

Statistiques descriptives

Analyse de données

Algorithmes

(1) INTRODUCTION :
RESEAUX DE NEURONES, APPLICATIONS

Jérôme Lacaille
Expert Émérite Safran

STATISTIQUES ET INTELLIGENCE ARTIFICIELLE

CHAMPS ALÉATOIRES, MACHINES DE BOLTZMANN, IA

CALCUL PARALLÈLE

ANALYSE D'IMAGES

LES MACHINES DE BOLTZMANN

On dispose d'un ensemble de sites $i \in S$ sur un graphe.

Le graphe est défini par un système de voisinage

$$N_i = \{j \in S ; j \text{ voisin de } i\}.$$

En chacun de ces sites on définit une variable aléatoire

$$x_i \text{ on pose } X = (x_i)_{i \in S}$$

Dans le cas des machines de Boltzmann classiques

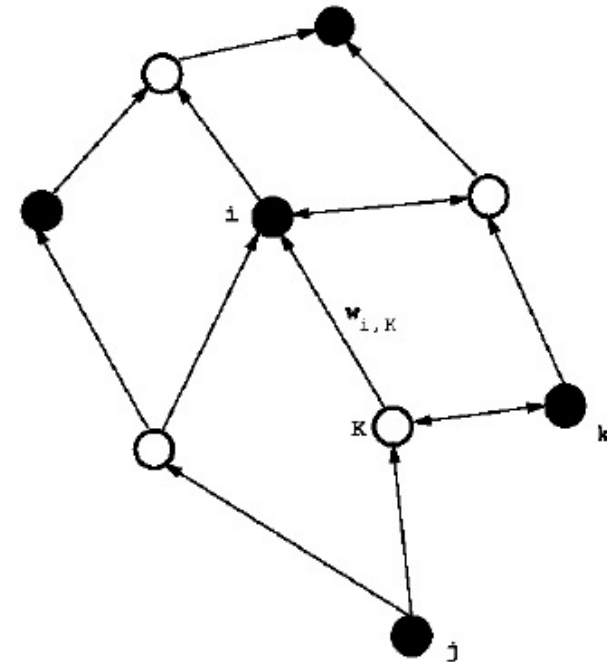
$$x_i \in \{0,1\}$$

On définit un champ markovien par ses probabilités conditionnelles

$$P(x_i | N_i) = f_w(x_j ; j \in N_i)$$

La fonction f_w est une représentation paramétrique des probabilités conditionnelles du champ sur son système de voisinage.

Exemple de connectivité à travers la définition de cliques



$$N_i = \bigcup_{\substack{K \in K \\ w_{i,K} \neq 0}} K$$



LES MACHINES DE BOLTZMANN

On montre que ce champ aléatoire $P_M(X)$ est unique, qu'on peut en écrire une formulation (théorème de Hamersley-Cliford) et que l'on peut définir des mécanismes de relaxation (Glauber, Métropolis) convergeant vers lui.

La machine de Boltzmann est un réseau de neurones exploitant ce mécanisme de relaxation. On implémente une technique d'apprentissage permettant d'adapter le paramétrage w du champ.

Si on fixe une partie des sites en imposant une entrée et une sortie on peut créer un champ qui approxime une probabilité conditionnelle modélisant une fonction Entrée → Sortie.

$$P_M(\text{Sortie} \mid \text{Entrée})$$

Champ aléatoire

$$P_M(y) = \sum_x P_M(x) Q(x, y)$$
$$P_M(x) = \frac{1}{Z} \exp[-H_w(x)/T]$$

Echantillonneur markovien

$$P(x^{t+1} = y \mid x^t = x) = Q(x, y) = \prod_i \Pi_i(x_{N_i}, y_i)$$
$$\Pi_i(x_{N_i}, y_i) \propto \exp\left[\frac{1}{T} \sum_K w_{i,K} J_K(x_K, y_i)\right]$$

Machines de Boltzmann

$$\lambda = \sum_x L(x) \log \frac{L(x)}{M(x)}$$
$$\frac{\partial \lambda}{\partial w} = \frac{1}{T} \left(E_L \left[\frac{\partial H}{\partial w} \right] - E_M \left[\frac{\partial H}{\partial w} \right] \right)$$



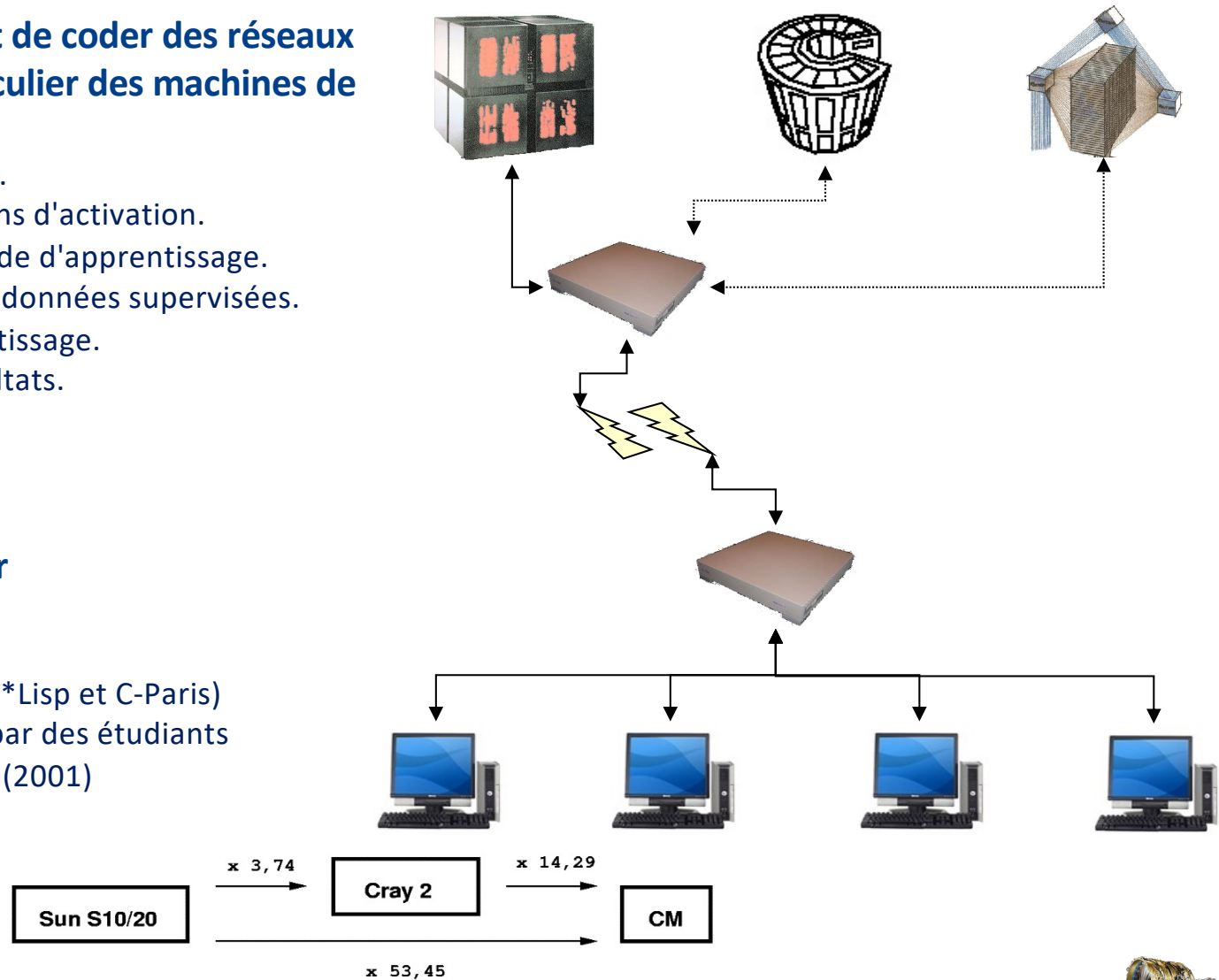
CALCUL PARALLÈLE

→ Le logiciel StoNe permet de coder des réseaux de neurones et en particulier des machines de Boltzmann.

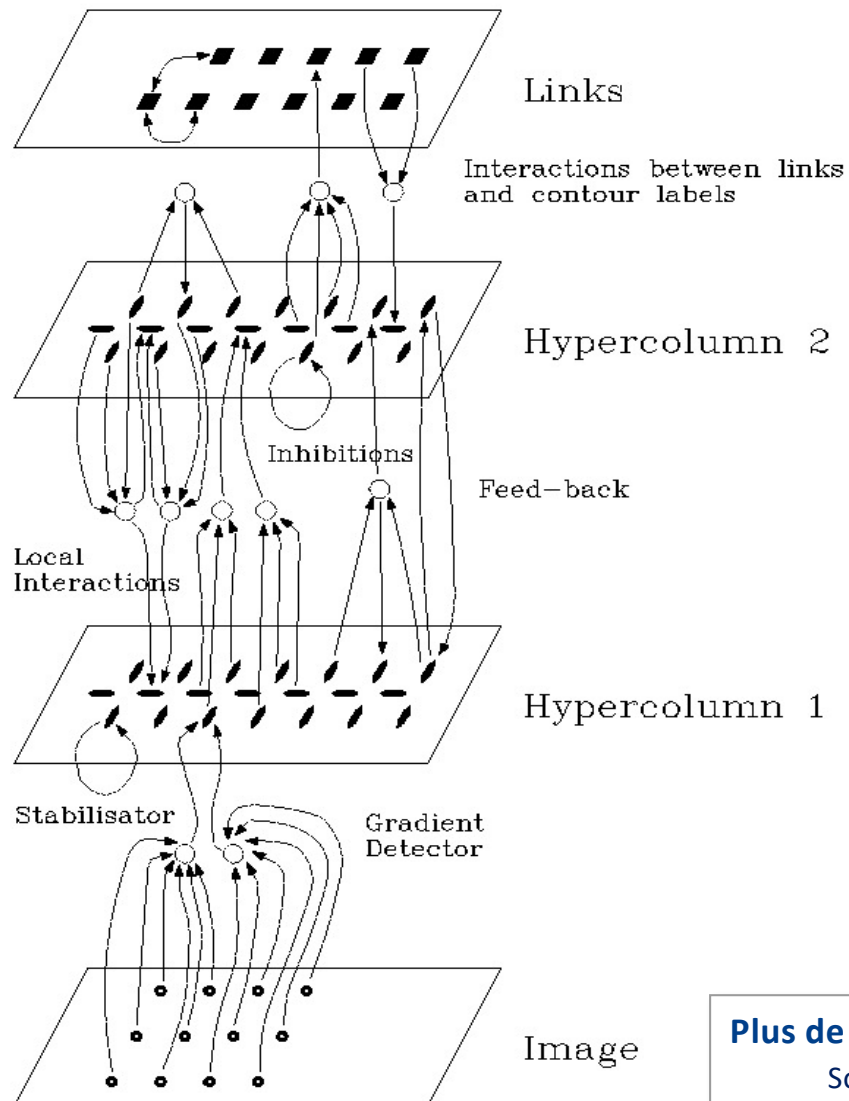
- Description du réseau.
- Définition des fonctions d'activation.
- Sélection de la méthode d'apprentissage.
- Connexion de liste de données supervisées.
- Exécution de l'apprentissage.
- Visualisation des résultats.

→ StoNe est implémenté sur

- Station Sun (C)
- Cray 2 (Fortran 77)
- Connection Machine (*Lisp et C-Paris)
- Maspar (C/Pascal) - par des étudiants MACS/Institut Galilée (2001)



APPLICATION EN IMAGERIE Détection de contours



Plus de trois millions de cellules.
Solution implémentée chez Renault.



L'ANALYSE DE DONNÉES INDUSTRIELLES

INTÉGRATION LOGICIELLE, TRAVAIL EN ÉQUIPE

ANALYSE D'INFLUENCE

SCORING

SURVEILLANCE

PRONOSTIC ET HEALTH MONITORING (PHM)

DATALAB

MIRIAD-PROCESS (1992-2005)

INDUSTRIALISATION D'ALGORITHMES

Besoin

→ Industrialisation d'algorithmes mathématiques

- **Mathématique**
 - Organisation et structure du code
 - Incorporation de codes externes
 - Développement d'algorithmes
 - Documentation méthodologique
 - Test méthodologiques
 - Optimisation algorithmique
- **Informatique**
 - Programmation de composants
 - Tests et benchmarks
 - Optimisation de performances
 - Documentation utilisateur
 - Assemblage de logiciels

Réponse

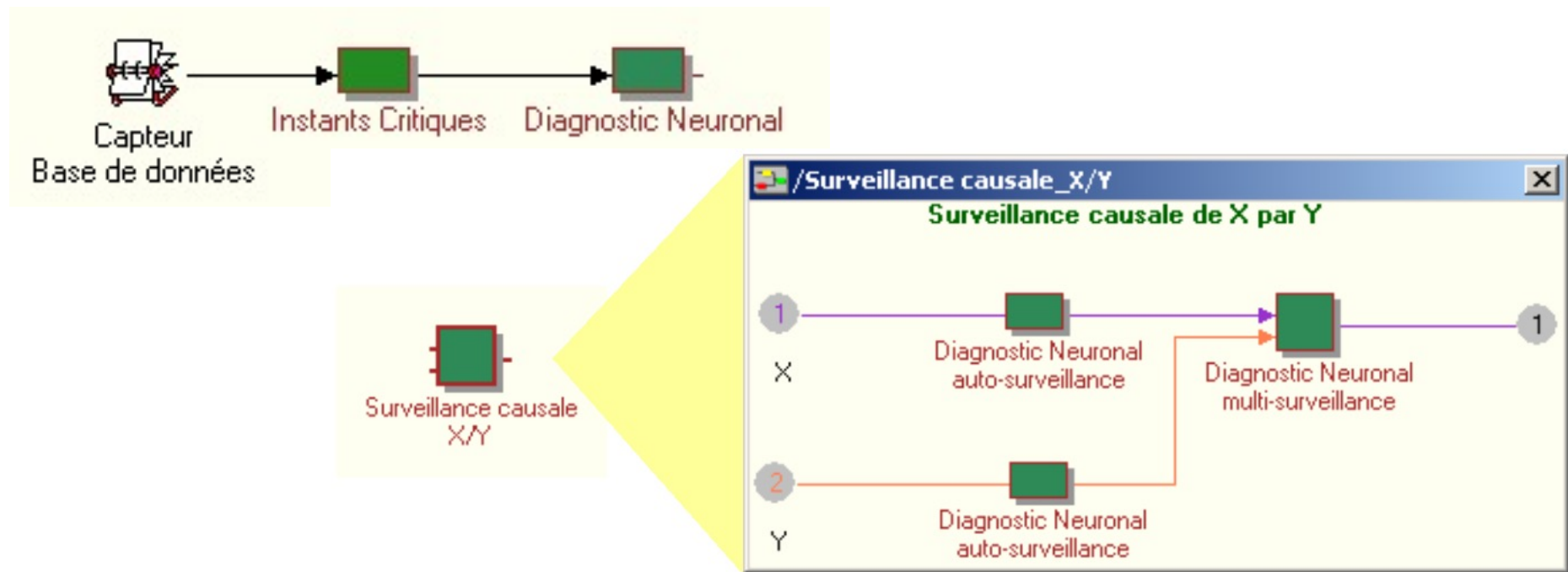
→ Atelier de développement

- **Conception d'algorithmes**
 - Environnement de développement mathématique
 - Normes, structure
 - Contrôle de la qualité
- **Partage de codes**
 - Gestion d'une base de connaissances
 - Interface interactive
- **Production de logiciels**
 - Compiler des applications
 - Intégrer des solutions
 - Déployer dans l'entreprise

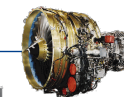
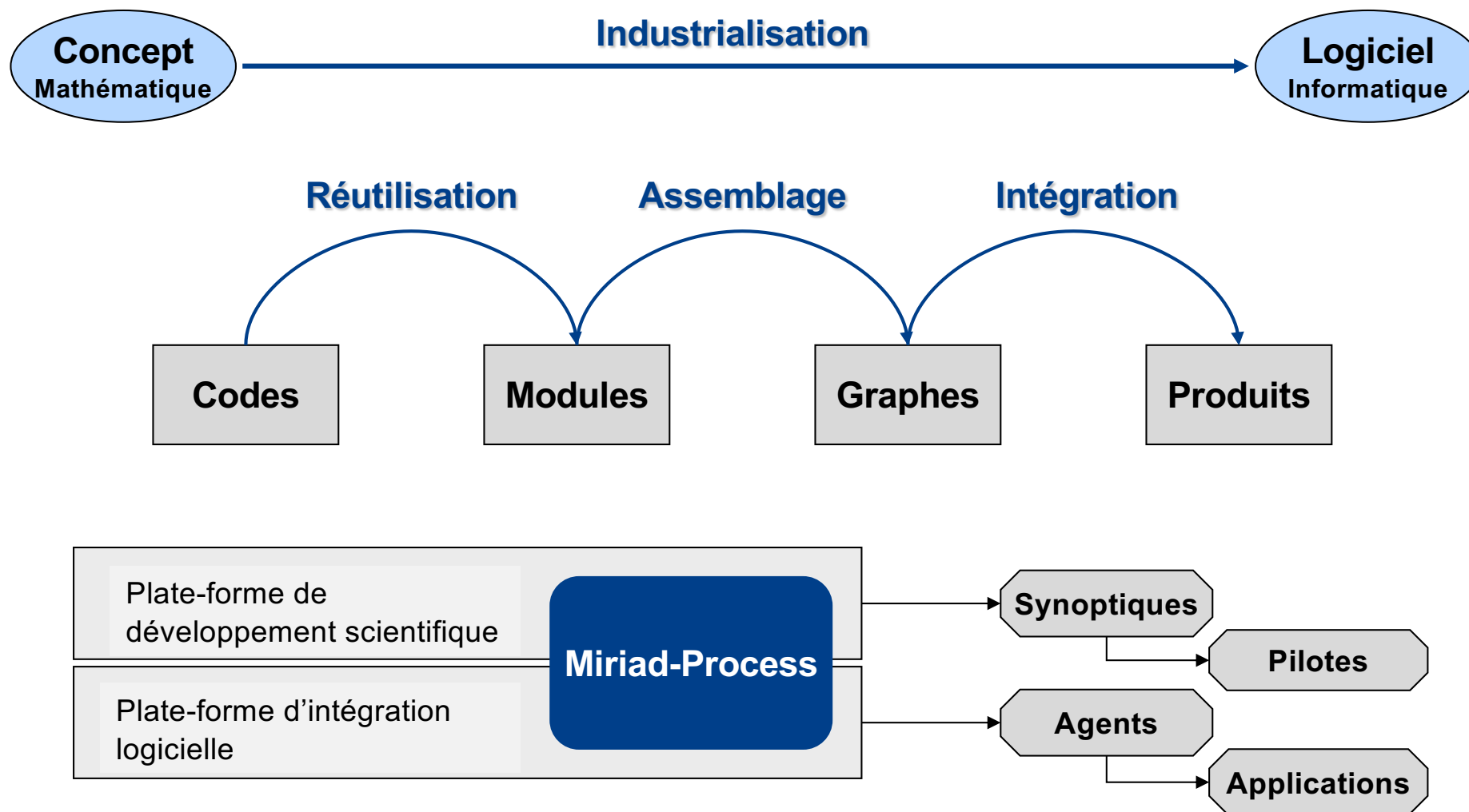


MIRIAD-PROCESS CONCEPT

- Archiver des algorithmes mathématiques
- Gérer les algorithmes auto-adaptatifs
 - Apprentissages par observation des données
- Séquencer des graphes de traitements complexes
 - Combinaisons de plusieurs algorithmes adaptatifs
 - Parallélisme et asynchronisme du séquencement



MIRIAD-PROCESS L'ATELIER



MIRIAD-PROCESS SÉQUENCEMENT

→ Transfert

- Les données d'entrée sont traitées et transmises

→ Attente

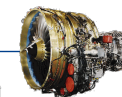
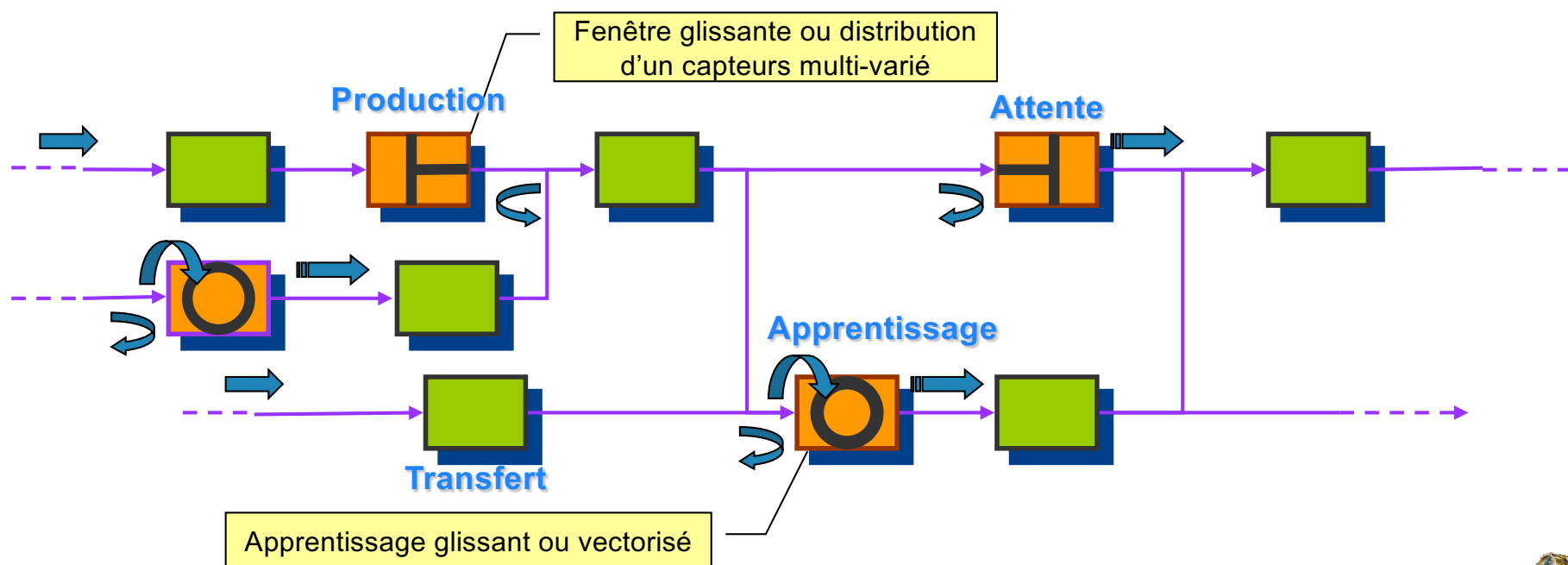
- Le module accumule des données jusqu'à en avoir assez pour effectuer un calcul

→ Production

- L'état du module lui permet de générer des données

→ Apprentissage

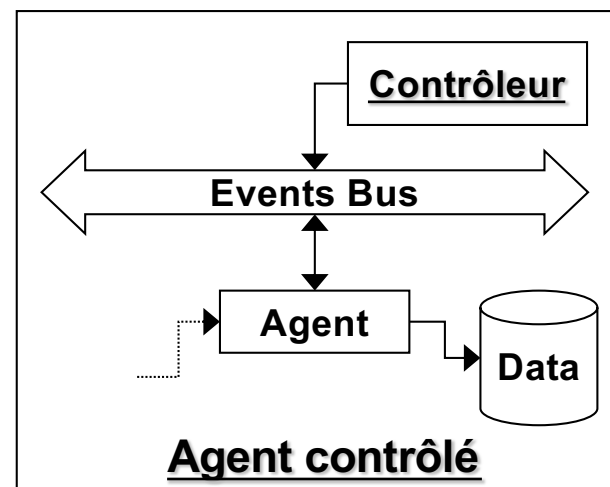
- Le module accumule des données pour son calibrage



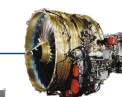
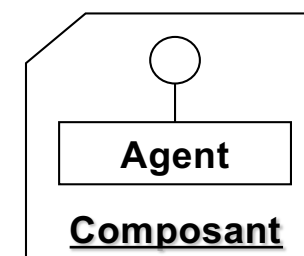
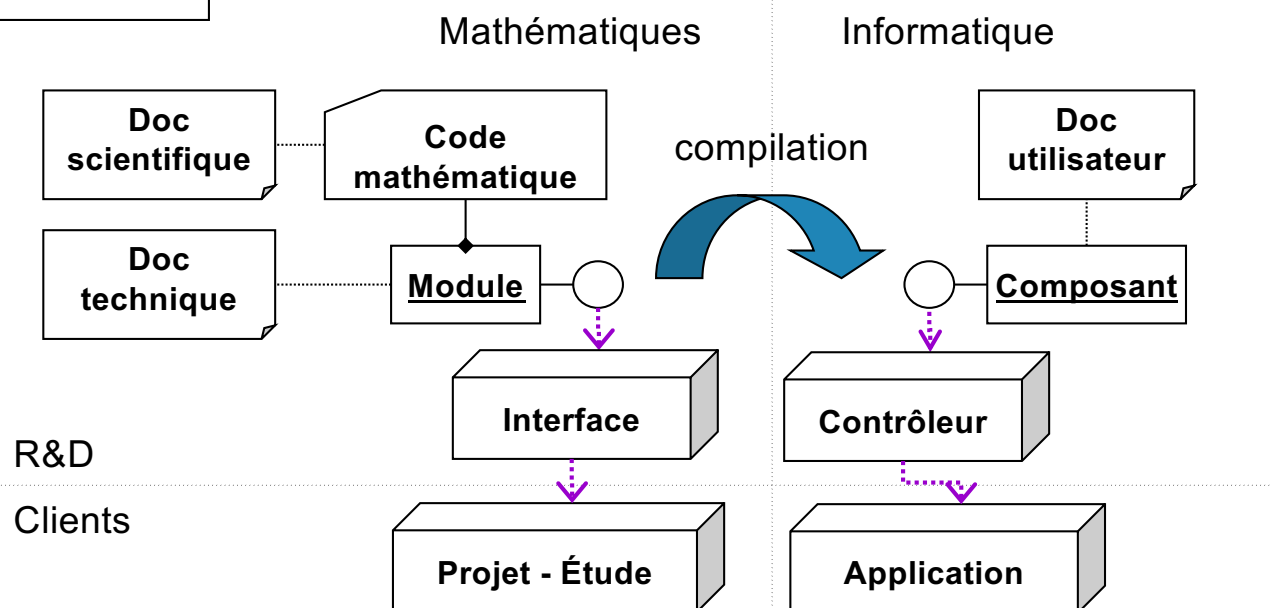
MIRIAD-PROCESS COMPILATION D'AGENTS

<u>Module</u>
Initialisation Paramétrage Transfert Apprentissage Affichage
Paramètres État Statut

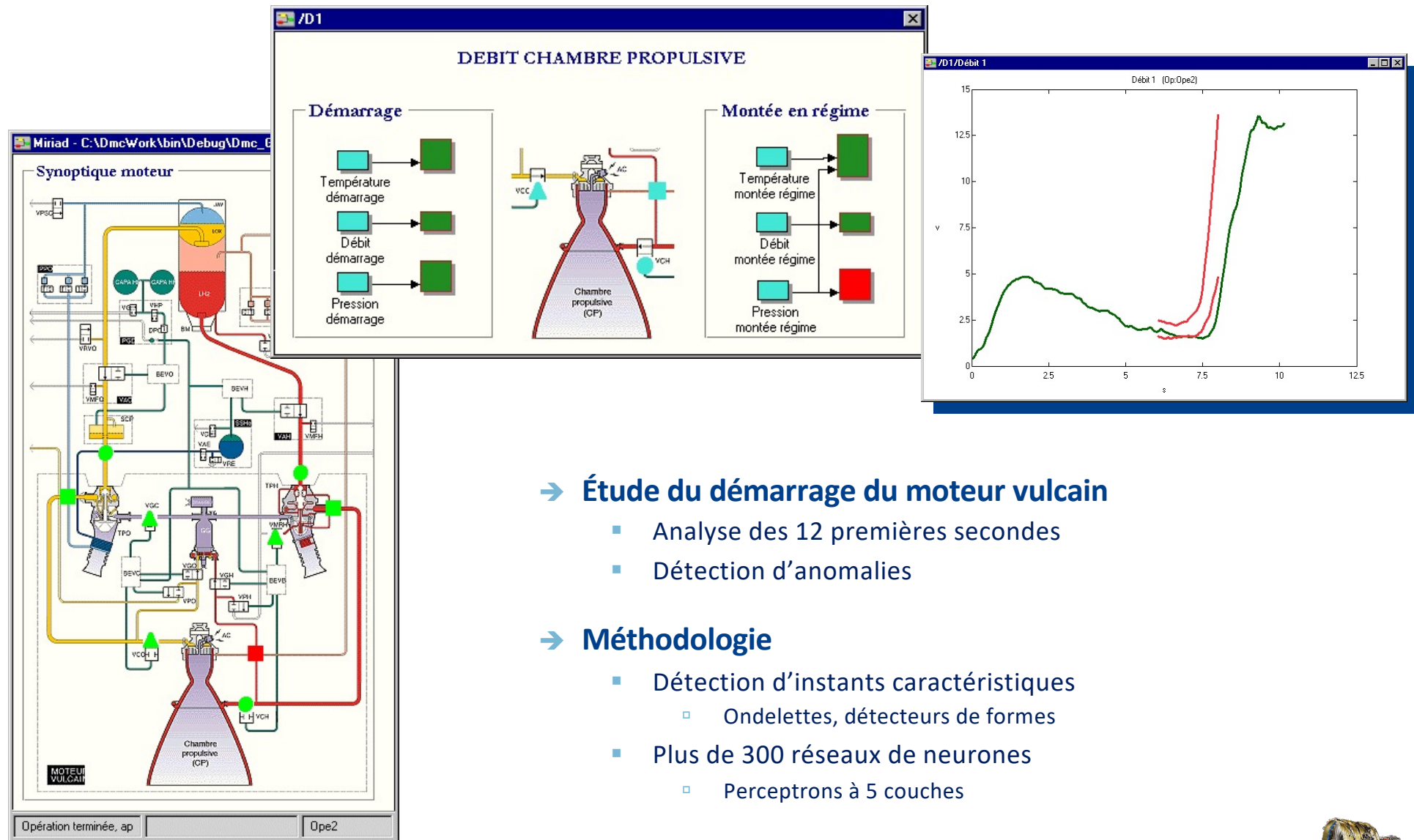
Séquenceur



<u>Agent</u>
Paramétrage Exécution Apprentissage Lancement d'agent Arrêt
Paramètres Statut



RÉALISATION SYNOPTIQUE SEP



→ Étude du démarrage du moteur vulcain

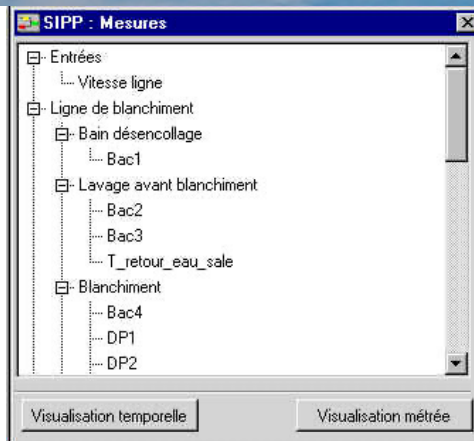
- Analyse des 12 premières secondes
- Détection d'anomalies

→ Méthodologie

- Détection d'instants caractéristiques
 - Ondelettes, détecteurs de formes
- Plus de 300 réseaux de neurones
 - Perceptrons à 5 couches

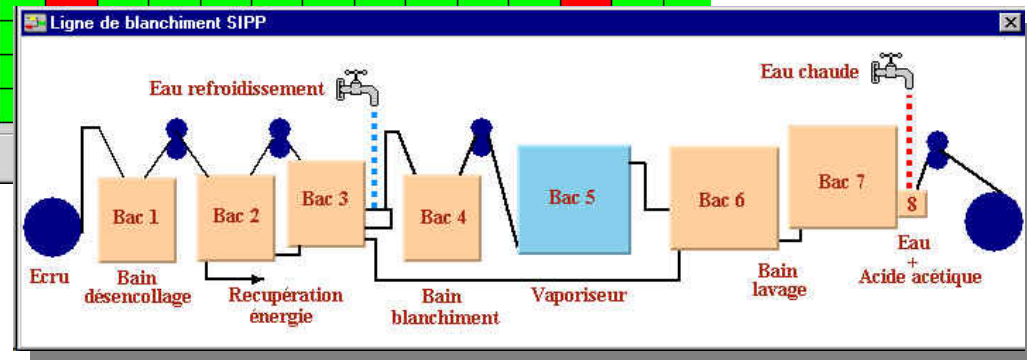
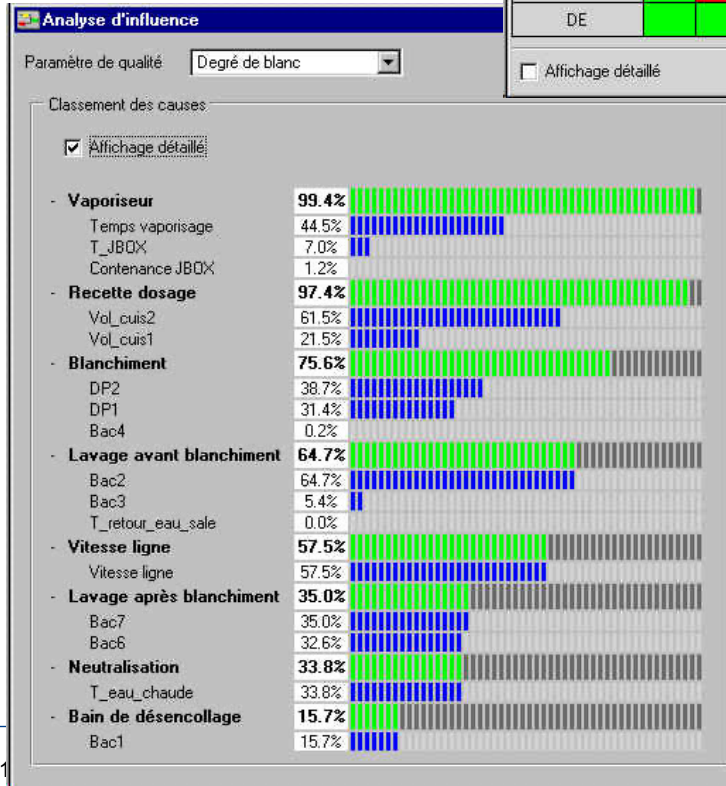


RÉALISATIONS PILOTE TEXTILE



Qualité - Suivi de qualité

N° du Prélèvement	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Prélèvement (m)	5	303	605	908	1203	1504	1803	2103	2403	2703	3006	3303	3603	3904	4204	4503	4736
Peroxyde résiduel																	
Degré de Blanc																	
Tegewa																	
Hydrophilité																	
pH																	
DP																	
Force chaîne																	
Force trame																	
All Chaîne																	
All Trame																	
DE																	



→ Procédé de blanchiment de textile

- Étude de la qualité du procédé
- Remontée aux causes de dégradation

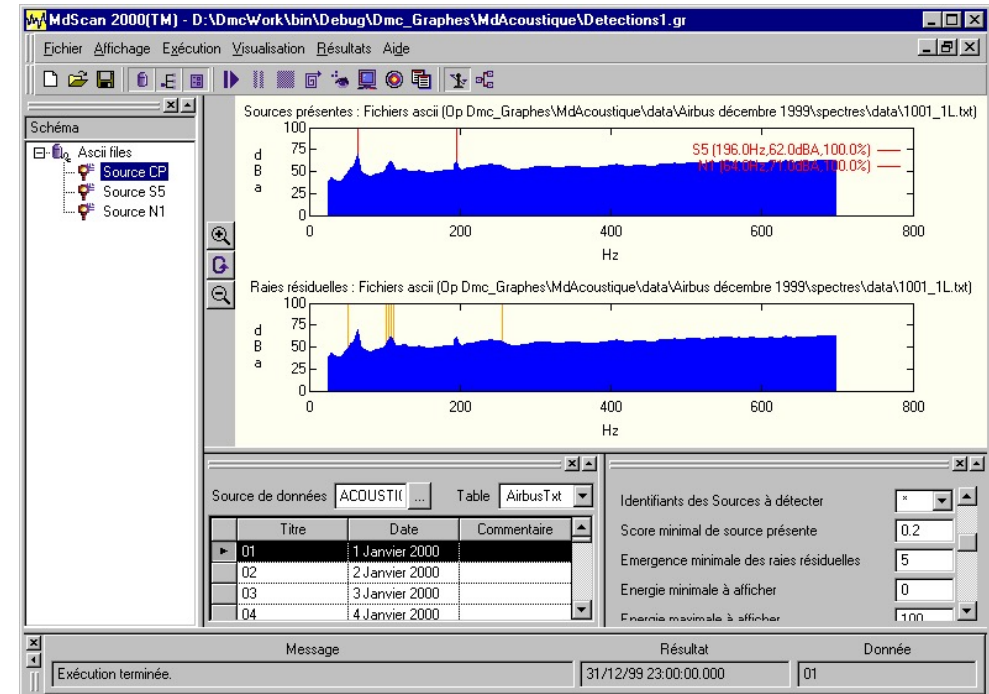
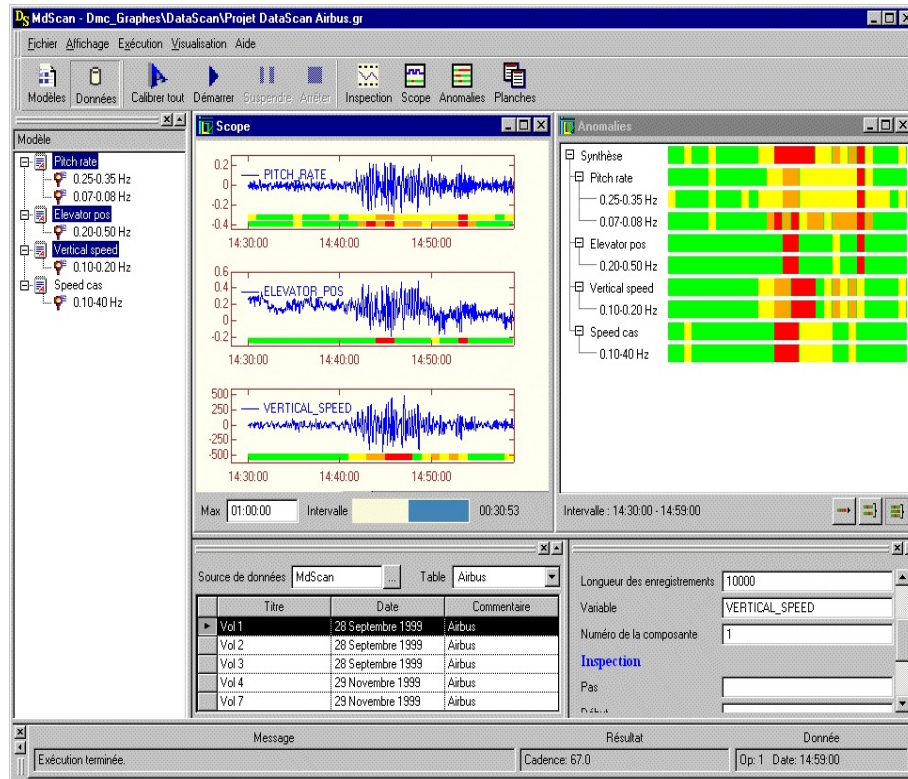
→ Méthodologie

- Analyse d'influence



RÉALISATIONS MDSCAN – SCAN ACOUSTIQUE

1
4



→ MdScan

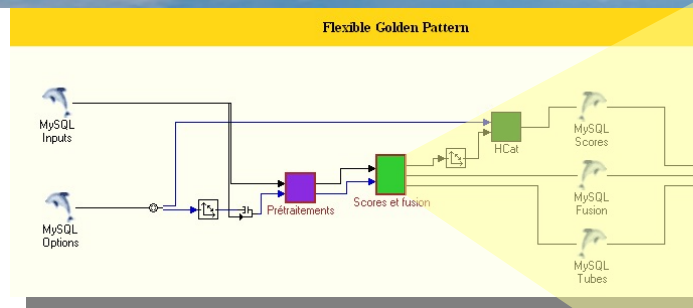
- Détection de formes
 - Temporelles (TraderForce)
 - Fréquentielles (Dassault)
 - Acoustiques (Airbus)
- Diagnostics d'anomalie
- Synthèses de diagnostics

→ Méthodologies

- Détecteurs spécifiques
 - Courbures, Fourier, ondelettes
 - Réseaux de neurones
- Diagnostics d'anomalies
 - Modèles stochastiques
- Fusion de scores
 - Modélisation de distributions

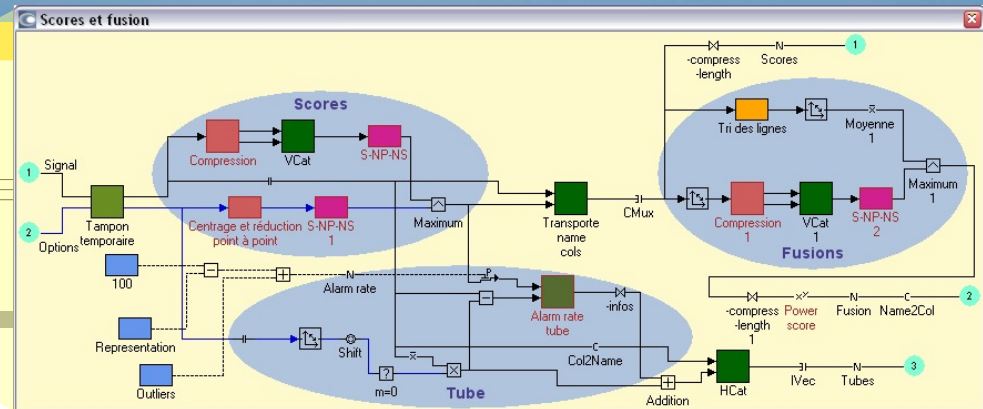
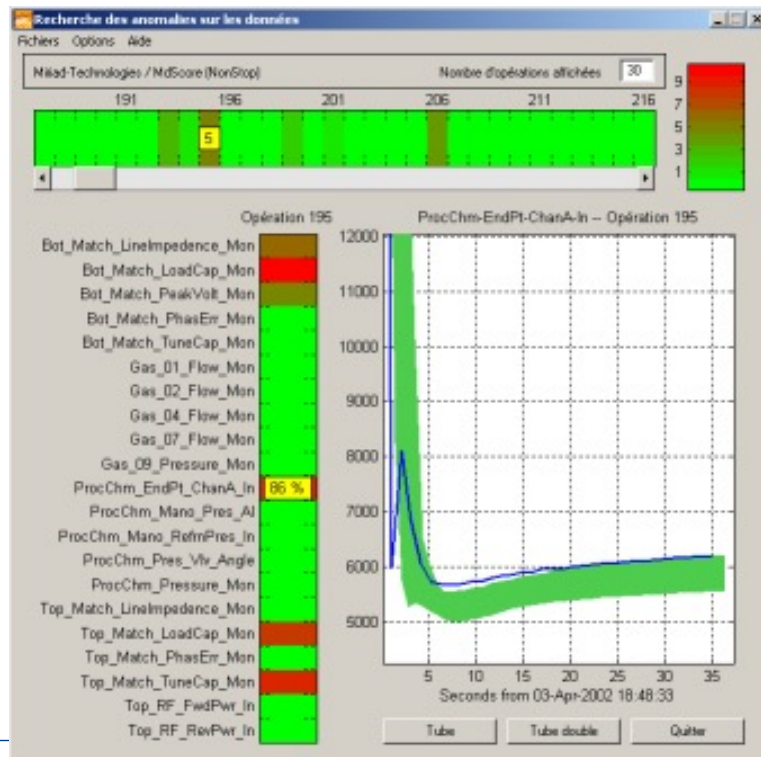


RÉALISATIONS FLEXIBLE GOLDEN PATTERN



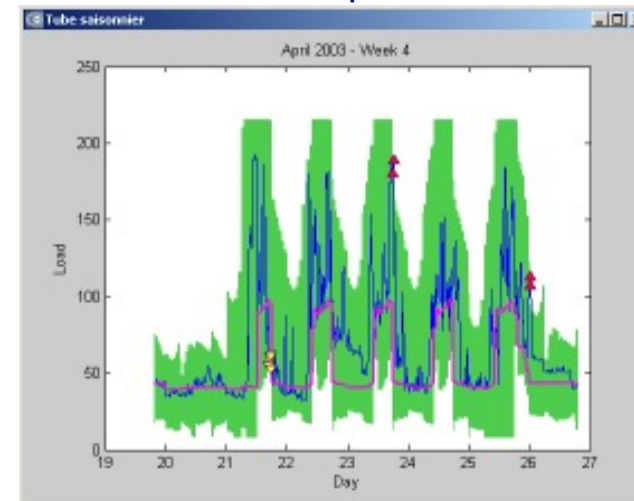
Repérage d'anomalies :

Micro-électronique (SI-Automation)



Suivi des taux de charge :

Réseau informatique (Procédé auto-adaptatif)

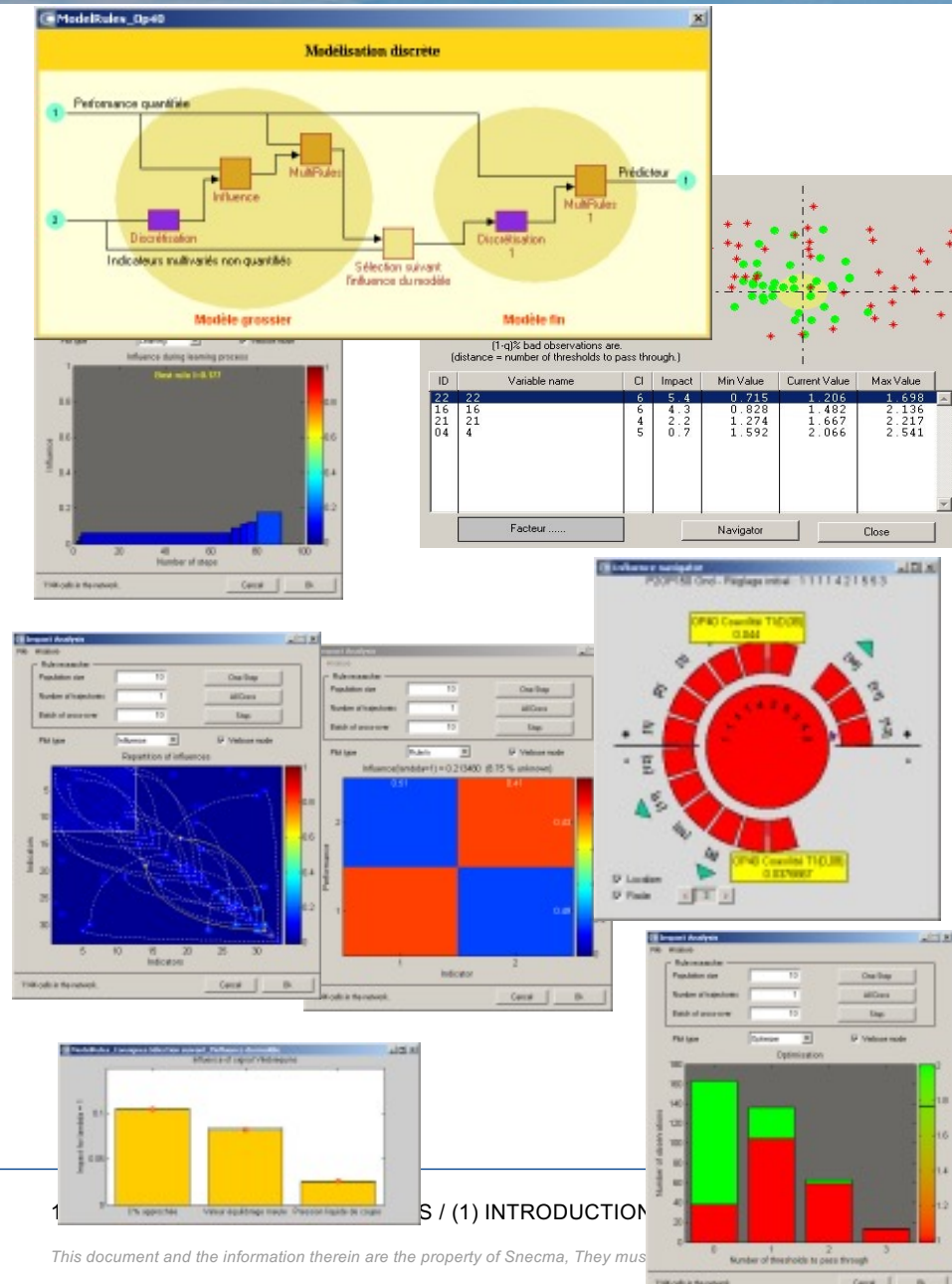


Contrôle de compresseurs d'oxygène :

Détection Sur-consommation (Air Liquide)



RÉALISATIONS ANALYSE D'IMPACT



→ Un outil algorithmique de défrichage de données.

- Études (Renault F1, Renault Cléon, Michelin, ...).
- Analyse a priori (Rhodia, HPCI, CTAS, Dior, Rhône-Poulenc, ...).

→ Interface graphique interactive

- Démonstrations originales.
- Une intégration complète dans Miriad-Process.
- Add-On Excel.
- Application d'analyse générique.

→ On en dérive plusieurs composants :

- Monitoring de procédés industriels (Rhodia, ST-Microélectronique).
- Générations de règles (Ilog), initialisation de la structure de réseaux de neurones.
- Production d'alarmes.



ANALYSE D'INFLUENCE

→ Exemple

- Un constructeur automobile dispose d'une machine-outil ayant plus d'une centaine de réglages possibles. Il veut optimiser les réglages pour améliorer la performance.

→ Mesure de l'influence des facteurs sur la performance

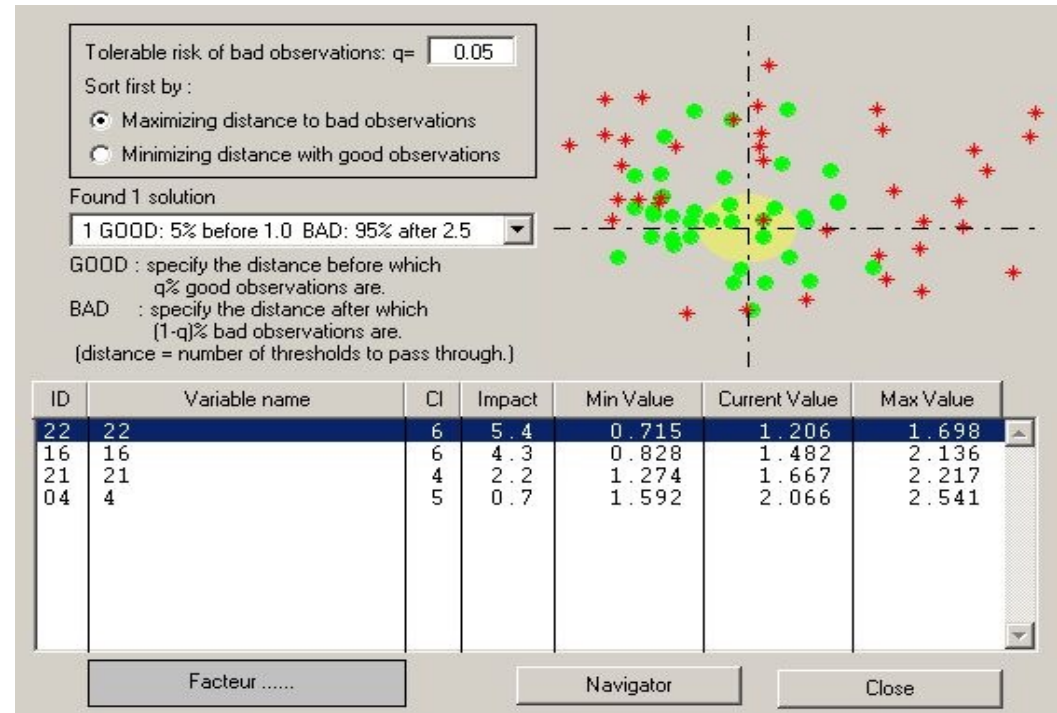
- Une première analyse mono-variée aide la sélection des variables.

→ Sélection de facteurs influents

- On recherche le nombre minimum de facteurs, qui **groupés**, maximisent l'influence sur la performance.

→ Modélisation robuste

- Proposition de zones de réglages optimales.



PR 143



SCORING

→ Normalisation

- La normalisation consiste à se débarrasser des conditions d'acquisition pour disposer de mesures comparables.
- On utilise ensuite des méthodes de régression avec éventuellement une sélection de variable (LASSO) et/ou une méthode robuste (Ridge / ElasticNet) pour supprimer les relations entre mesures.
- Le résidu de régression est utilisé pour la surveillance.



PR 033

→ Flexible Golden Pattern

- Le FGP est une méthode de tube dynamique permettant de définir des bornes de confiance qui dépendent d'autres mesures.



FGP AEC/APC



PR 120

→ Le suivi de tendance

- On utilise des modèles de processus temporels pour anticiper une évolution hors norme.



PR 116



ANALYSE DE LA DÉFECTIVITÉ

→ Contrôle de la défektivité dans une fab de semi-conducteurs.

- Il s'agit de repérer les sources de défektivités
 - Parmi les ateliers d'usinage
 - Parmi les équipements
 - Dans la logistique
 - Les recettes
 - Les matériaux ...

→ La fab produit plus de 3000 wafers par jour.

- Les technologies ont une durée de vie de 6 mois.

→ La qualité est mesurée en fin de process de fabrication

- Calcul d'un rendement de l'usine
 - Peut se dériver en technologies et devices.

→ Sampling intermédiaire de mesure de qualité.

- 20% des wafers (au plus) sont mesurés.



ISMI

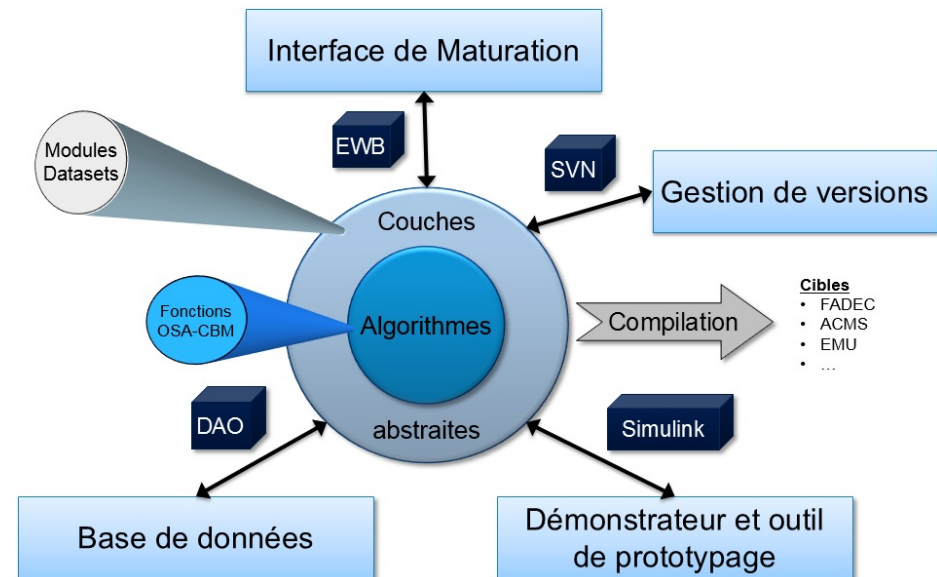


SAMANTA

→ **SAMANTA = Snecma Algorithms Maturation and Test Application.**

→ **Samanta est une plate-forme analogue à Miriad-Process mais conçue sous Matlab/Simulink.**

- Peu de ressources en développement informatique à Snecma.
- Développement d'algorithmes
 - Encapsulation de modules génériques.
 - Graphes de traitements.
 - Macro-modules déployables sur des calculateurs de bord ou sur le système sol.
- Maturation d'algorithmes
 - Intégration d'un cahier de laboratoire E-Workbook (IDBS).



→ **Matlab Expo** [[Présentation de SAMANTA](#)]

→ **Etude du démarrage d'un moteur** [[PR 033](#)].

→ **Identification de capteur fautif** [[PR 036](#) et [PR 120](#)].

→ **Détection d'anomalies à la volée** [[PR 059](#), [PR 099](#) et [PR 148](#)].

→ **Visual Mining / Cartographie** [[PR 060](#)].

→ **Anticipation d'usure** [[PR 116](#)]

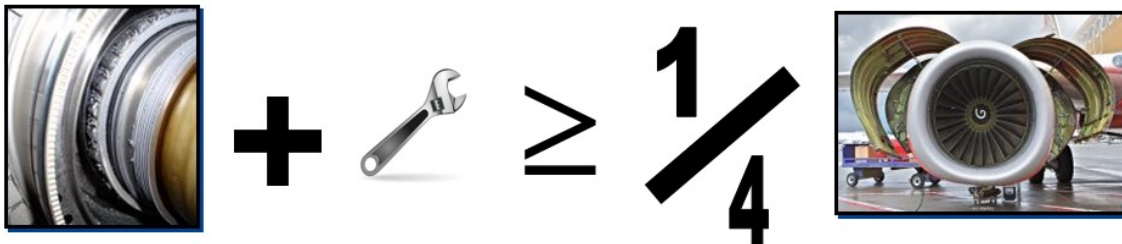


PHM & DATALAB

→ PHM : Prognostic & Health Monitoring

- Première mission : développer des algorithmes pour la surveillance des moteurs d'avions.
- Les algorithmes doivent être pérennes (50 ans de durée de vie d'un moteur).
- Intégration des outils algorithmiques Samanta au processus de développement Snecma.

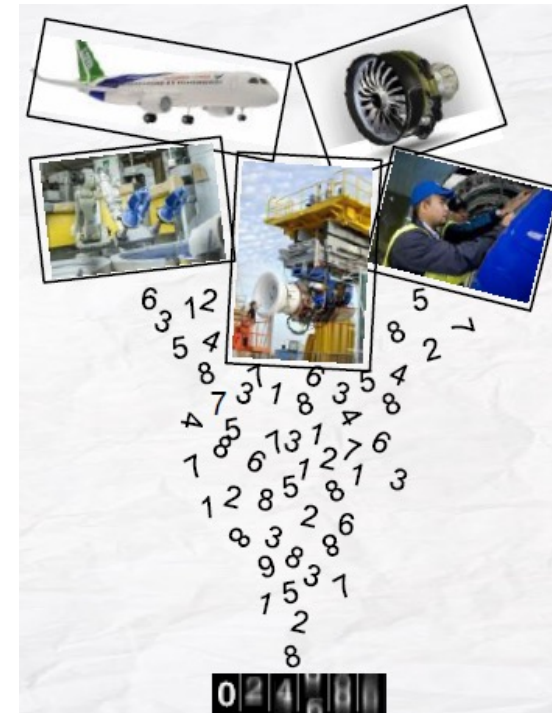
[\[PR 094\]](#)



→ DATALAB

- Conception et animation d'un laboratoire d'analyse de données à Snecma.

[\[PR 152\]](#) [\[PR 159\]](#)



ETUDES INDUSTRIELLES

MIRIAD TECHNOLOGIES

SI AUTOMATION / PDF SOLUTIONS

SNECMA / SAFRAN

MIRIAD - ETUDES INDUSTRIELLES

Rhone-Poulenc	Diagnostic d'anomalie en raffinage.	Modélisation neuronale conditionnelle. Implémentation sous Miriad-Process d'une solution permettant de gérer simultanément plus de 300 réseaux de neurones analysant les dépendances entre capteurs suivant différentes phases du processus de raffinage.
Saint Gobain	Analyse de la qualité du verre en production.	Définition théorique et spécification de l'algorithme d'analyse d'influence. Conception de l'algorithme de synchronisation de capteurs. Mise en œuvre informatique d'un pilote et installation dans l'usine d'Auvelais.
IFTH (SIPP, SAIC)	Analyse du procédé de blanchiment des textiles.	Recherche des causes de dégradation du degré de blanc à l'aide des algorithmes et des modules Miriad-Process d'analyse d'influence.
Christian Dior	Influence de la matière première et des paramètres sur la qualité des poudres.	Utilisation des modules d'analyse d'influence.

MIRIAD - ETUDES INDUSTRIELLES

Cognac Martel	Anticipation de l'instant de coupe.	Modules de modélisation et utilisation de l'interface Miriad-Process pour la conception du prototype.
Renault F1	Optimisation de la structure des moteurs.	Utilisation des algorithmes d'analyse d'influence pour trouver les assemblages les plus performants.
Air Liquide – Ales	Anticipation de l'instant de vidage de bouteilles d'oxygène.	Conception, réalisation de la modélisation et implémentation de l'algorithme de calcul de la probabilité de franchissement d'un objectif sous un horizon donné.
Rhodia – HPCII	Détection d'anomalie en production du latex.	Conception et programmation de l'algorithme de modélisation et de détection.
Rhodia – PPMC	Modélisation et contrôle de la viscosité.	Utilisation des modules de prédiction neuronale et d'analyse d'influence. Sélection automatique des mesures influentes, gestion de la robustesse et du taux de généralisation.



MIRIAD - ETUDES INDUSTRIELLES

Dassault	Détection d'anomalies vibratoires sur le Rafale.	Conception d'une plate-forme (Scan) dédiée à l'environnement vibratoire et implémentation d'algorithmes spécialisés.
EADS – Airbus	Détection et identification de sources sonores pendant les vols d'essais.	Conception d'une solution spécialisée dans l'aéronautique. Utilisation des modules de modélisation neuronale. conception d'algorithmes de détection de formes dans le domaine fréquentiel. Spécification, conception et développement de la technologie du push-up de Miriad (détection automatique et non supervisée de nouvelles sources sonores).



MIRIAD - ETUDES INDUSTRIELLES

DiagOnline	Détection automatique de formes chartistes	Développement d'algorithmes de détection de forme spécialisés. Conception d'algorithme de reconnaissance de forme neuronal utilisant un apprentissage supervisé interactif. Utilisation du logiciel MdScan en ligne.
CPR	Gestion automatique d'un fond de placement sur des futures.	Conception d'un algorithme de modélisation par ondelettes glissante et d'un algorithme de détection automatique de retournement de tendance. Exploitation des détections par la mise en œuvre d'une stratégie de mise utilisant les probabilités de retournement de tendances.
SGAM	Analyse multi-variée des cours journaliers.	Conception d'un modèle de traitement multi-varié de cours financiers. Utilisation d'un processus markovien pour la stratégie de mises.



MIRIAD - ETUDES INDUSTRIELLES

Renault	Optimisation du réglage d'une machine outil en usinage des vilebrequins à Cléon.	Utilisation de l'analyse d'influence, construction d'un modèle stochastique, conception d'un pilote, application à l'étude de la machine Gustina en utilisant un processus markovien pour l'introspection du modèle.
Acetex	Conduite de Catalyse	Utilisation des modules Miriad-Process de modélisation neuronale.
Air Liquide – CRCD	Détection de sur-consommation sur des compresseurs d'oxygène.	Développement de module de modélisation contextuelle et de détection d'outliers sous Miriad-Process.
Air Liquide – CTAS	Détection automatique et non supervisée de défauts de soudage.	Développement d'une solution algorithmique sous Miriad-Process. Installation d'un prototype.



MIRIAD - ETUDES INDUSTRIELLES

SIEMA	Surveillance du fonctionnement et diagnostics d'anomalies du comportement des aiguillages.	Modules de modélisation neuronaux, modules de synthèses multi-variées par tubes de confiance conditionnels.
SI-Automation	Analyse et surveillance en ligne d'équipements de production en micro-électronique.	Conception, spécification et développement de l'algorithme du « Flexible Golden Pattern » (score de normalité des opérations d'usinage). Conception d'un algorithme de scoring non paramétrique. Utilisation des modules de modélisation neuronale.
ST-Micro-electronics	Exploitation des mesures de défektivité pour l'optimisation de la fabrication de semi-conducteurs.	Conception, spécifications et développement d'un système temps réel utilisant une architecture distribuée pour déployer des algorithmes de modélisation, d'anticipation et de diagnostic. Utilisation des codes d'analyse d'influence pour le contrôle de la population d'agents algorithmiques.



MIRIAD - ETUDES INDUSTRIELLES

Solios Carbone	Modélisation de la densité d'une anode.	Le code développé est un capteur logiciel permettant d'estimer la densité d'une anode à partir des mesures temps-réel effectuées lors de la fabrication.
Solios Environnement	Modélisation du taux de rejets polluants.	Conception d'un capteur logiciel permettant de modéliser le taux de déchets polluants en fonction des paramètres de production et des données d'environnement comme le taux d'hygrométrie, la température, etc.
Raffineries de Dunkerque (Sherpa)	Etude la densité du goudron résiduel.	Des mesures coûteuses et peu fréquentes du taux de pénétration du goudron sont progressivement remplacées par un capteur logiciel estimant cette valeur.
Altis semi-conducteurs	Boucle de contrôle « run-to-run »	Mise en place d'une stratégie de contrôle d'équipements chez Altis.
Micron	Diagnostic automatique d'anomalie	Installation en ligne de l'algorithme FGP (Flexible Golden Pattern)

A SUIVRE

PROBABILITÉS ET STATISTIQUES