

Appel à candidature pour thèse CIFRE - Saint Gobain : « Méthodes statistiques pour la génération automatique de lois de commande efficaces sans modèle »

1 Résumé du sujet de thèse (sujet détaillé en annexe)

De nombreuses étapes de fabrication dans le groupe Saint Gobain impliquent de transformer de manière très précise des produits semi-finis, par exemple en chauffant ou en appliquant des forces sur des plaques de verres. La plupart du temps, aucun modèle satisfaisant de la transformation n'est disponible. Les différents réglages (puissance de chauffe, position des outils, ..) sont effectués par des régulateurs très simples mais pas forcément très pertinents (typiquement : des régulateurs PID prenant une seule variable en entrée) ou par des intervenants humains sur la base de leur seule expertise. Durant ou après ces transformations, de grandes quantités de données (parfois plusieurs milliers de variables pour chaque pièce) sont collectées. Ces données peuvent aussi bien être des mesures physiques en temps réel (comme des évolutions de température) que des contrôles qualités a posteriori (comme des mesures très précises de l'état de surface). Des séries temporelles sur plusieurs années sont disponibles sur certaines chaînes de production.

Le but de la thèse est de proposer une méthode automatique de génération de lois de commande efficaces (dans l'idéal : optimales) pour chacune de ces étapes de fabrication. Le terme « lois de commande » est à comprendre dans un sens très large qui inclut le calcul de la valeur d'un contrôle à un instant donné, mais aussi la détection de défaut ou d'usure des outils et la programmation d'opérations de maintenance préventive. Le mot « automatique » sous-entend qu'on espère limiter le recours à l'expertise humaine uniquement aux étapes initiales de compréhension du problème métier.

2 Profil recherché

Le/la candidat(e) retenu(e) aurait idéalement une double compétence (probabilités/statistique et automatique). En pratique, seules des connaissances de base sur l'analyse en composantes principales et les méthodes de classification (hiérarchique, partition *K-means*, etc.) seront nécessaires dès le début, toutes les autres compétences (en probabilités, statistiques et automatique) pourront être acquises en cours de thèse.

Le/la candidat(e) retenu(e) sera amené(e) à communiquer en anglais (communications scientifiques écrites et orales). Une compréhension basique du français est souhaitable. Des formations complémentaires en français et en anglais sont possibles pendant la thèse.

Seul un master 2 est formellement requis. Aucune condition d'âge, de genre ou de nationalité.

3 Financement, calendrier, localisation

Type de financement :	CIFRE	
Durée :	36 mois, début dès que possible	
Entreprise :	Saint Gobain	
Laboratoire :	Institut Élie Cartan de Lorraine	
École doctorale :	IAEM (ED 77)	
Localisation :	Nancy ou Paris, dans des proportions à négocier	
Co-directeurs de thèse :	Thomas Chambrion (MCF HDR 26, contrôle)	50 %
	Radu Stoica (PR 26, probabilités/statistique)	50 %

4 Contact

Les dossiers de candidatures (lettre de motivation, CV et relevés de note de master, éventuellement : publications, mémoires, recommandations,...) sont à faire parvenir par voie électronique à

Sujet de thèse : « Méthodes statistiques pour la génération automatique de lois de commande efficaces sans modèle »

1 Motivation industrielle

1.1 Existant

De nombreuses étapes de fabrication dans le groupe Saint Gobain impliquent de transformer de manière très précise des produits semi-finis, par exemple en chauffant ou en appliquant des forces sur des plaques de verres. La plupart du temps, aucun modèle satisfaisant de la transformation n'est disponible. Les différents réglages (puissance de chauffe, position des outils, ..) sont effectués par des régulateurs très simples mais pas forcément très pertinents (typiquement : des régulateurs PID prenant une seule variable en entrée) ou par des intervenants humains sur la base de leur seule expertise.

Durant ou après ces transformations, de grandes quantités de données (parfois plusieurs milliers de variables pour chaque pièce) sont collectées. Ces données peuvent aussi bien être des mesures physiques en temps réel (comme des évolutions de température) que des contrôles qualités a posteriori (comme des mesures très précises de l'état de surface). Des séries temporelles sur plusieurs années sont disponibles sur certaines chaînes de production.

1.2 Problèmes posés

Le but de la thèse est de proposer une méthode automatique de génération de lois de commande efficaces (dans l'idéal : optimales) pour chacune de ces étapes de fabrication. Le terme « lois de commande » est à comprendre dans un sens très large qui inclut le calcul de la valeur d'un contrôle à un instant donné, mais aussi la détection de défaut ou d'usure des outils et la programmation d'opérations de maintenance préventive. Le mot « automatique » sous-entend qu'on espère limiter le recours à l'expertise humaine uniquement aux étapes initiales de compréhension du problème métier.

2 Difficultés et verrous scientifiques

Les difficultés sont multiples et pour un certain nombre originales.

2.1 Difficultés de modélisation

La première difficulté est que l'on ne connaît pas de modèle pour la transformation étudiée. Cette non-connaissance est à comprendre dans un sens très fort :

- les variables physiquement pertinentes ne sont pas connues. Elles sont sans doute mesurées ou reconstituables à partir des nombreuses mesures, mais vraisemblablement noyées dans une masse de données redondantes d'intérêt moindre.
- les échelles temporelles ne sont pas connues, et la fréquence d'acquisition des séries temporelles n'a que peu de chance de correspondre aux constantes de temps du système (constantes de temps qui pourraient d'ailleurs fortement varier d'une variable à une autre).
- des phénomènes physiques complexes (chocs, micro-fissuration, changement de phases, ...) peuvent se produire pendant la transformation. La modélisation de ces phénomènes fait intervenir des dynamiques peu régulières (a minima : non linéarité, vraisemblablement : défaut de continuité, apparition d'hysteresis). L'identification d'un modèle linéaire par des techniques standard risque d'être insuffisante.
- les défaillances (y compris temporaires) de capteurs sont inévitables au vu de la durée des périodes étudiées et de l'environnement parfois hostile auquel ils sont confrontés. Évaluer la qualité des données, et les sélectionner en conséquence, sera un préalable indispensable à la modélisation.

2.2 Problème de commande

La difficulté de la recherche de lois de commande augmente avec la complexité du modèle. On aura donc sans doute intérêt à privilégier les modèles pas forcément les plus réalistes, mais sur lesquels on pourra obtenir les meilleurs résultats. Si il devient nécessaire d'introduire des dynamiques très irrégulières, il faudra s'assurer que l'on garde une certaine maîtrise de ces irrégularités (par exemple en les localisant).

2.3 Verrous scientifiques

Une littérature abondante est disponible pour résoudre l'essentiel des difficultés évoquées précédemment pour peu que les autres aient été résolues. Par exemple, un modèle de dynamique étant connue, on sait en général extraire les variables les plus pertinentes pour proposer une dynamique très proche mais beaucoup plus simple (réduction de modèle). De la même façon, la forme d'un modèle étant connue (liste des variables et formes analytique de la dynamique), des techniques très efficaces sont disponibles pour l'identification de dynamique à partir de données fiables; la génération de stratégies de commande efficaces ou la détection de capteurs défaillants est envisageable de manière automatique dès que la dynamique est connue.

La grande difficulté viendra de la nécessité de résoudre simultanément tous ces problèmes.

2.4 Méthodes envisagées et programme de travail

Une méthode en plusieurs phases est envisagée, incluant une première phase d'élaboration du modèle (basée sur des outils statistiques), suivi d'une phase d'analyse de la contrôlabilité des modèles obtenus (basée sur des méthodes de contrôle géométrique). Il conviendra d'examiner la pertinence d'une méthode itérative où on alternerait les phases 1 et 2.

2.5 Complémentarité de méthodes

Il sera crucial que chacune des étapes soit exécutée en gardant en vue le but global du projet. Par exemple, il serait contre-productif de proposer des modèles à la dynamique tellement complexe ou irrégulière qu'elle en devient difficilement simulable numériquement, ou de proposer au final des lois de commande insuffisamment robustes aux erreurs de modèle, ou faisant intervenir des commandes physiquement irréalistes. Pour ces différents critères (forme possible des pathologies du modèle, lois de commande physiquement admissibles) il semble difficile de se passer de l'avis d'un expert humain.

3 Programme scientifique

3.1 Formalisation mathématique

On s'intéresse à la transformation d'une pièce entre les instants 0 et T .

L'état d'une pièce à un instant t est noté par x_t , un point de \mathbf{R}^{n_x} . (Le vecteur x_t recense les différents paramètres pertinents de la pièce, par exemple des dimensions, un état de surface, un profil de températures,...) La pièce arrive dans un état x_0 , qui est soit mesuré soit supposé égal à un nominal connu. En général, on ne disposera que de x_0 et de x_T (pas de connaissance complète de la pièce aux instants intermédiaires $0 < t < T$). La pièce terminée sera acceptable si $x_T \in V$ où V est un sous-ensemble donné de \mathbf{R}^{n_x} .

Un certain nombre de paramètres physiques sont mesurés au cours de la transformation. On regroupe ces paramètres au temps t dans un vecteur y_t de \mathbf{R}^{n_y} (qui peut contenir par exemple des températures, des forces, des pressions, ...). On n'a pas accès directement à y_t , mais juste à des mesures $\hat{y}_t = y_t + B_1(\omega, t) + B_2(\omega, t)$ où les bruits aléatoires B_1 et B_2 sont de nature différente. B_1 est un bruit de mesure standard, de moyenne nulle ou presque et à faible variance, qui correspond au fonctionnement standard d'un bon capteur. B_2 modélise un bruit très élevé mais à très faible probabilité, qui correspond au comportement défectueux d'un capteur. En général, on disposera de nombreuses mesures $y = (\hat{y}_{t_1}, \hat{y}_{t_2}, \dots, \hat{y}_T)$ collectées tout au long de la transformation.

Les valeurs des contrôle au temps t sont regroupées dans un vecteur u_t de \mathbf{R}^{n_u} . u_t peut contenir aussi bien des variables discrètes qui restent constantes tout au long de la transformation (modèle d'outil utilisé, nombre de brûleurs dans un four, présence ou non de tel accessoire) que des variables discrètes (état marche/arrêt d'un brûleur,...) ou continues (débit de gaz, force exercée par un outil,...) susceptibles de varier pendant la transformation. L'ensemble de ces contrôles, aux différents instants est regroupé dans un vecteur $u = (u_{t_1}, \dots, u_T)$.

On dispose d'une grande collection d'enregistrements (x^i, \hat{y}^i, u^i) , pour i dans $1, \dots, N$. Le but est d'utiliser ces données pour en

1. déduire une application F telle que $x_T = F(x_0, y, u)$.
2. F étant connue, en déduire une application G qui à chaque instant retourne le contrôle optimal u_t en fonction de x_0 et y_t : $u_t = G(x_0, y_t)$.

3.2 Programme prévisionnel de travail

Dans un objectif d'efficacité, on pourra (au moins en début de thèse) s'appuyer sur une application pratique (trois suggestions sont données dans les sections suivantes) pour laquelle on dispose à la fois de données réelles et de la possibilité, au besoin, de faire appel à une expertise humaine pour vérifier/ infirmer des hypothèses de modélisation. Concrètement, le programme de travail des premiers mois pourrait ressembler à :

1. **Réduction de données** disponibles pour en extraire une liste de quelques dizaines de variables utiles, identifier leur plage de fonctionnement, distinguer la mesure de paramètres physiques de consignes d'actionneurs. Ce dernier point demandera un recours à l'expertise humaine. Éventuellement, on pourra séparer à ce stade les mesures des signaux de défaillance.
2. **Agrégation de données : critère physique.** À l'aide d'algorithmes de type agrégation à centre mobile (de type K-means, éventuellement adaptés pour tenir compte de la masse de données disponibles), regrouper les données nettoyées par paquets P_1, P_2, \dots où les paramètres physiques mesurés \hat{y} sont proches. L'objectif est de ne comparer des systèmes qu'à conditions expérimentales voisines de manière à mieux distinguer l'effet des contrôles.
3. **Agrégation de données : type de contrôle** Diviser les paquets de données rassemblés dans l'étape 2 en sous-paquets

$$\underbrace{P_{1,1}, P_{1,2}, \dots}_{P_1}, \underbrace{P_{2,1}, P_{2,2}, \dots}_{P_2}, \dots$$

correspondant à des modes de contrôle différents (grandeurs discrètes constantes sur u).

4. **Identification par sous-paquet** Sur le paquet $P_{i,j}$, on approchera donc $F(x_0, y, u)$ par $A_{ij}x_0 + B_{ij}\hat{y} + C_{ij}u$. Pour cela, on va procéder à une identification de dynamique linéaire pour chacun des sous-paquets $P_{i,j}$ obtenus à l'étape 4. Cela revient à trouver des applications linéaires A, B, C qui minimisent une fonctionnelle d'erreur du type $\sum_{l \in P_{i,j}} \|x_T^l - A_{ij}\hat{y}^l - B_{ij}u^l\|$ ou $\sum_{l \in P_{i,j}} \|x_T^l - A_{ij}\hat{y}^l - B_{ij}u^l\|^2$. La taille des données risque d'imposer une méthode d'optimisation stochastique. Une question intéressante est de savoir si la détection des mesures par des capteurs défaillants peut avoir lieu à ce moment. La réponse est probablement positive avec une méthode d'optimisation adaptée.
5. **Ajustement de dynamique** Si la dynamique obtenue au point 4 n'est pas satisfaisante, il sera nécessaire de scinder le sous-paquet P_{ij} en sous-sous paquets ou complexifier le modèle et réitérer le cycle identification/complexification/scission de paquets jusqu'à obtenir une dynamique satisfaisante.
6. **Détermination de la commande** Rechercher, au moins numériquement, des lois de commande pour le modèle identifié au point 5. Il sera à ce stade possible de prendre en compte des critères à minimiser signalés par un expert humain.

3.3 Prolongements scientifiques

Le programme scientifique annoncé dans le paragraphe précédent a pour but de fournir une ligne directrice pour garantir l'obtention d'un résultat à l'issue de la thèse. Chaque étape peut être abordée naïvement, par exemple en s'appuyant sur [5], [2] et [3] pour les points 1, 2, et 3 puis sur [3], [4] pour les points 4, 5 et [1] pour le point 6. Une première question importante sera de déterminer quelle combinaison de méthodes usuelles donne les meilleurs résultats.

Par la suite, toutes les étapes du programme précédent peuvent donner à des investigations plus poussées. Citons par exemple de plus faibles hypothèses sur la régularité du modèle ou du bruit, la recherche d'algorithmes éventuellement moins précis mais plus rapides pour une première analyse, l'optimisation pour un critère donné, l'amélioration de la robustesse des lois de commande... Le choix des questions étudiées se fera en fonction de l'intérêt industriel et des goûts et compétences du doctorant.

4 Application 1 : commande d'un plateau verrier

4.1 Existant

Une certaine chaîne de production comporte, entre autres, un four où sont chauffés des plateaux de verre. Dans l'idéal, tous les plateaux entrant dans le four ont les mêmes caractéristiques, et tous les plateaux ont en sortie de four le même historique thermique. Dans la réalité, les caractéristiques des plateaux en entrée varient légèrement (composition, profil de température, positionnement dans le four, ..), et la commande de régulation de puissance du four est effectuée avec un correcteur PID prenant en entrée la température de l'air dans le four. En conséquence, l'injection de puissance est très importante lors du démarrage du four, et plus faible quand le four est chaud. De plus, le correcteur PID est paramétré pour optimiser différents critères (durée de chauffe, coût énergétique, ..) pas forcément compatibles avec l'exigence d'uniformité d'historique thermique des plateaux.

4.2 Problème posé

Certaines mesures physiques sont effectuées en continu, d'autres (homogénéité thermique du plateau) uniquement en sortie de four. Est-il possible d'utiliser cette masse de données, après une éventuelle préparation, pour

1. construire automatiquement, ou en tout cas avec le minimum d'intervention humaine, un modèle (faisant intervenir des variables pertinentes) de plateau dans le four ?
2. détecter en temps réel la défaillance d'un capteur ?
3. proposer une loi de commande de la puissance du four meilleure que le PID actuellement utilisé ?

5 Application 2 : contrôle qualité pour la préparation de pare-brise

La deuxième application concerne une étape de fabrication de pare-brise, où les bords de plaques de verre planes sont meulés. Le contrôle est la force exercée sur la meule et d'autres paramètres. L'objectif est d'obtenir la plaque la plus uniforme possible.

6 Application 3 : contrôle en temps continu d'un dispositif d'injection plastique

La troisième application consiste dans le contrôle de vis utilisée pour l'injection plastique (la vis pousse un granulat chauffé autour de 250 degrés dans un moule). Les pièces produites (huit par seconde) sont vérifiées optiquement individuellement. Actuellement, un agent humain vérifie le fonctionnement du dispositif toutes les huit heures et y apporte des corrections (température, pression, changement des outils) en fonction de son expertise. L'objectif est d'arriver à une adaptation en continu des différents paramètres physiques au vu des mesures réalisées en sortie sur les pièces précédentes.

Références

- [1] C. BRIAT, *Linear Parameter-Varying and Time-Delay Systems*, Springer Verlag, 2015.
- [2] R. O. DUDA, P. E. HART, AND D. G. STORK, *Pattern Classification, 2nd Edition*, Wiley, 2000.
- [3] T. HASTIE, R. TIBSHIRANI, AND J. FRIEDMAN, *The elements of statistical Learning data mining, inference and prediction*, Springer, 2009.
- [4] R. HORST AND P. M. PARDALOS(EDS.), *Handbook of Global Optimization, Volume 1*, Springer, 1995.
- [5] G. SAPORTA, *Probabilités, analyse des données et statistique*, Editions Technip, 2006.