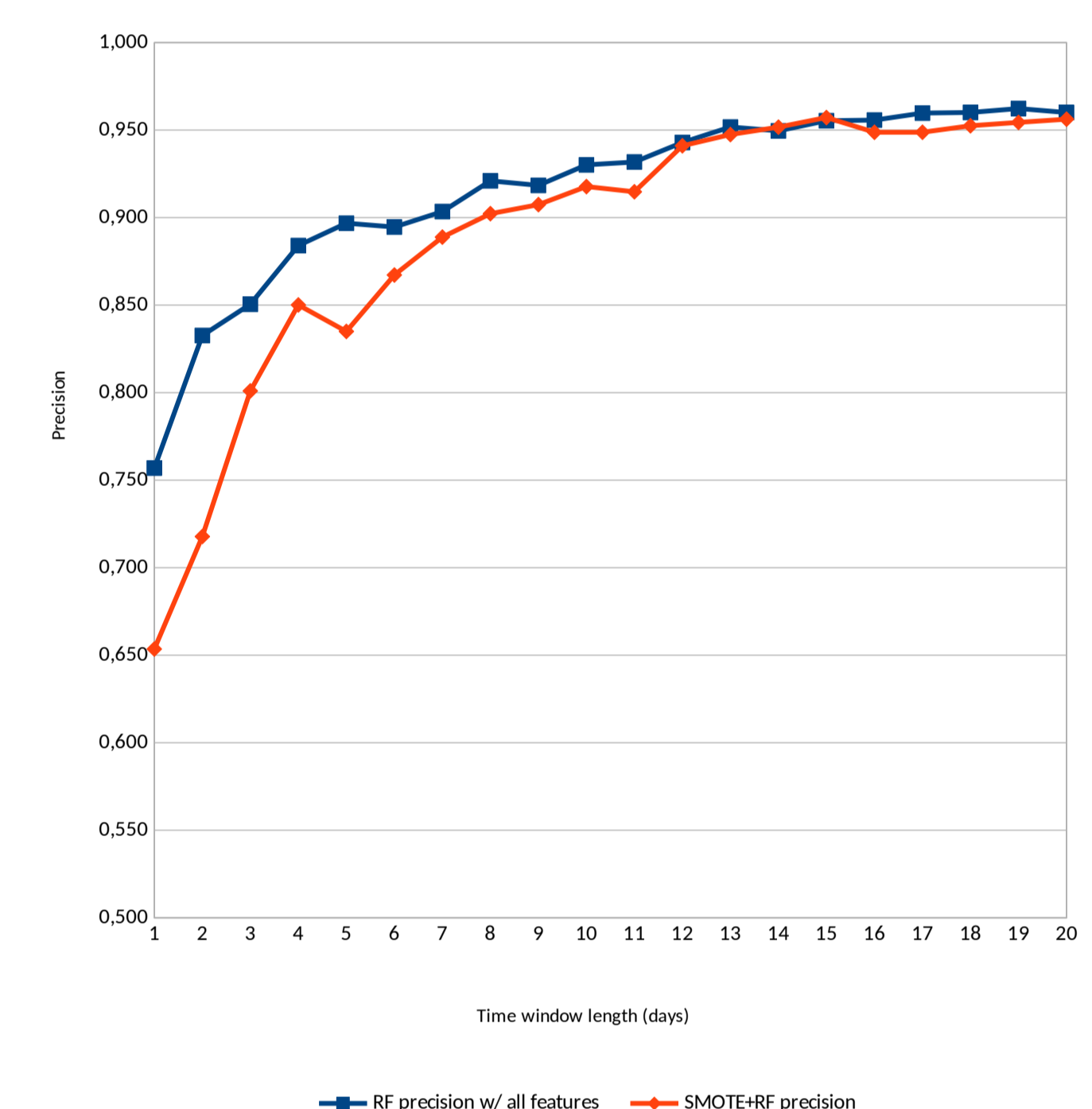
Overview of Federated Active Learning process

General solution framework

- ▶ Bring together **Federated Learning** and **Active Learning**.
- ▶ **Federated Learning:** learns a model in central server while dataset is distributed among many hosts. Offers fine control on communication budget but cannot guarantee the performance of learned model at individual host level.
- ▶ **Active Learning:** learn a model in a semi-supervised fashion with the true labels being accessible on request to a third party at some cost. Offers a way to control the model uncertainty under a budget but is not easily distributable.
- ▶ **Proposed solution: Federated Active Learning**
- ▶ Combines both paradigms with a central aggregate model that leverages data from multiple flights **without transferring raw data** and individual on-board models that can self-evaluate and make requests to the central model when needed to **guarantee a minimum level of performance**.

Context and problematic

- ▶ **Unplanned maintenance** implies very high expenses. **Online predictive maintenance** promising to reduce costs.
- ▶ **On-board computational resources** limited in storage and processing power. But **collected data** data are plentiful.
- ▶ **Ground computational resources** with high capacity storage and high processing power. But access to **historical data** not always relevant (different systems, different sensors, ...).
- ▶ **Modern communication techniques** allow for nearly always-up data link to the ground but with limited bandwidth and high cost.
- ▶ **Research question:** How to leverage available resources for the best prediction possible under a communication budget ?



Performance comparison of sampling techniques

[1] N. Aussel, Y. Pététin, S. Chabridon, Improving Performances of Log Mining for Anomaly Prediction Through NLP-Based Log Parsing. MASCOTS 2018.

[2] N. Aussel, S. Jaulin, G. Gandon, Y. Petetin, E. Fazli, S. Chabridon, Predictive Models of Hard Drive Failures Based on Operational Data. ICMLA 2017.

Software Measurement Management aided by Machine Learning

INDUSTRIAL CONTEXT

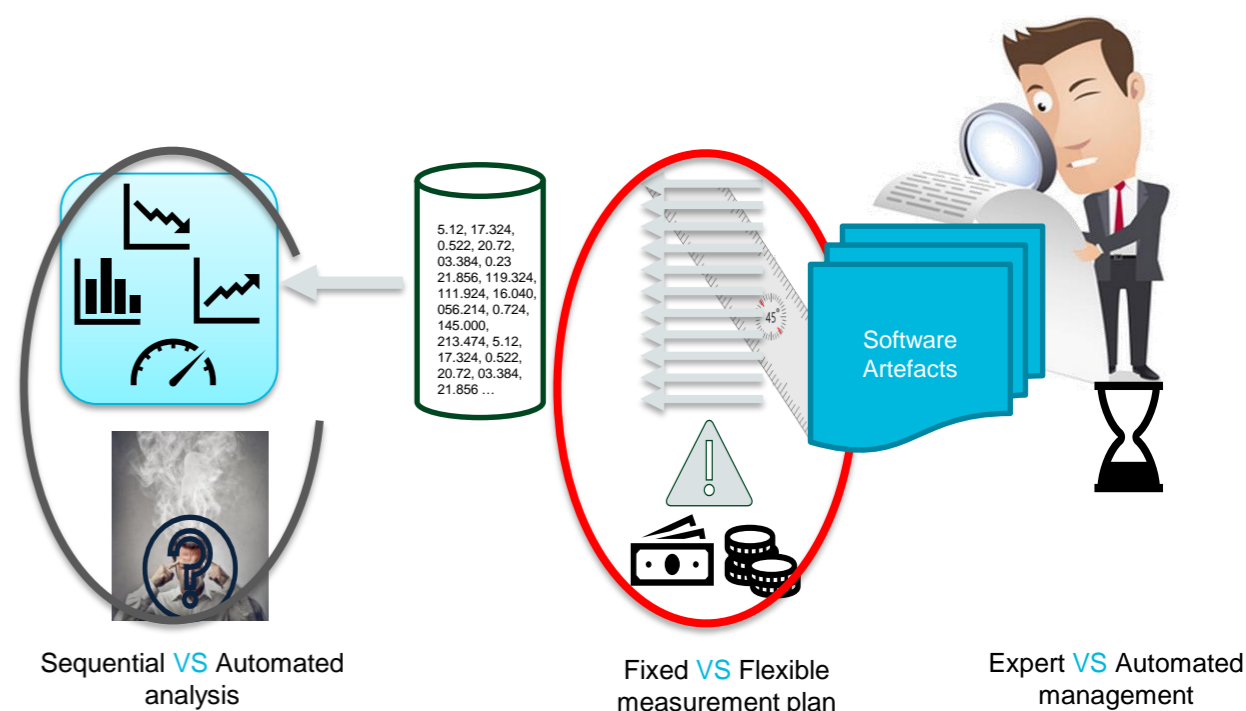
MEASURE project

ITEA3 MEASURE project :

- Improvement of software measurement process
- Increase the measurement process quality and efficiency

Heavier and costly measurement management load:

- Expert dependency for measurement management
- Costly measurement process
- Big amounts of data to analyze
- Reduce the software measurement management load
- Reduce the measurement process costs
- To aid to the analysis



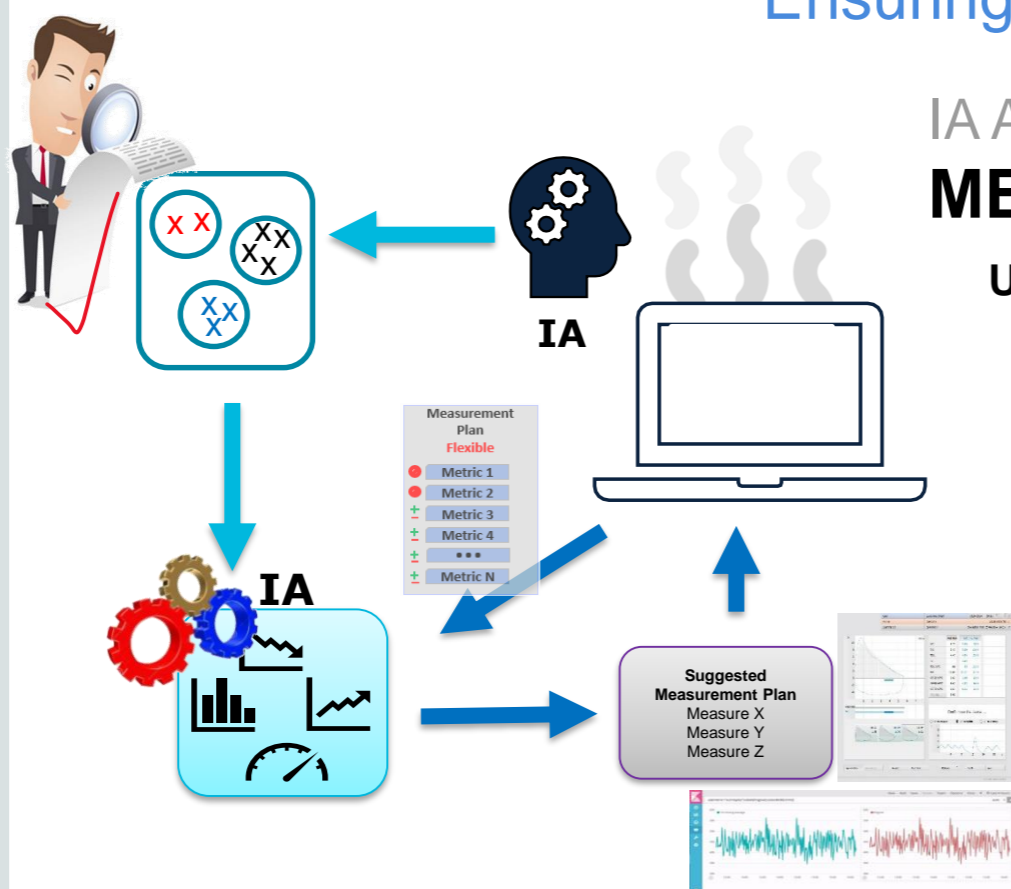
Parties prenantes



Auteurs

Sarah DAHAB
Stéphane MAAG

Ensuring relevant evaluation continuously at runtime



IA ADVANTAGES

METRICS SUGGESTER TOOL

Unsupervised-based analysis model elaboration

- Automated learning of software historic by using X-Means algorithm
- In order to generate an analysis model for the analysis module
- Expert intervention only to validate the generated model
- Reduce the expert load and management time cost

Automated Measurement Analysis and Suggestion

- Based on the generated analysis model : Automated measurement analysis at runtime by using SVM algorithm
- Lightest and "smart" analysis
- Suggestion at runtime of new measurement plan according to the analysis result
- Optimize the measurement process to the software needs
- Reduce the measurement cost and the analysis harshness

Partenaires



METRICS SUGGESTER TOOL ARCHITECTURE

USE CASE RESULTS

The Metrics SuggesterTool is an independent web application based on the Django framework. The measurements analysis and suggestion through the SVM and RFE algorithms are implemented using the Scikit-learn library.

The whole process consists of the following steps

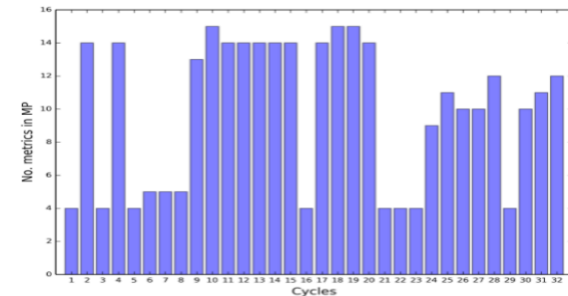
- Elaboration of the training file by providing the historical measurement data to generate data model
- Configuration of the initial classifier by using the previous data model to train the initial classifier.
- Configuration of the suggestion by providing the correlation data model with the initial configuration.
- The result of the classification process is then used to suggest changes in the measurement plan.

Experimentation on varied number of vectors.

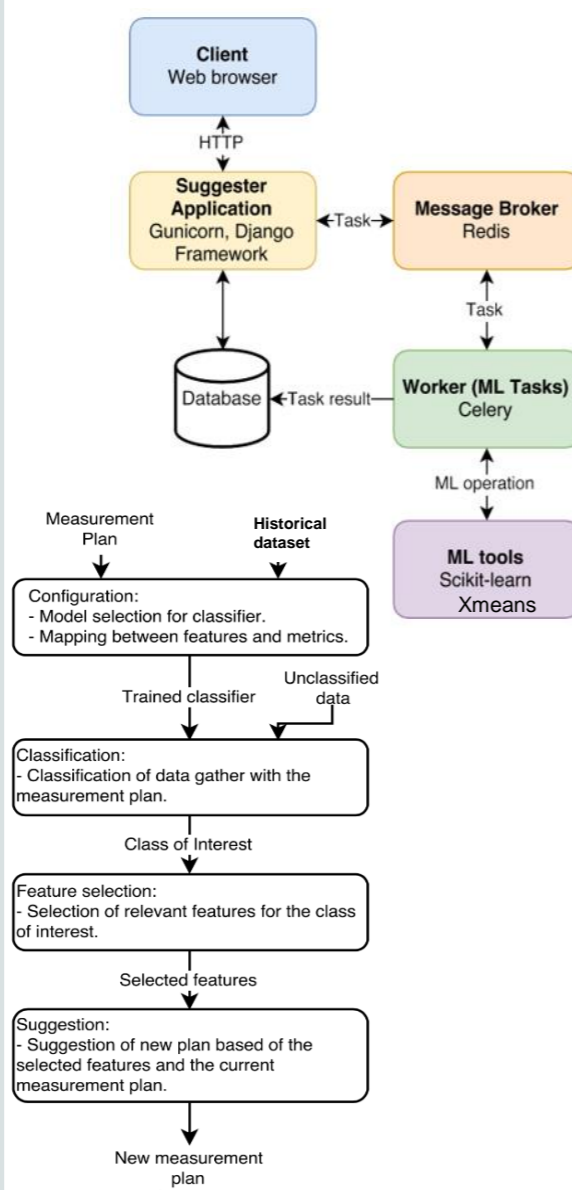
File	Data	Distribution	X-MEANS	Time (s)
1	50	3 groups	2 clusters	0,03
2	100	6 groups	6 clusters	0,05
3	1000	6 groups	6 clusters	0,08
4	10000	6 groups	6 clusters	0,15

Unsupervised clustering results

Experimentation on 16 million vectors divided on 32 subset of 500000.



Number of metrics of the suggested plan in each cycle



[1] Sarah A. Dahab, Juan Jose Hernandez Porras, Stephane Maag. A Novel Formal Approach to Automatically Suggest Metrics in Software Measurement Plans. ENASE 2018: 283-290
[2] Sarah A. Dahab, Stephane Maag. Suggesting Software Measurement Plans with unsupervised learning data analysis. ENASE 2019: in proceeding

[1] Ramesh, M. R., & Reddy, C. S. (2016). Difficulties in Software Cost Estimation: A Survey. International Journal of Scientific Engineering and Technology, Volume, (5), 10-13.
[2] Kitchenham, B. (2010). What's up with software metrics?—A preliminary mapping study. Journal of systems and software, 83(1), 37-51.

Adaptive Collaborative Topic Modeling for Online Recommendation

Marie Al-Ghossein¹, Pierre-Alexandre Murena^{1,2}, Talel Abdesslem¹, Anthony Barré³, Antoine Cornuéjols²

¹LTCI Télécom ParisTech (France); ²UMR MIA518 - AgroParisTech INRA (France); ³AccorHotels (France)



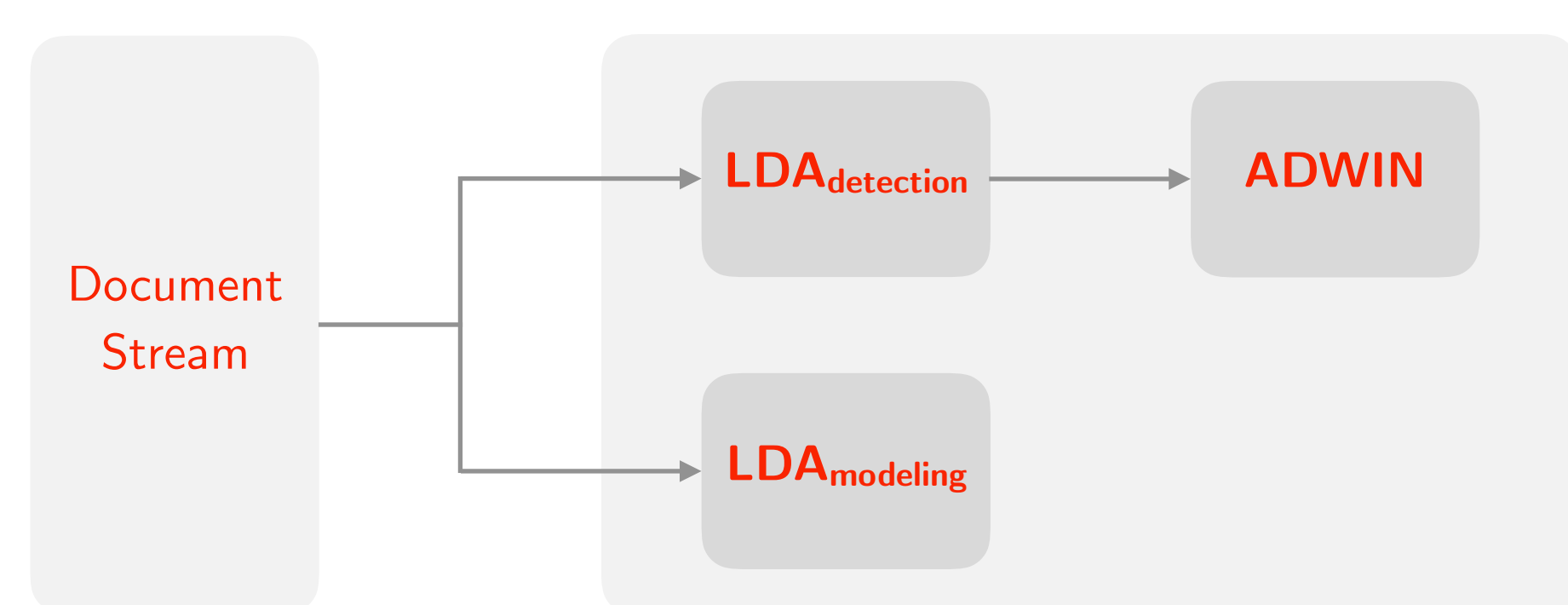
Abstract

Collaborative Filtering (CF) mainly suffers from rating sparsity and from the cold-start problem. Auxiliary information like texts and images has been leveraged to alleviate these problems, resulting in hybrid Recommender Systems (RS). Due to the abundance of data continuously generated in real-world applications, it has become essential to design online RS that are able to handle user feedback and the availability of new items in real-time. These systems are also required to adapt to drifts when a change in the data distribution is detected.

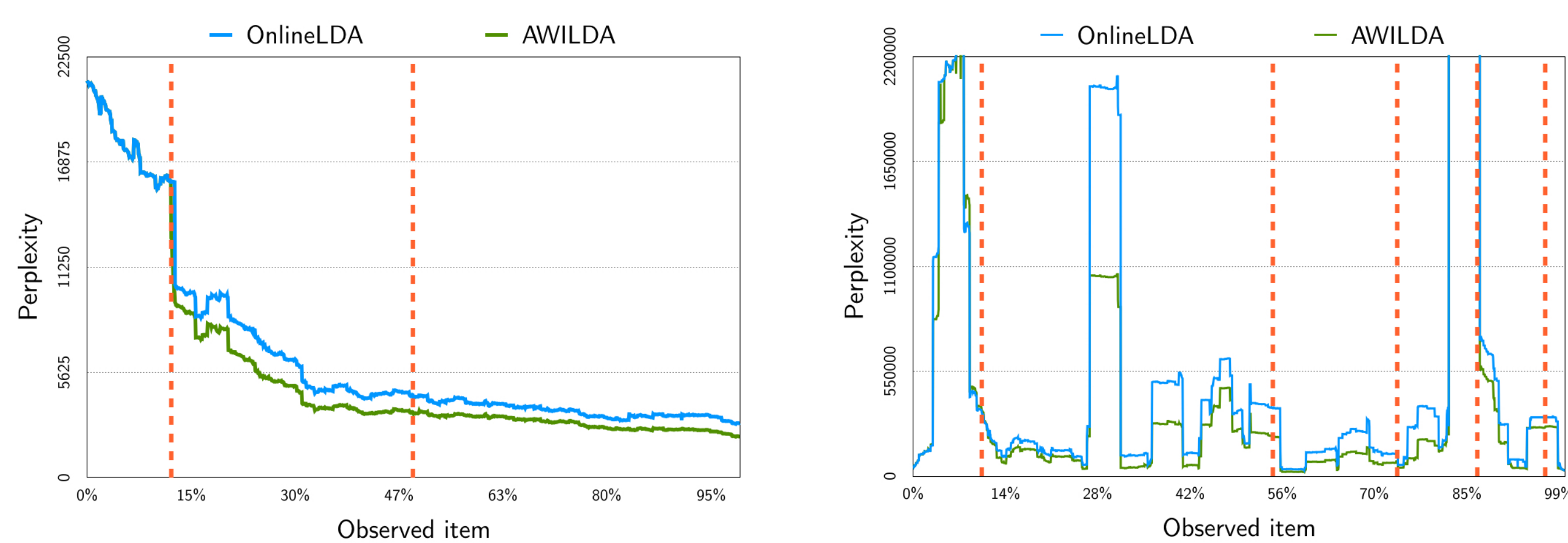
In this paper, we propose an adaptive collaborative topic modeling approach, CoAWILDA, as a hybrid system relying on adaptive online Latent Dirichlet Allocation (AWILDA) to model new available items arriving as a document stream and incremental matrix factorization for CF. The topic model is maintained up-to-date in an online fashion and is retrained in batch when a drift is detected, using documents automatically selected by an adaptive windowing technique. Our experiments on real-world datasets prove the effectiveness of our approach for online recommendation.

AWILDA

AWILDA is an algorithm designed to handle drift detection in streams of documents. Topics of a document are modeled using online LDA model and the drifts are detected by ADWIN algorithm using the likelihood.



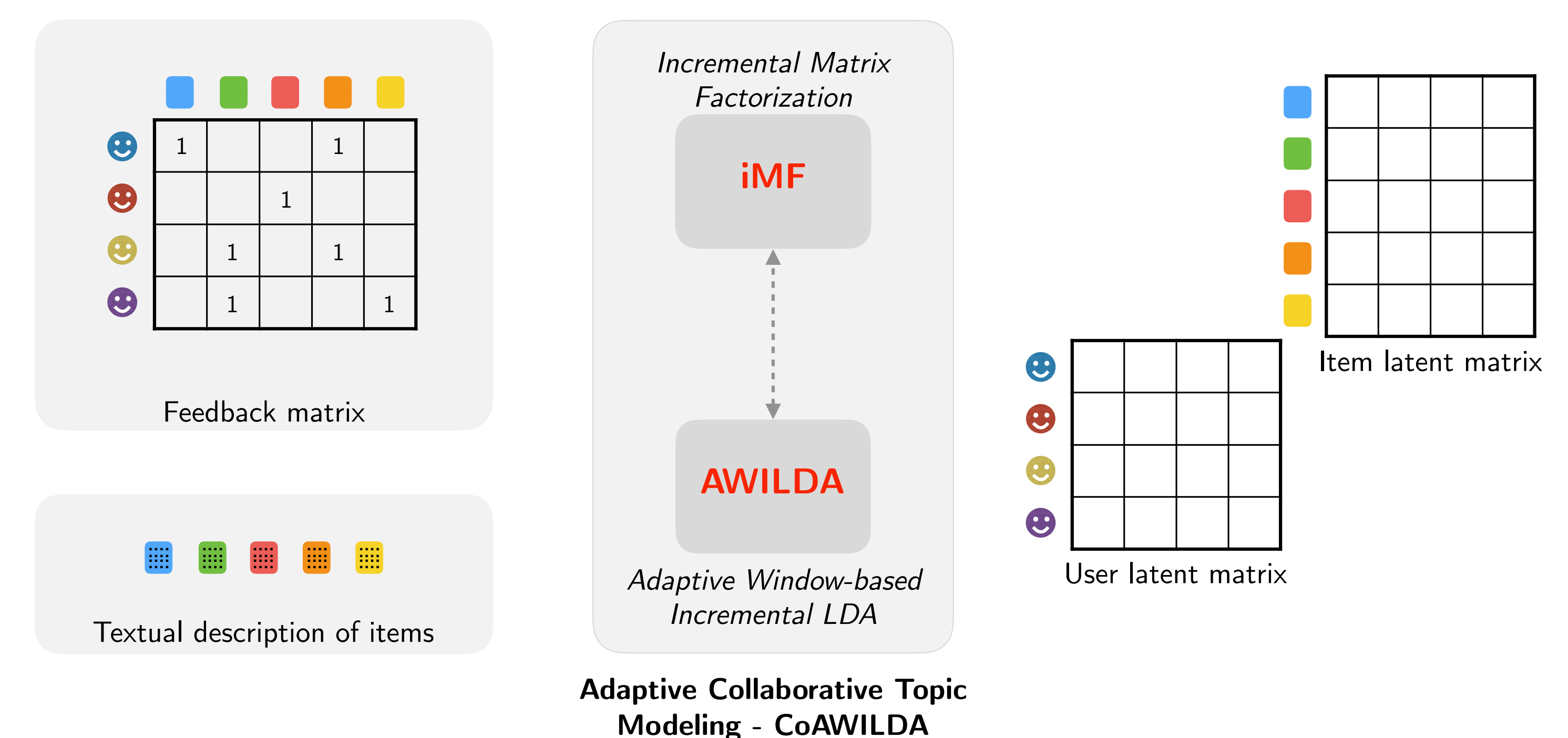
Experimental results: Perplexity is evaluated for classical online LDA and our adaptive approach, AWILDA. The following figures show the evolution of perplexity for the two datasets, where a lower perplexity indicates a better capacity of generalization. The dashed vertical lines correspond to detected drifts.



CoAWILDA for Adaptive Collaborative Topic Modeling

Context: We consider the problem of online recommendation of textual items (for instance news articles) using a **hybrid approach**. This task induces several major issues which makes the use of classical approaches insufficient: handling observations in real-time, continuously learning from data streams and updating models, addressing the item cold-start problem as soon as items are available...

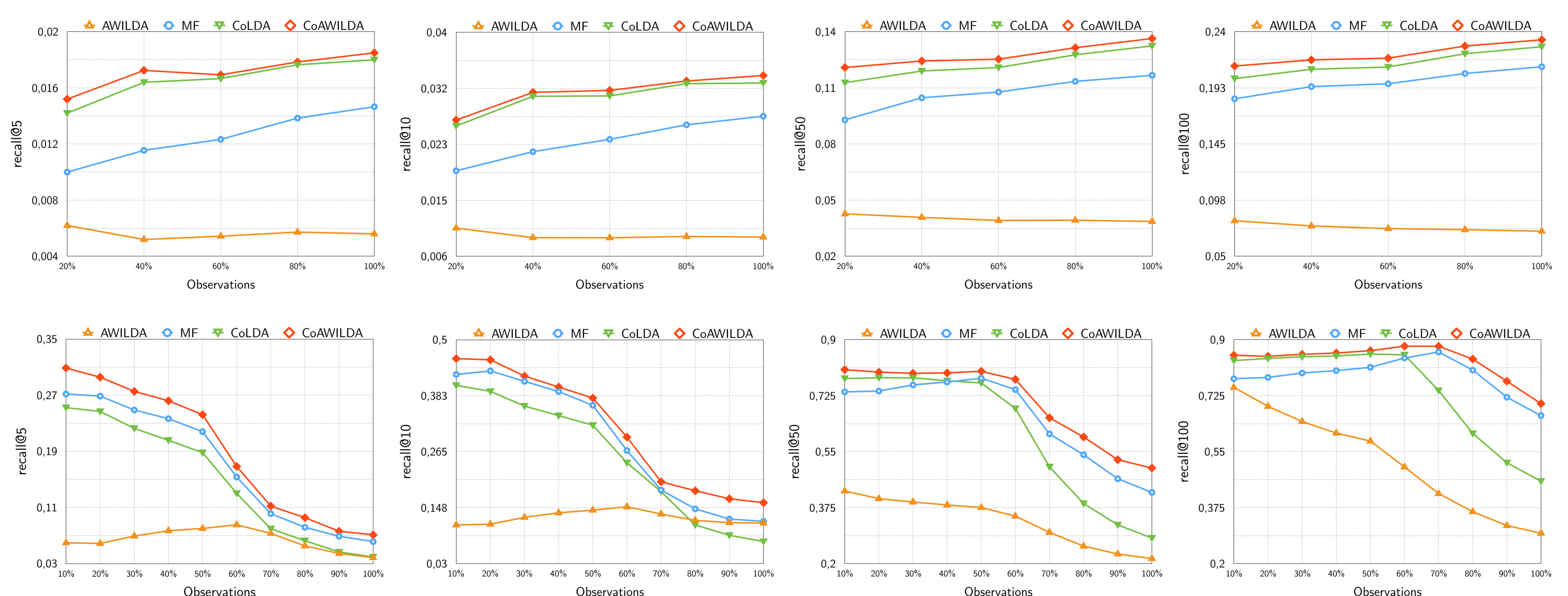
CoAWILDA is an incremental hybrid recommendation method which combines **incremental matrix factorization** for online collaborative filtering and AWILDA for online item representation. AWILDA combines **online LDA** with **drift detection**, in order to model changes in the topic distributions and compositions over time.



CoAWILDA: Experimental Results

The **MovieLens-100k** dataset gathers 100,000 ratings from 1,000 users on 1,700 movies, spanning over 18 months. Movies become available according to their reported release date and are represented by their English abstract extracted from DBpedia.

The **plista** dataset captures interactions collected during the month of February 2016 on several German news portals. It gathers 32,706,307 interactions from 1,362,097 users and 8,318 news articles, where the date of an article corresponds to its publication date.



The figures show the recall@N over time for CoAWILDA and the other compared methods using prequential evaluation (*test-then-learn*) and highlight the importance of the drift detection module.

AWILDA: Content-based method based on adaptive online LDA model (AWILDA); **MF:** Incremental Matrix Factorization for collaborative filtering only; **CoLDA:** Hybrid method using a fixed LDA model.

Conclusion

We propose a simple and generic method, combining the advantages of several state of the art methods. The experiments display the necessity of proposing drift adaptation for recommendation in online contexts. **Perspectives:** Including drift detection in the user models, in order to reflect changes in users' interests.

[1] João Vinagre, Alípio Mário Jorge, and João Gama. Fast incremental matrix factorization for recommendation with positive-only feedback. In UMAP, 2014.

[2] Albert Bifet and Ricard Gavaldà. Learning from time-changing data with adaptive windowing. In SDM, 2007

[3] Pierre-Alexandre Murena, Marie Al-Ghossein, Talel Abdesslem, and Antoine Cornuéjols. Adaptive window strategy for topic modeling in document streams. In IJCNN, 2018.

Parties prenantes



Auteurs

Jessica DACLEU
Jenny FAUCHEU
David DELAFOSSE,
Mihaela JUGANARU-
MATHIEU

Partenaires

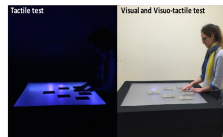


Table d'évaluation sensorielle
Nappomatic®

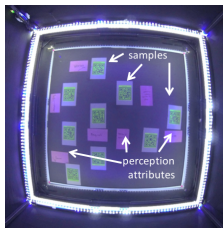
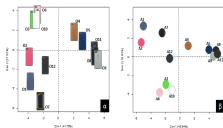
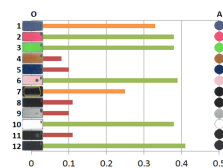


Image d'une évaluation



Clustering d'objets et d'échantillons
de matériaux



Distances dans l'espace textuel
entre la perception des objets et des
matériaux

Méthodologie, données collectées et traitements des spécialistes

Approche analyse sensorielle

- ▶ **Est-ce que présenter un objet fini (coque de téléphone) ou un échantillon de matière change la perception ? Si oui, comment ?**
- ▶ **Expériences** – selon un protocole précis :
 - présenter à 23 utilisateurs 12 échantillons ou 12 objets
 - les faire placer sur la Nappomatic®
 - demander à caractériser les groupes formés avec des mots libres.
- ▶ **Traitement des données avec des outils standards** (à l'exception des mots, donc basé uniquement sur les coordonnées x et y)
 - Clustering assez ressemblant entre les deux types : échantillons versus objets

Collaboration interdisciplinaire : analyse sensorielle et data science

Validation de l'expérience

- ▶ nombre d'utilisateurs suffisants et résultats fiables
- ▶ pas de différences entre les testeurs

Analyse du texte produit

- ▶ trouver une équivalence via une fonction de similarité entre les représentations numériques et textuelles
- ▶ les clusterings des objets et, respectivement, des échantillons sont très ressemblants

Exploration approfondie dans l'espace textuel

- ▶ Distances très différentes entre la perception objet et la perception échantillon
- ▶ Aucune explication basée sur les données disponibles

Avis de l'expert

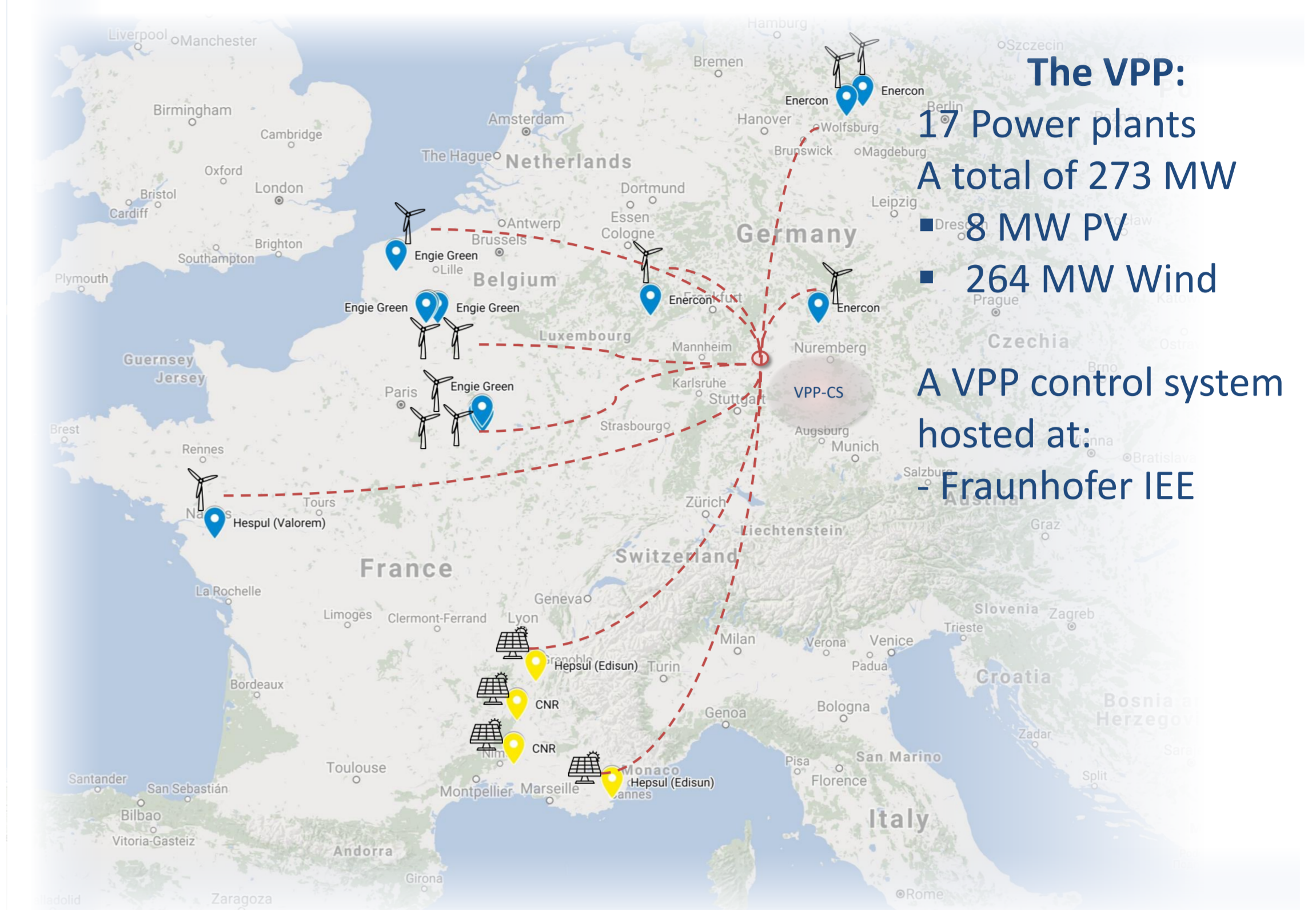
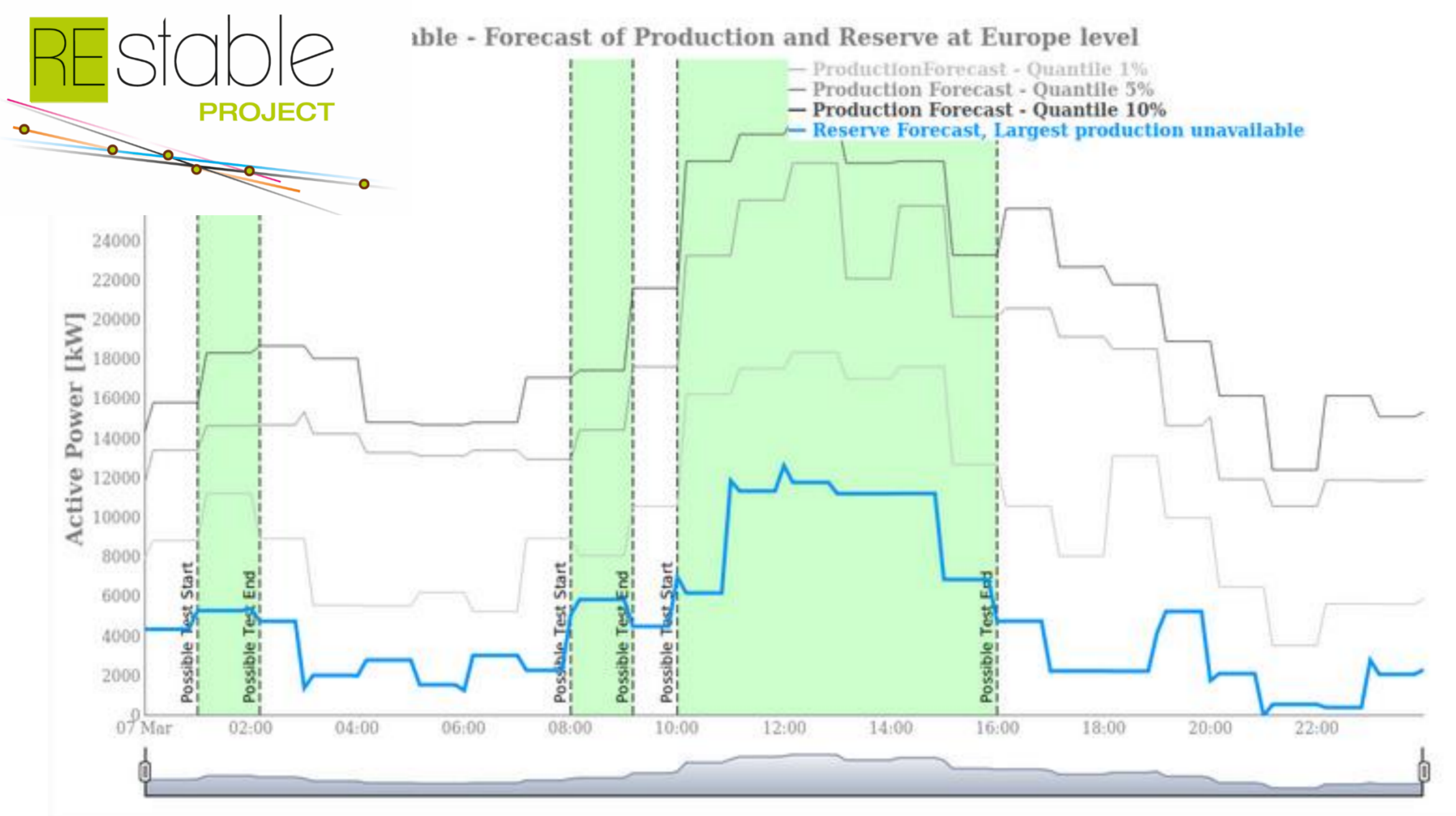
- ▶ distances grandes pour des matériaux de type imitation

Apport primordial

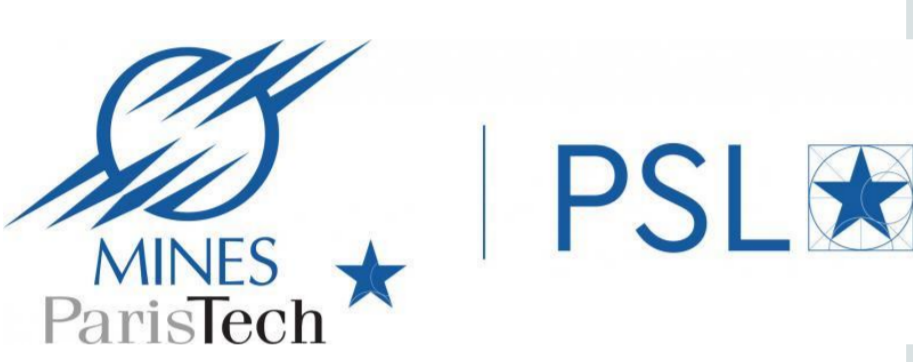
Conclusion de la collaboration

- ▶ Une approche IA ou Science de données en tout domaine technique doit s'appuyer sur des données et sur les compétences des experts du domaine

Forecasts of aggregated renewable production



Parties prenantes



Auteurs

Andrea MICHIORRI
Simon CAMAL
Romain DUPIN

Partenaires



External producers



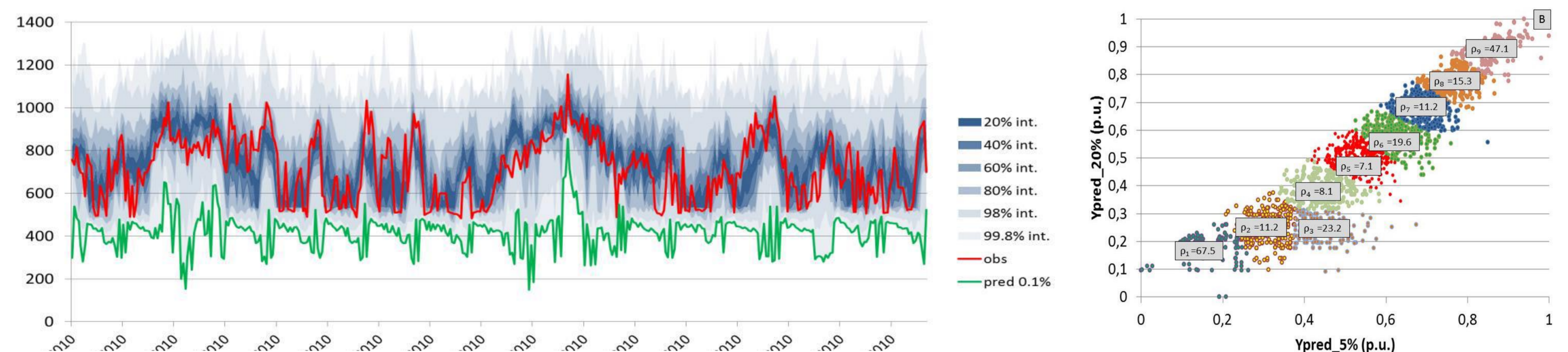
Financement



Challenges and objectives

- ▶ **CHALLENGE:** Renewables do not offer ancillary services due to intermittency and requested quality of service
- ▶ **OBJECTIVE:** To develop and test the technology necessary for allowing renewable resources to provide ancillary services to grid operators without the need of auxiliary generators, storage or other investments.
- ▶ **OBJECTIVE:** Focus on 4 fundamental ancillary services: 1) Frequency containment reserve (FCR), 2) Frequency restoration reserve (FRR), 3) Replacement reserve (RR), 4) Voltage support (VS)
- ▶ **OBJECTIVE:** Attention is paid to the integration into current market schemes with: a) Bidding strategies, b) Intra-day, day-ahead and weekly forecasts of conventional power plants.
- ▶ **FORECASTS:** Forecasts of production and market conditions are used to determine the amount of services that the VPP can provide.
- ▶ **FORECASTS:** Three horizons (intra-day, day-ahead and week-ahead) and 4 aggregation levels are considered (power plant, BRP, TSO, VPP)
- ▶ **FORECASTS:** Attention is paid to forecast low value of production with a high probability of realisation and to the calculation of realistic scenarios: **Bottom 1% with a 0.1% resolution**

Forecasts of transmission line ampacity



Challenges and objectives

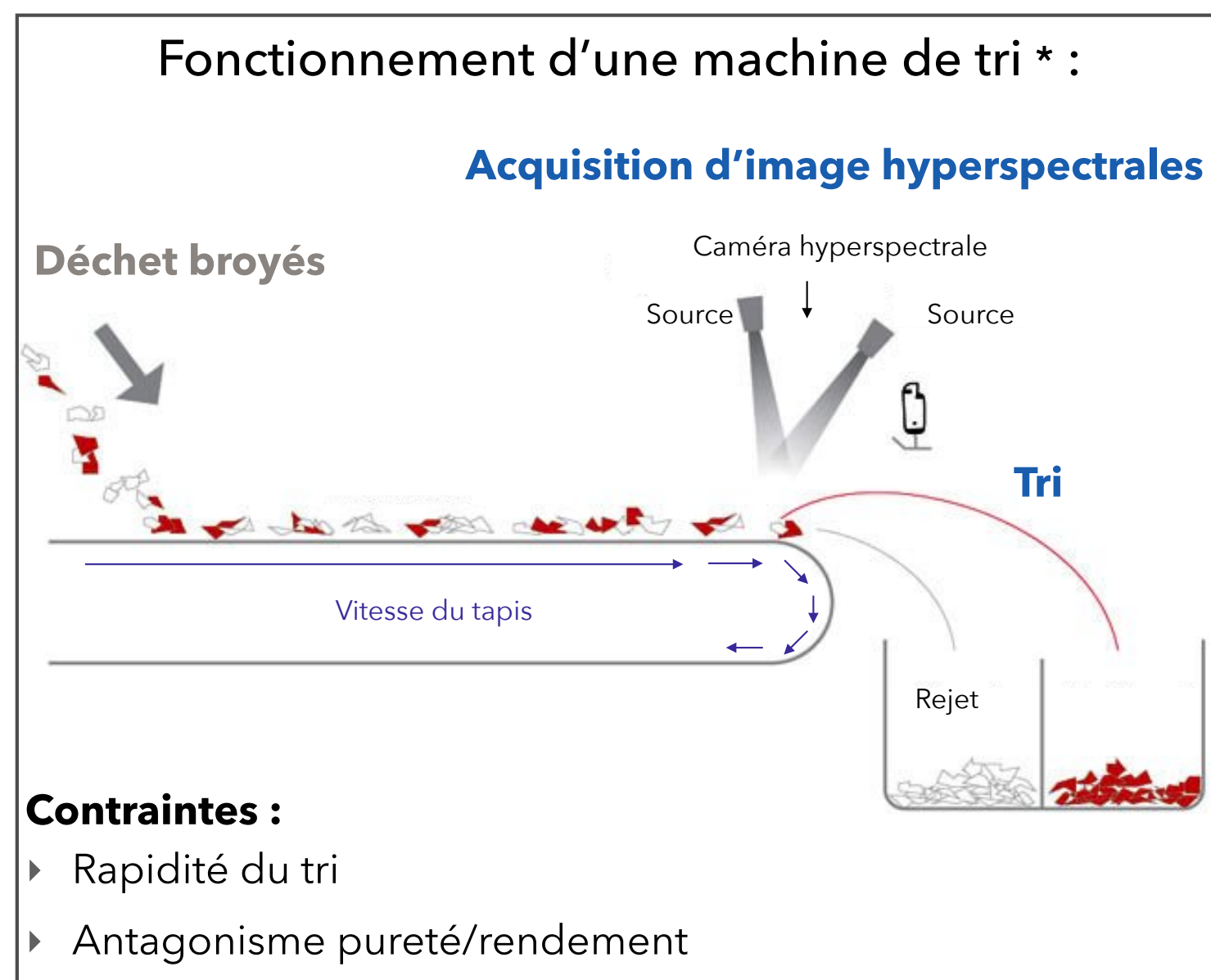
- ▶ **CHALLENGE:** overhead lines ampacity (maximum current) depends on weather conditions. The possibility of low winds, high temperatures and high solar irradiance determine its value
- ▶ **CHALLENGE:** Ampacity underestimation causes assets under-utilisation, renewable curtailment and recourse to peak generators. Overestimation cause conductor's elongation, with possible contacts and disastrous consequences
- ▶ **OBJECTIVE:** to accurately forecast in day ahead weather conditions over the whole line and assign the highest safe value for the line ampacity
- ▶ **OBJECTIVE:** To take into account the network operator's risk profile, increase the ampacity and reduce the risk.
- ▶ **FORECASTS:** Forecasts are realised by training random forests models for different time horizons. Inputs are: historical weather measurements, numerical weather predictions and ampacity estimations
- ▶ **FORECASTS:** Attention is paid to forecast low value of production with a high probability of realisation and to the calculation of realistic scenarios: **Bottom 1% with a 0.1% resolution**
- ▶ **FORECASTS:** this is done by replacing the tale of the forecasted distribution with a new forecast from a parametric model. Its parameters are optimised for clusters of observations characterised by their 5% and 20% quantiles values.

Contact : andrea.michiorri@mines-paristech.fr

Pourquoi trier les plastiques avant de les recycler ?

- ▶ Perte de propriétés mécaniques dès quelques pourcentages d'impuretés.
- ▶ Traitements différents nécessaires pour chaque type de plastiques à cause des incompatibilités.

Tri de fragment des plastiques à partir de données hyperspectrales dans un contexte industriel



Technologies spectroscopiques actuelles :

Proche Infra rouge
UV - Visible

Application aux plastiques non noirs



Verrou scientifique :
L'absorption du noir de carbone dans le proche infrarouge rend difficile la reconnaissance des plastiques noirs.

Technologie en développement :

Moyen Infra rouge

Application aux plastiques quelconques

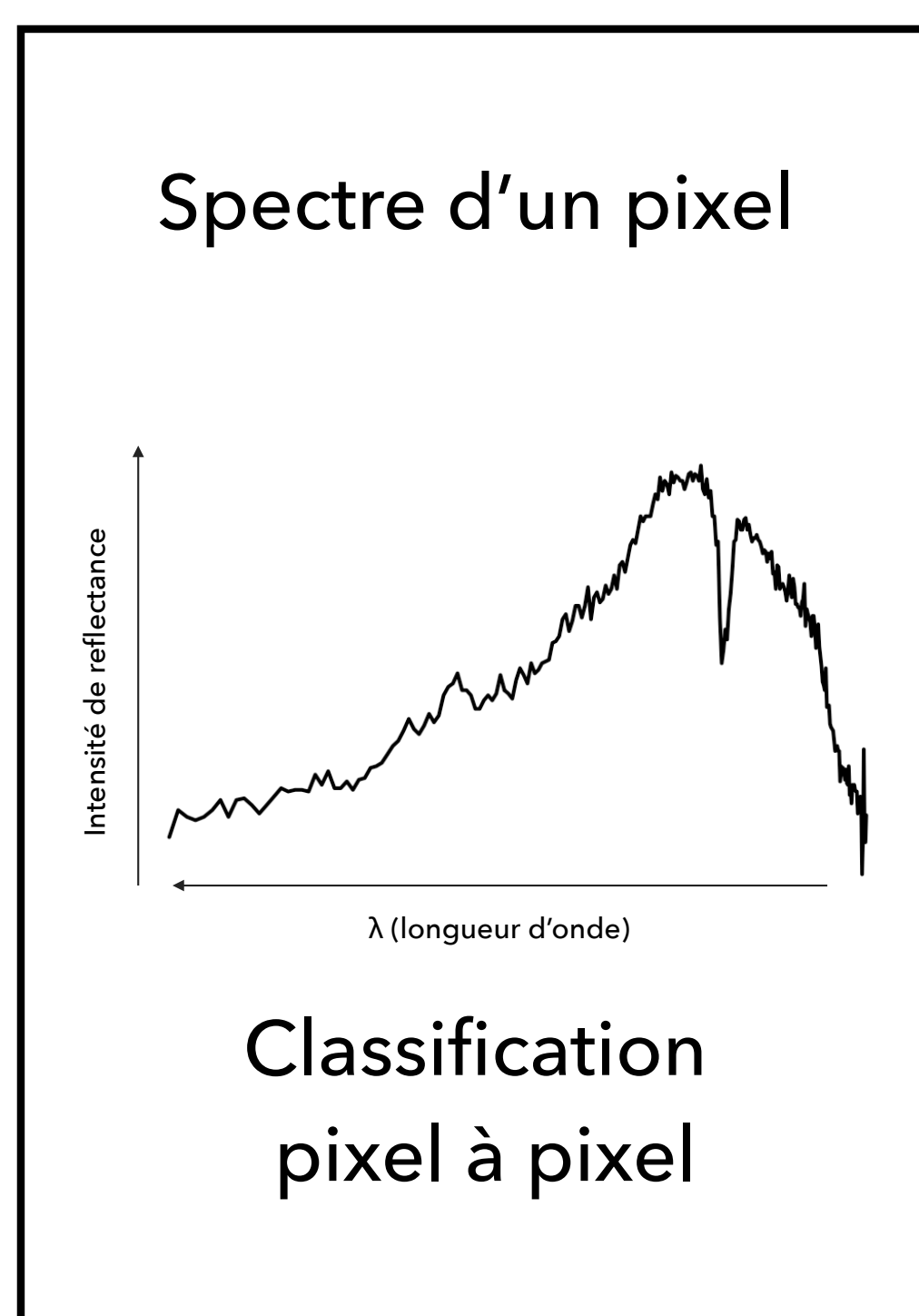
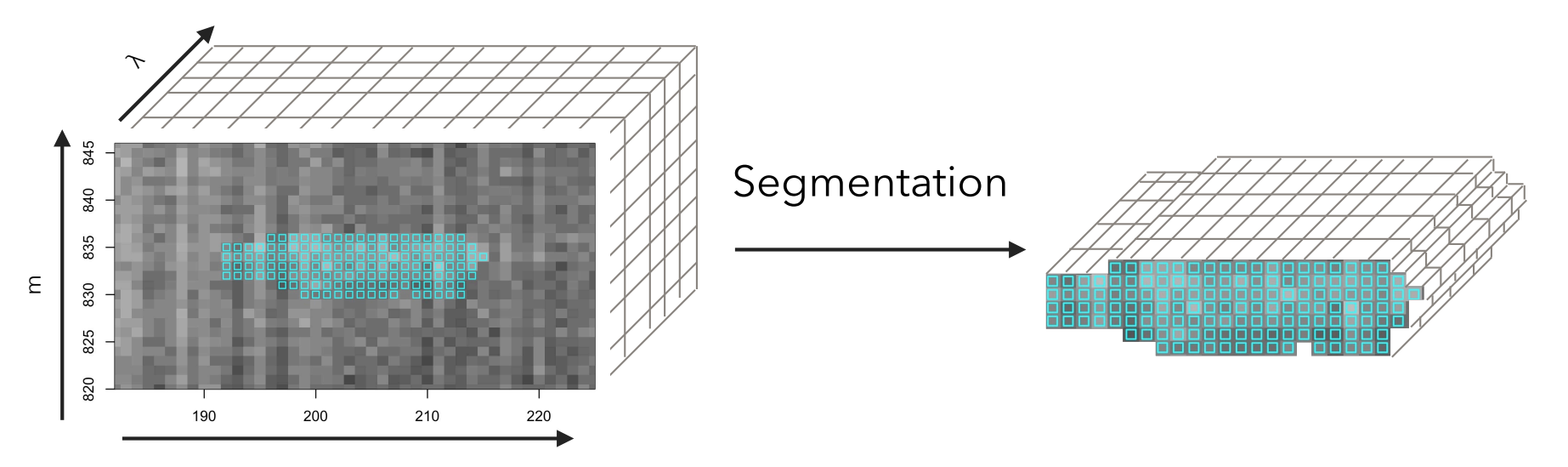
Problèmes de la technologie hyperspectrale moyen infra rouge appliquée aux plastiques

- ▶ Perturbations, présence de charges, impureté sur les objets, vieillissement des plastiques ...
- ▶ Absence de différences spectrales évidentes sur la plage de longueur d'onde disponible ...

Etapes du processus pour chaque objet sur le tapis

Segmentation :
Détection et isolation l'objet du tapis

Entrée : Image hyperspectrale acquise en temps réel (cube $m \times n \times \lambda$).
Classification binaire peu coûteuse « tapis » contre « objet ».
Sortie : l'ensemble de pixels connexes correspondants à l'objet.



Constat

Performances insuffisantes des classifieurs classiques, AFD, PLS-DA et SVM dans le cas des plastiques non vierge

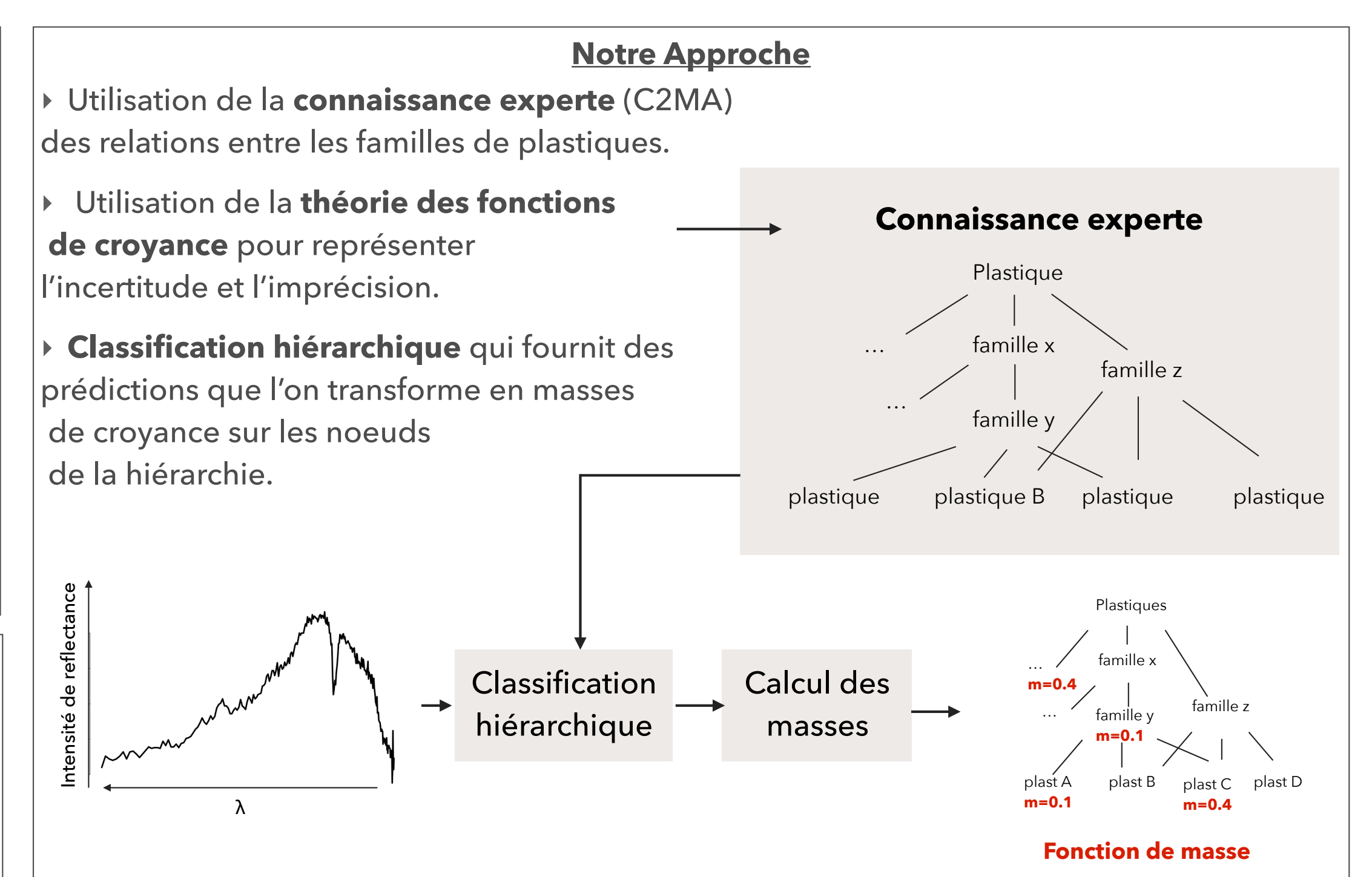
Classes réelles	Classes estimées				
	Plastique A	Plastique B	Plastique C	Plastique D	Plastique E
Plastique A	2340	0	0	0	0
Plastique B	0	2321	0	0	0
Plastique C	1	0	1172	0	3
Plastique D	0	0	0	2229	0
Plastique E	0	0	0	0	1274

Matrice de confusion AFD pour les plastiques vierges

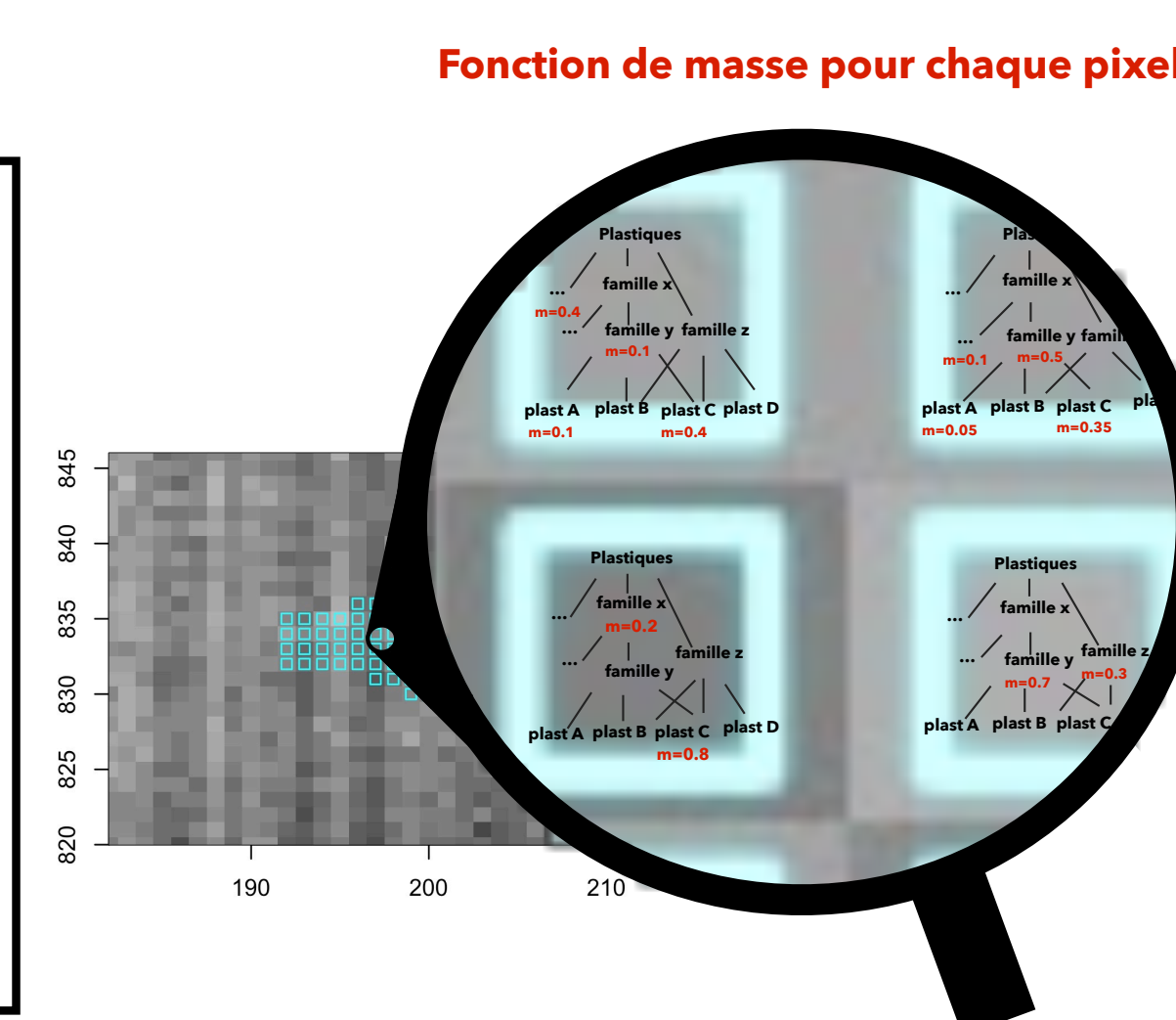
Classes réelles	Classes estimées			
	Plastique A	Plastique B	Plastique C	Plastique D
Plastique A	109	130	1	24
Plastique B	25	223	0	21
Plastique C	2	4	172	23
Plastique D	1	0	0	134

Analyse

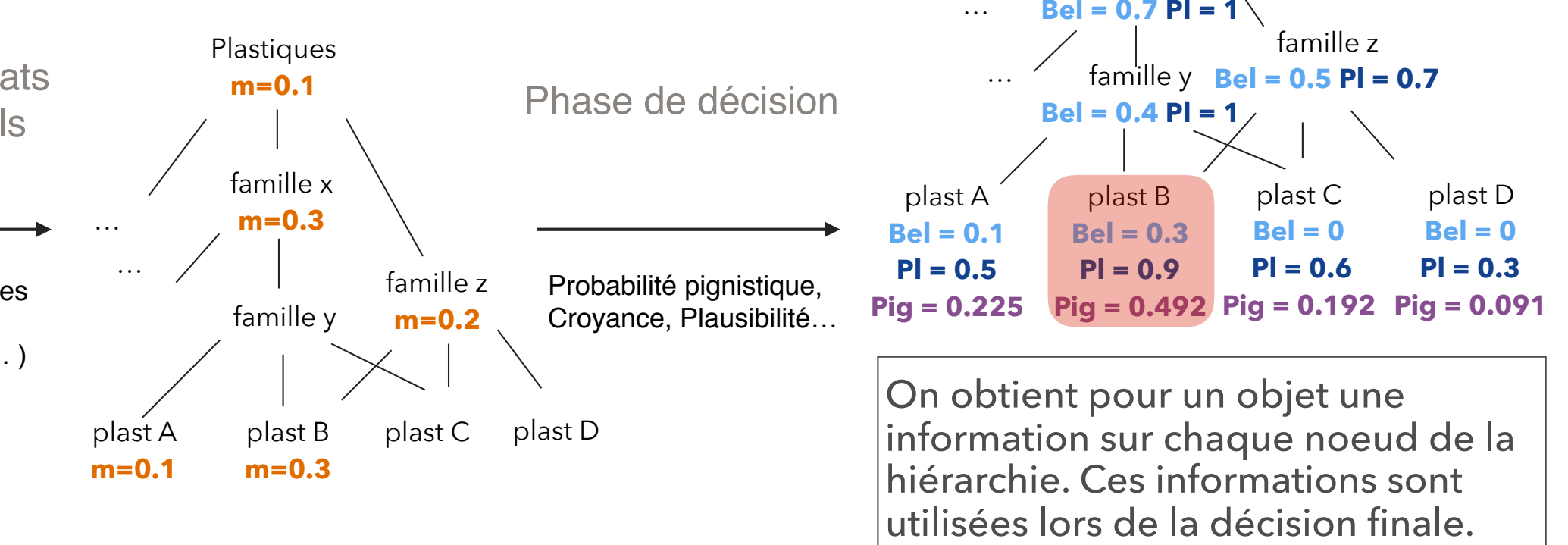
- ▶ Incertitude (imperfection de l'information)
- ▶ Imprécision (manque d'information)
- ▶ Alimentation de la base de données : couverture des classes à identifier (manque de données)



Objet : Ensemble de pixels
Fusion de toutes les informations des pixels



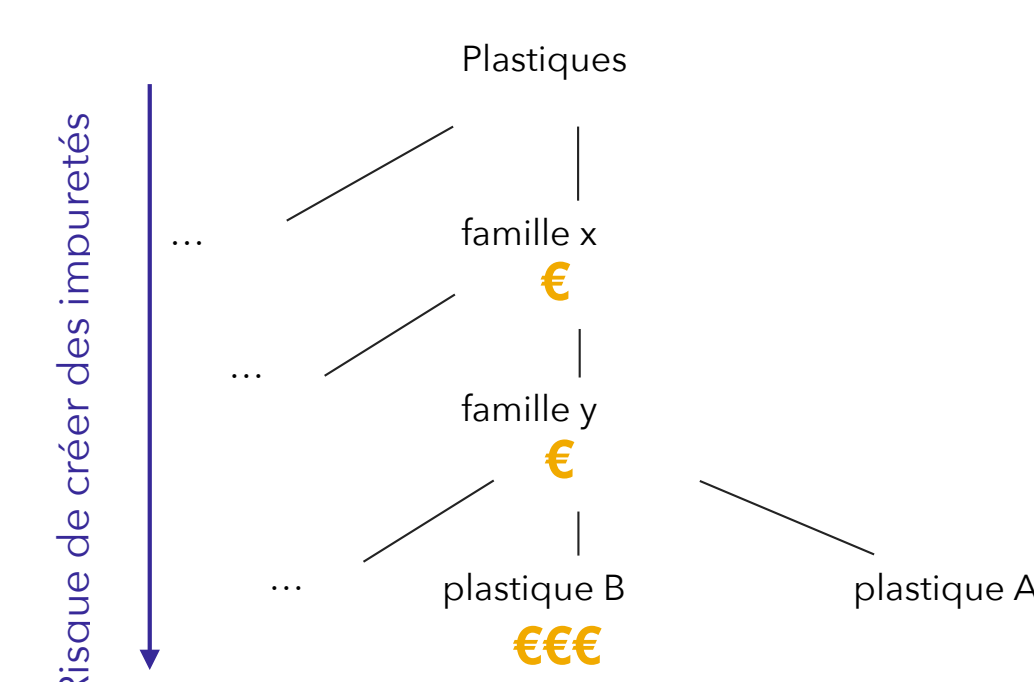
Fusion des résultats de tous les pixels
Combinaison de masses (Dempster-Shafer, R. Yager, D. Dubois...)



Décision finale du tri de l'objet

Compromis entre « pureté » du tri et le rendement économique.

- ▶ Plus le tri d'un plastique est proche des feuilles (classes les plus spécifiques), meilleure est sa rentabilité. En effet une **matière séparée finement** est plus facilement recyclable donc plus chère en général. (€)
- ▶ Plus le tri d'un plastique est proche des feuilles, plus les **risques d'erreurs** sont élevés. (↓)



Et donc ?

L'ambition de cette thèse est de proposer un algorithme capable de prendre des décisions à partir de **données imprécises et incertaines** acquises dans un **contexte industriel** pour trier des plastiques notamment les noirs. Les classifieurs classiques n'ayant pas permis d'obtenir des résultats satisfaisants, nous proposons d'utiliser la **connaissance experte** liée au processus, notamment en exploitant les familles chimiques des plastiques à travers une **classification hiérarchique**.

Auteur

Lucie Jacquin

Encadrants

Abdelhak Imoussaten
François Troussat

Directeurs

Didier Perrin
Jacky Montmain

Partenaires



Réseaux de Neurons Récurrents Multi-tâches pour l'Analyse Automatique d'Arguments

Parties prenantes



IMT Mines Alès
École Mines-Télécom

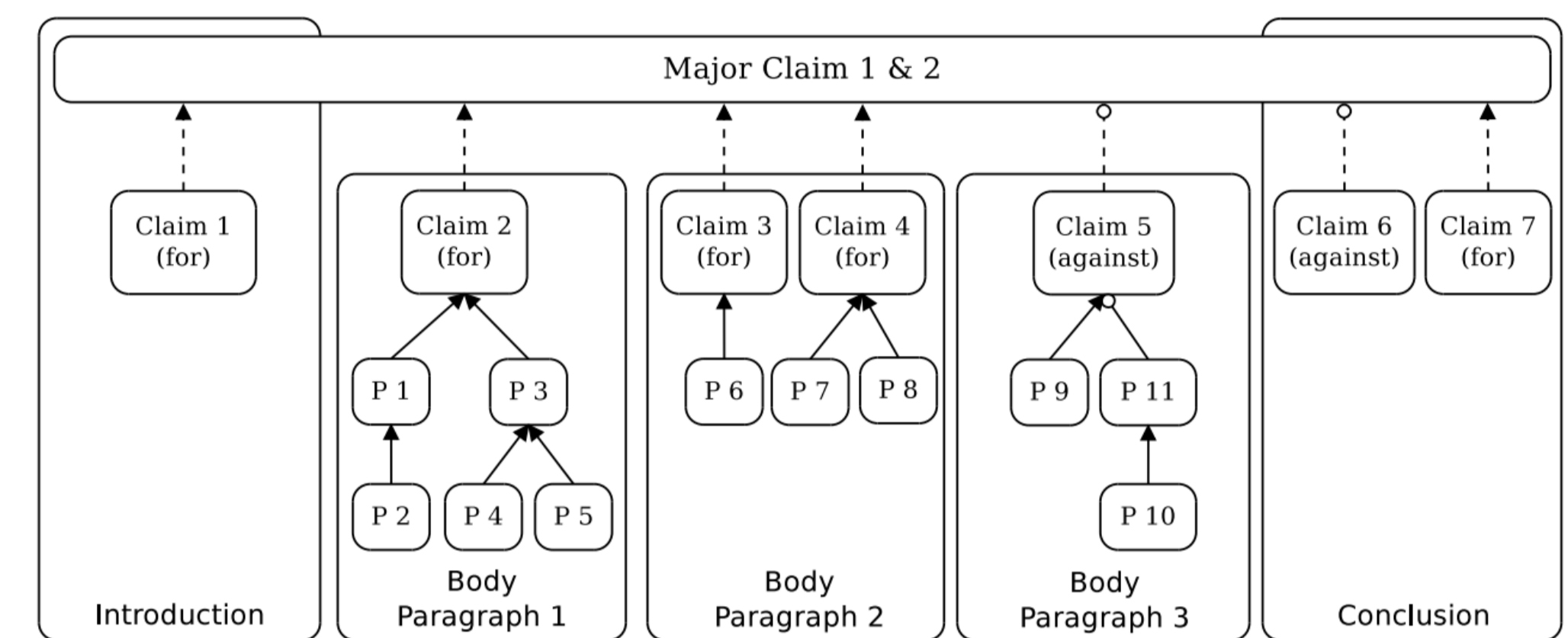


Auteurs

Jean-Christophe MENSONIDES
Sébastien HARISPE
Jacky MONTMAIN
Véronique THIREAU

Définitions et objectifs

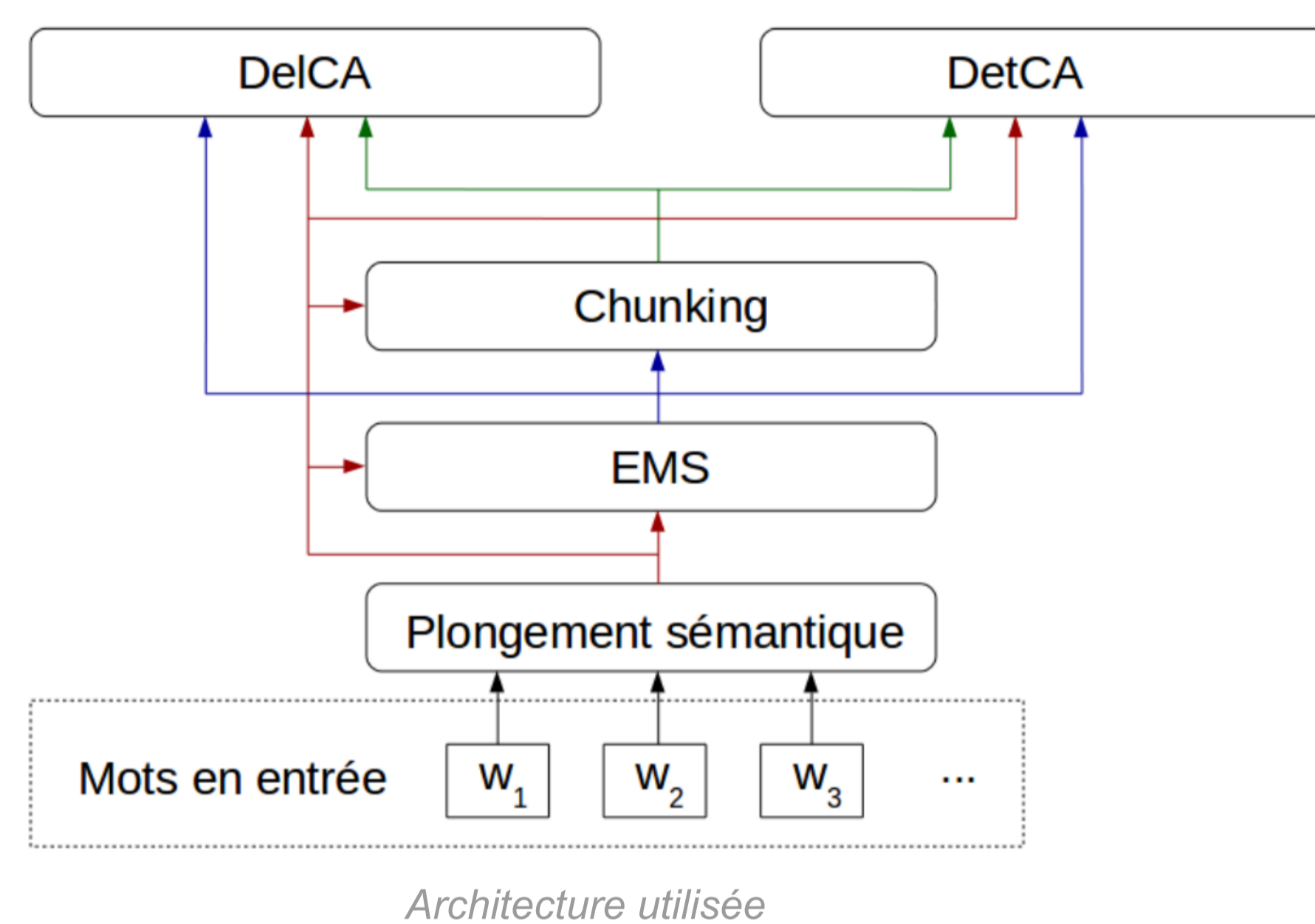
- ▶ **Argumentation** – Ensemble de techniques visant à faire adhérer un interlocuteur à un point de vue qui lui est présenté, en construisant un raisonnement à base d'arguments.
- ▶ **Argument mining supervisé** – Obtenir un graphe structuré de composants argumentatifs à partir de textes non structurés, en ayant recours à des techniques de traitement automatique du langage dans un contexte d'apprentissage supervisé.
- ▶ **Micro-structure argumentative** – Analyse limitée aux interactions au sein d'un même texte. Les éventuelles interactions entre différents textes d'un corpus ne font pas l'objet de cette étude.
- ▶ **3 types de composants argumentatifs**
 - Conclusions majeures : elles reflètent le point de vue global d'un auteur sur un sujet disserté.
 - Conclusions intermédiaires : elles représentent des affirmations qui ne pourraient être acceptées sans justifications complémentaires.
 - Prémises : elles sont le socle permettant de justifier les conclusions intermédiaires avancées.
- ▶ **Chaîne de traitements proposée par Stab et Gurevych (2017)**
 - Délimitation des frontières des composants argumentatifs.
 - Détermination du type de chaque composant argumentatif.
 - Déterminer l'existence de relations de support ou d'attaque entre chaque paire de composants argumentatifs.
- ▶ Stab et Gurevych proposent une approche à base de Support Vector Machine et de caractéristiques manuellement définies par des experts. **Peut-on s'affranchir de la définition de ces caractéristiques en utilisant des techniques d'apprentissage profond ?**



Exemple de structure argumentative d'une dissertation.
Source : Stab et Gurevych, 2017

Approche proposée

- ▶ **Un modèle multi-tâches** – Partager un sous-ensemble de paramètres communs à plusieurs tâches auxiliaires. L'architecture utilisée, découpée en plusieurs couches, est illustrée ci-dessous.



- ▶ **Plongement sémantique** – Associer une représentation vectorielle à chaque mot du texte à traiter.
- ▶ **Etiquetage morpho-syntaxique (EMS)** – Identifier la classe morpho-syntaxique de chaque mot.

Industry observers expect a wide divergence in performance
 NN NNS VBP DT JJ NN IN NN

Nous encodons chaque texte à l'aide d'un réseau de neurones récurrent bidirectionnel de type Gated Recurrent Unit (GRU) et estimons la probabilité d'associer la bonne étiquette EMS k au mot d'indice t à l'aide d'un perceptron multicouches :

$$h_t = [\overrightarrow{GRU}(x_t); \overleftarrow{GRU}(x_t)] \quad f_{c_t} = \text{relu}(W_{fc}h_t + b_{fc})$$

$$p(y_t = k | h_t) = \frac{\exp(W_{sm}f_{c_t} + b_{sm})}{\sum_{c_1} \exp(W_{sm}f_{c_t} + b_{sm})}$$

Avec x_t la représentation vectorielle du mot d'indice t , W et b les matrices et vecteurs de paramètres, relu la fonction Unité de Rectification Linéaire, et c_1 l'ensemble des classes possibles pour l'étiquette EMS.

- ▶ **Chunking** – Identifier la classe de chaque groupe grammatical – réseau de neurones récurrents + perceptron multicouches.

Industry observers expect a wide divergence in performance
 B-NP I-NP B-VP B-NP I-NP I-NP B-PP B-NP

- ▶ **Délimitation des frontières des composants argumentatifs (DelCA)** – Déterminer au mot près les frontières de chaque composant argumentatif – réseau de neurones récurrents + perceptron multicouches.
- ▶ **Détermination du type de chaque composant argumentatif (DetCA)** – réseau de neurones récurrents + mécanisme d'attention + perceptron multicouches.

Expérimentations et résultats

- ▶ Jeux de données utilisés

- Tâches EMS et Chunking : corpus issu de la tâche partagée CoNLL-2000
- Tâches DelCA et DetCA : Argument Annotated Essays (version 2) (Stab et Gurevych, 2017)

- ▶ Baseline et résultats obtenus

Tâche	F1-score obtenus par Stab et Gurevych (2017)	F1-score humain
DelCA	0.867	0.886
DetCA	0.826	0.868

Macros f1-scores obtenus sur les tâches DelCA et DetCA par Stab et Gurevych (2017) et des agents humains

Tâche	w/o EMS & chunking	w/ EMS & chunking
DelCA	0.5934	0.8688
DetCA	0.7529	0.7911

Macros f1-scores obtenus sur les tâches DelCA et DetCA

- ▶ La colonne **w/o EMS & chunking** présente les résultats obtenus sur les tâches DelCA et DetCA sans entraînement spécifique des couches EMS et chunking. La colonne **w/ EMS & chunking** présente les résultats obtenus sur les tâches DelCA et DetCA avec entraînement spécifique des couches EMS et chunking.
- ▶ Les résultats obtenus sont prometteurs et proches de ceux obtenus par Stab et Gurevych (2017), montrant qu'il est en effet possible de s'affranchir de caractéristiques manuellement définies par des experts en utilisant des techniques d'apprentissage profond.
- ▶ Les performances obtenues par la version **w/ EMS & chunking** du modèle surpassent celles de la version **w/o EMS & chunking**, mettant en évidence l'intérêt d'intégrer des tâches auxiliaires (i.e de l'apprentissage multi-tâches) pour traiter les tâches DelCA et DetCA.

Références :

Jean-Christophe Mensonides, Sébastien Harispe, Jacky Montmain et Véronique Thireau, Réseaux de Neurons Récurrents Multi-tâches pour l'Analyse Automatique d'Arguments, *CNIA & RJCIA 2018*, 78, 2018.

Christian Stab and Iryna Gurevych, Parsing argumentation structures in persuasive essays, *Computational Linguistics*, 43(3):619–659, 2017.



Institut Mines-Télécom

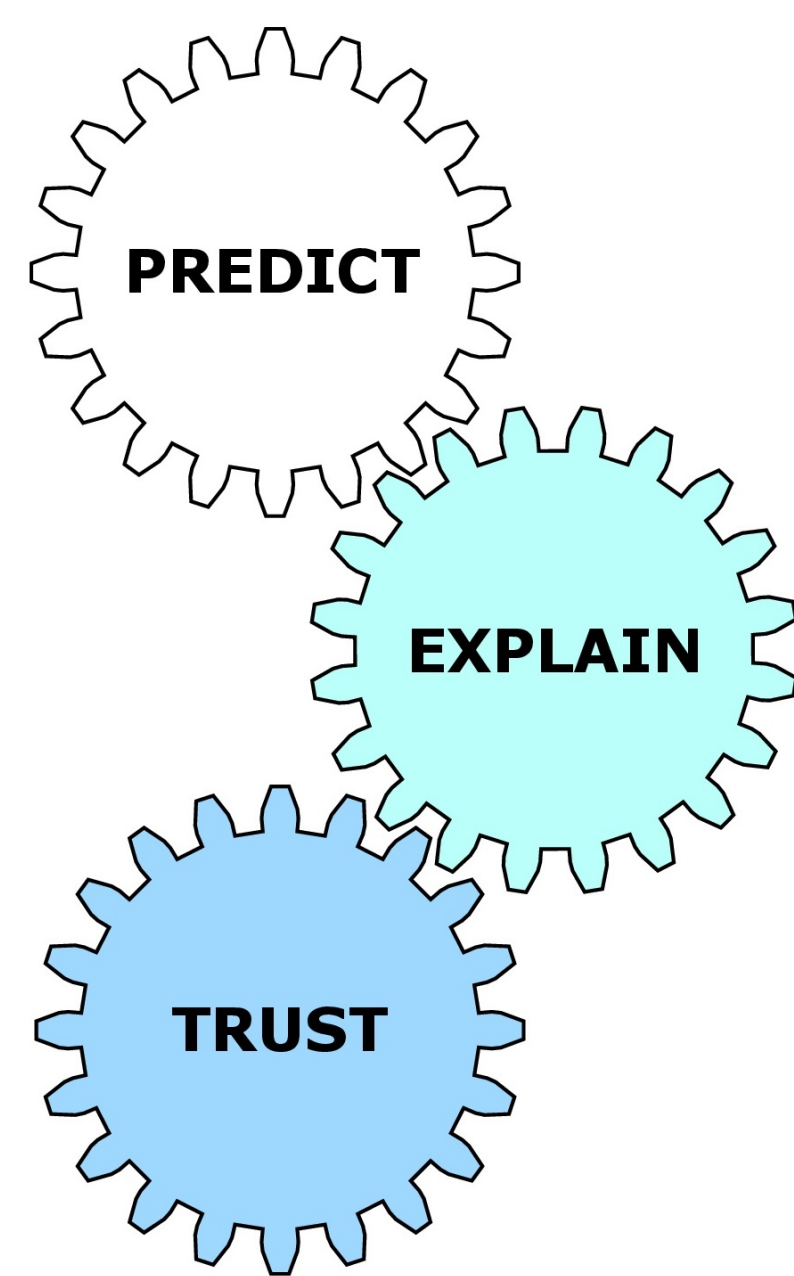


Learning Ensembles of Monotone Boolean Functions For Interpretable Classification

Introduction and Goals

The need of interpretability in critical applications

- ▶ **Understanding a prediction** – Black box models are widely used to perform a huge variety of tasks due to their exceptional predictive performance. However, there are some application areas where *understanding the reason behind the prediction* of a model is at least as essential as the prediction itself.
- ▶ **Predictive Rule Learning** – We developed a novel interpretable classification system that provides *interpretability as a set of rules*, while maintaining predictive performance comparable to black box models.
- ▶ **Predictive Maintenance and Interpretability** – We apply such a system in the context of *data-driven predictive maintenance* of IT services.
 - Every time the model detects an anomaly, it also outputs one or more rules to justify its prediction.
 - Our model learn automatically from data, adapting and discovering new rules during its life cycle.



Interpretable classification workflow
Source : Graziano Mita

Overview of our proposal

Ensembles of Boolean Rule-sets (E-BRS)

- ▶ Our model (E-BRS) exploits some properties of **monotone Boolean functions** to build a boundary separating positive and negative samples.
- ▶ The boundary can be naturally mapped to a set of rules expressed as a function of the input features.
- ▶ A small set of compact rules should be preferred to favor interpretability.
- ▶ We use **an ensemble of rule-based classifiers** that operate on a randomized set of features to ensure high predictive performance.
 - The final set of rules is the union of the rules.
 - In this way, the ensemble is as interpretable as a single weak learner.
- ▶ E-BRS can be tuned to meet specific **performance-interpretability tradeoffs**.

Experimental evaluation

Comparisons and results

- ▶ We run E-BRS on several datasets and made a comparison with many other methods to evaluate both its predictive performance and interpretability.
- ▶ Here, we report a selected subsets of experiments and methods.

DATASET	RBF-SVM	CART	RIPPER-K	E-BRS
ADULT	63.76 (1.33)	62.05 (1.17)	62.78 (1.64)	70.74 (1.26)
BANK	46.77 (2.21)	48.11 (2.08)	44.58 (2.99)	58.31 (2.10)
ILPD	45.52 (9.03)	41.88 (10.66)	18.73 (19.10)	57.31 (7.19)
SAP HANA	96.09 (2.32)	92.04 (2.27)	95.77 (2.88)	96.12 (1.86)

10-fold cross validated F1-score (standard deviation in parenthesis)
Source : Graziano Mita

Parties prenantes



Auteurs

Graziano Mita
Paolo Papotti
Maurizio Filippone
Pietro Michiardi

Partenaires



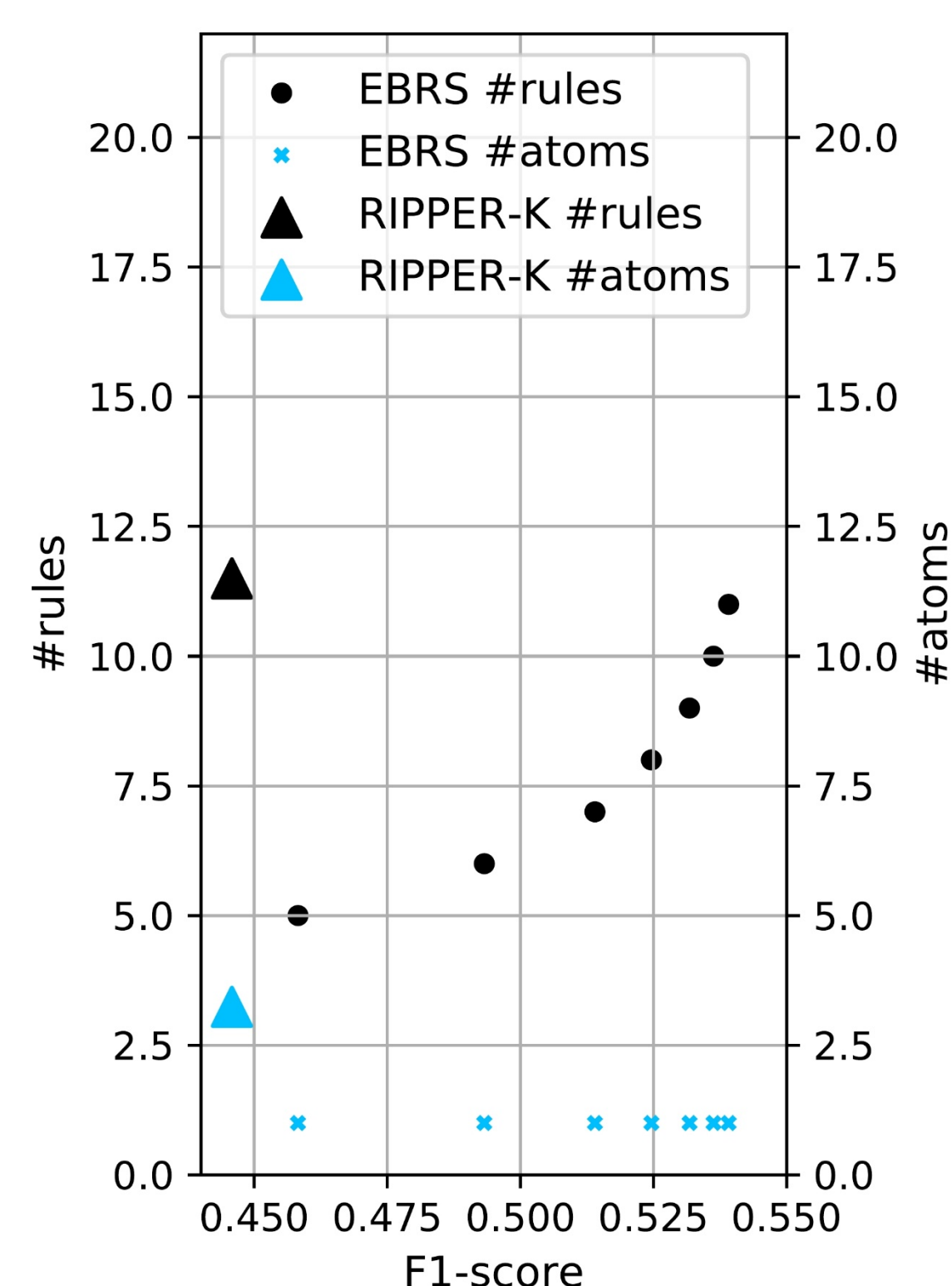
```

R1: if RAM>0.98 and
    (glibc=2.1 or glibc=2.2)
    then crash=1

R2: if ActiveUsers > 10^5 and
    DBsize > 100GB
    then crash=1

R3: else
    crash=0
  
```

Example of Rule-set.
Source : Graziano Mita



Interpretable Rule-sets
for the BANK dataset
Source : Graziano Mita