

# Analyse de l'accroissement de l'erreur d'estimation d'indicateurs de monitoring en fonction de la distance au domaine d'apprentissage.

- Application à la surveillance du démarrage d'un moteur d'avion -

A. AUSLOOS<sup>1</sup>, P. BEAUSEROY<sup>2</sup>, E. GRALL-MAËS<sup>2</sup>, A. GRALL<sup>2</sup>, J.R. MASSE<sup>1</sup>

<sup>1</sup>SNECMA, Etablissement de Villaroche, Rond Point René Ravaud, 77550 Moissy-Cramayel.

<sup>2</sup>Institut Charles Delaunay, Université de Technologie de Troyes, 12 rue Marie Curie, BP 2060, 10010 Troyes.

{alexandre.ausloos, jean.remi.masse}@snecma.fr, {pierre.beuseroy, antoine.grall, edith.grall}@utt.fr

**Résumé** – Ce papier présente une analyse des erreurs d'estimation des indicateurs de monitoring appliquée la séquence de démarrage d'un turboréacteur. L'approche de surveillance mise en place est basée sur l'étude des résidus entre la mesure des indicateurs et leurs estimations par des modèles de régression appris sur une base de données. En dehors de la région couverte par la base de données, la validité des modèles n'est plus garantie ce qui limite l'applicabilité du système de monitoring. Les connaissances sur le comportement des démarrages laissent supposer que l'erreur d'estimation n'augmente que modérément autour de l'espace d'apprentissage. Ce papier étudie donc l'évolution des résidus en fonction de la distance à la zone d'apprentissage des modèles. L'objectif est de statuer sur la possibilité de l'élargissement de l'espace de validité des modèles sans dégrader les performances du processus de détection.

**Abstract** – This paper deals with the estimation errors of monitoring indicators for a turbofan start sequence. The proposed monitoring approach is based on the analysis of the residuals between the indicators measurements and their estimations by regression models built on a database. Outside the area covered by the database, the models validity is not guaranteed, which limits the monitoring system applicability. The knowledge on the start behavior allows supposing that the estimation errors increase not so much around the learning space. Therefore, this paper studies the evolution of these residuals in function of the distance to the model learning area. The aim is to work out the possibility to expand the model validity area without degrading the detection process performances.

## 1. Introduction

Les défaillances du système de démarrage d'un avion surviennent à un moment critique pour les compagnies, à savoir quand les passagers sont à bord et que l'avion doit partir. Ces problèmes de démarrage provoquent en général des retards et, dans les pires cas, des annulations de vols. La mise en évidence précoce de la dégradation permet la mise en place de maintenance préventive. Les bénéfices attendus sont donc une réduction du nombre de retards et annulations de vols, une augmentation de la satisfaction des clients et une diminution des coûts opérationnels.

Une étude de la phase de démarrage et de ses défaillances a permis la définition d'une série d'indicateurs représentatifs de l'état de santé des composants participant au démarrage du moteur [3]. La détection d'une détérioration est basée sur le résidu de chaque indicateur entre la grandeur mesurée et la valeur nominale théorique. Néanmoins, la valeur de chacun des indicateurs est sensible aux conditions opératoires du démarrage et il est impératif de tenir compte des paramètres environnementaux. La valeur nominale théorique de

chaque indicateur doit donc être estimée en fonction de mesures complémentaires représentatives des conditions opératoires. Une solution proposée dans [7] consiste à estimer cette valeur nominale à l'aide d'un modèle d'estimation par régression dont les paramètres sont appris sur l'ensemble des données d'apprentissage disponible. Cet ensemble ne couvre qu'une partie des conditions opératoires possibles. Par conséquent, lorsque les conditions opératoires sont en dehors de cet ensemble, la validité du modèle peut être remise en cause, ce qui limite le domaine d'applicabilité de la méthode de monitoring. Toutefois les connaissances sur le comportement des démarrages permettent de faire l'hypothèse que l'influence des paramètres associés aux conditions opératoires sur les valeurs réelles des indicateurs se fait de façon continue. Ainsi, la déviation par rapport au modèle appris est probablement continue et il est possible d'envisager d'étendre un peu le domaine d'applicabilité des modèles appris. Une étude de l'erreur du modèle d'estimation en fonction de la distance entre les conditions opératoires et celles du domaine d'apprentissage a été menée. L'intérêt et

certaines limites de l'approche utilisée ont été mis en évidence à l'aide d'une étude expérimentale. La démarche considérée et les résultats obtenus sont rapportés dans cette communication.

En section 2 le fonctionnement d'un démarrage pour mieux appréhender sa surveillance est décrit. Une approche générale de monitoring tenant compte des contraintes liées à l'influence des conditions opératoires et au domaine d'apprentissage incomplet est présentée en section 3. La section 4 donne une brève description des modèles d'estimation et de leurs performances. L'étude des performances des modèles en fonction de l'éloignement des conditions opératoires par rapport au domaine d'apprentissage des modèles d'estimation est exposée en section 5. Une conclusion fait l'objet de la section 5.

## 2. Présentation du problème de l'application

Une brève description du fonctionnement d'un turboréacteur en phase de démarrage est donnée dans la partie 2.1. Puis les pannes les plus fréquentes qui empêchent le démarrage du moteur sont présentées dans la partie suivante et les indicateurs pour détecter et localiser ces pannes sont définis dans la partie 2.3.

### 2.1 Description de la phase de démarrage

L'opération de démarrage d'un moteur d'avion fait intervenir une séquence qui nécessite la participation de plusieurs systèmes sous la responsabilité d'opérateurs différents :

- L'avionneur (Boeing, Airbus).
- Le motoriste (Snecma, General Electric, Pratt & Whitney, Rolls-Royce).
- Les équipementiers (Hispano-Suiza, ...).

Par conséquent la connaissance sur les composants participant au démarrage est dispersée entre plusieurs sociétés ce qui rend parfois difficile la récolte d'informations. Les principales étapes d'un démarrage standard sont listées ci dessous ainsi que le système mis en oeuvre:

- Alimentation des canalisations en air (système d'air),
- Ouverture de la vanne à air (système de démarrage),
- Enroulement par le démarreur du compresseur haute pression, noté CoHP (système moteur),
- Injection de carburant quand le CoHP dépasse un certain régime (système carburant),
- Allumage dans la chambre de combustion par les bougies (système d'allumage),
- Vérification de l'évolution normale des performances via les capteurs (action du pilote).
- Fermeture automatique de la vanne à air au régime d'auto rotation, dit Cut Out (système de régulation),
- Atteinte du ralenti du moteur et vérification de la stabilisation des signaux des capteurs (action du pilote)

La figure 1 illustre l'évolution, lors d'un démarrage, de la température des gaz d'échappements (EGT) ainsi que le régime du CoHP, de 0 tour par minute jusqu'au ralenti.

La définition d'un processus de monitoring de la phase de démarrage exige la connaissance des pannes à surveiller. Le paragraphe suivant présente donc les principales causes des non démarrages.

### 2.2 Principales défaillances

Grâce aux experts de Snecma, au retour d'expérience en exploitation et aux publications [1, 2], il a été possible d'identifier les causes les plus fréquentes des non démarrages. Elles sont synthétisées dans le tableau 1 :

TAB. 1 : Causes principales de non démarrage.

Composants		Modes de défaillance
C1	Hydro Mechanical Unit (HMU)	Problèmes de débit carburant
C2	Vanne à air	Problèmes d'ouverture ou de fermeture
C3	Démarreur	Rupture d'arbre, problème d'embrayage, fuite d'huile, usure
C4	Bougies d'allumage	Pas d'étincelle
C5	Injecteurs	Colmatés
C6	Compresseur d'air	Pression faible
C7	Pompe à carburant principale	Perte de rendement

Les effets de ces défaillances sur le comportement du moteur ont été explorés dans le but de définir des indicateurs significatifs de l'état de santé du démarrage. Le paragraphe suivant présente les effets de trois pannes pour expliquer le choix des indicateurs ainsi que la définition des signatures de pannes qui serviront à la localisation.

### 2.3 Indicateurs et localisation

Un problème de débit carburant provoqué par le HMU (C1 dans tab. 1) peut perturber la phase d'allumage, l'accélération du moteur après l'allumage et la température des gaz d'échappement (EGT). Ce type de défaillance a donc des conséquences sur le temps d'allumage, le temps après l'allumage ( $t_2$  &  $t_3$  sur la figure 1) et la vitesse d'évolution de l'EGT (Gradient EGT).

Une vanne à air reste ouverte (C2 dans le tab. 1) peut raccourcir le temps de la dernière phase de démarrage. La durée  $t_3$  sera donc plus courte que d'habitude.

Un démarreur usé (C3 dans le tab. 1) fournit moins de couple au moteur. Donc l'accélération sera moins élevée que la normale, particulièrement pendant les premières secondes. Cette dégradation a donc des conséquences sur les coordonnées du point d'accélération maximale, la durée  $t_1$  et le gradient EGT.

Ce travail a été effectué pour toutes les défaillances du tableau 1. Ainsi il a été possible de définir les indicateurs les plus appropriés pour surveiller la capacité à démarrer.

Une liste de 8 indicateurs, notés de I1 à I8, a été sélectionnée :

- Coordonnées du point d'accélération maximale du compresseur haute pression (I1, I2),
- La durée de trois phases spécifiques, notées t1, t2, t3 sur la figure 1 (I3, I7, I8),
- Le temps d'allumage (I4),
- Le gradient moyen et maximal de la température des gaz d'échappement (I5, I6).

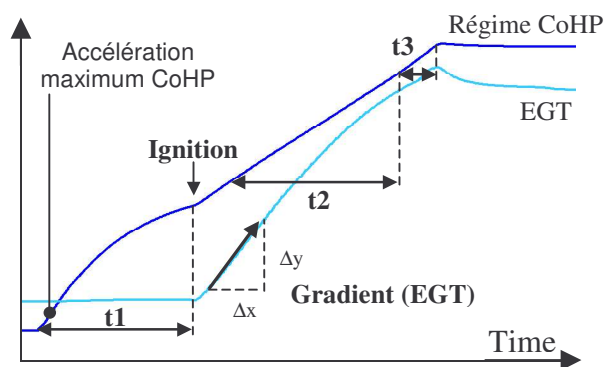


FIG. 1: Evolution du régime du compresseur haute pression (CoHP) et de la température des gaz d'échappement pendant un démarrage.

Les effets des défaillances ou des prémisses des dégradations ont été caractérisés individuellement. Pour chaque défaillance, le sens de variation de chacun des indicateurs a été déterminé : stable, en diminution ou en augmentation, et l'importance de la modification a été qualifiée de faible, moyenne ou grande.

Ces impacts ont été codés suivant la convention suivante : 0 signifie que l'indicateur n'est pas impacté par cette défaillance, une valeur positive (respectivement négative) signifie que l'indicateur augmente (respectivement diminue) et l'importance de la modification est décrite par les chiffres 1, 2 et 3. Une matrice d'identification des défaillances a ainsi été construite [3].

TAB. 2. Effets des défaillances de C1, C2, C3.

	I1	I2	I3	I4	I5	I6	I7	I8
C1	0	0	0	-1	+3	+3	-2	-1
C2	0	0	0	0	0	0	0	-1
C3	+2	-2	+1	0	-1	-1	+1	0

### 3. Méthode pour le Monitoring

Les indicateurs proposés pour surveiller la séquence de démarrage, en plus d'être sensible aux dégradations, sont sensibles aux conditions opératoires comme le détaille le paragraphe 3.1. Pour étudier l'influence de ces conditions et proposer une solution qui permette de tenir compte quantitativement des différentes influences, une base de données est disponible. Le paragraphe 3.2 présente quelques limites inhérentes de cette base de données. Enfin, le dernier paragraphe propose une approche

générale pour la surveillance de la capacité à démarrer du moteur.

#### 3.1 Influence des conditions opératoires

La difficulté de surveiller la séquence de démarrage avec les indicateurs vient du fait qu'ils ne dépendent pas seulement des dégradations mais également des conditions opératoires telles que la température extérieure, la pression atmosphérique ou encore la température d'huile.

Par exemple, les couples résistants du moteur augmentent quand l'huile refroidie. Donc un démarrage réalisé avec de l'huile froide sera plus lent qu'un démarrage effectué avec de l'huile chaude. Sans tenir compte de cette température, il est difficile de savoir si un temps de démarrage anormalement long a été observé à cause d'une dégradation ou d'une température d'huile froide. La figure 2 montre le temps de la première phase de démarrage, c'est à dire l'indicateur t1, en fonction de la température d'huile pour des démarrages sains. Le lien entre ces deux grandeurs est linéaire jusqu'à une certaine valeur puis constant. L'importante dispersion observée est due à influence des autres conditions opératoires.

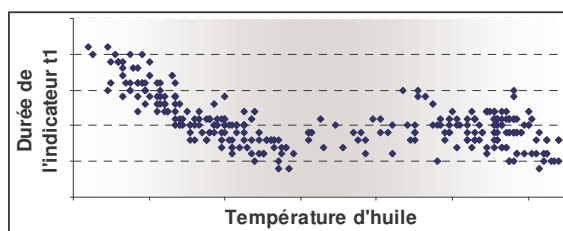


FIG. 2: Indicateur t1 en fonction de la température d'huile pour des démarrages sains.

Des essais sur des bancs avec un moteur ont montré que l'influence des conditions opératoires est importante, et a un impact qui n'est pas négligeable sur les indicateurs. Donc les conditions opératoires peuvent masquer l'effet des dégradations si elles ne sont pas prises en considération.

Une approche permettant de déterminer des indicateurs indépendants des conditions opératoires a été proposée.

#### 3.2 Limite de la base de données

Les données issues de démarrages réalisés avec les bancs d'essais de Snecma à Villaroche ont été utilisées pour apprendre des modèles prédictifs pour les différents indicateurs [5].

La base de données est constituée des mesures des indicateurs et des conditions opératoires réalisées avec 2 moteurs pour un total de 300 démarrages. Cependant les conditions opératoires associées à ces essais n'ont pas varié autant que cela est possible en exploitation. En particulier, aucune donnée n'est disponible pour des températures extérieures extrêmes, la pression atmosphérique a été presque constante et la pression d'air envoyée au démarreur a très peu changé.

L'espace dans lequel se répartissent les données d'entrée ne couvre donc qu'un faible pourcentage de la totalité de l'espace de variation possible.

L'approche de monitoring proposée concilie les deux contraintes décrites en 3.1 et 3.2 à savoir, tenir compte de l'influence des conditions extérieures et les limites de la couverture du domaine de variation des conditions opératoires.

### 3.3 Solution générale

La solution proposée est basée sur l'analyse du résidu entre la valeur de l'indicateur mesuré et la valeur nominale estimée à partir des données relatives aux conditions opératoires.

En considérant que les effets des dégradations et des conditions opératoires sont additifs, il est possible d'écrire que la valeur d'un indicateur est la somme des effets de la dégradation  $d$ , des conditions opératoires caractérisées par une fonction  $g$  de  $K$  paramètres  $op_k$  et d'une variable caractérisant l'erreur  $e$  :

$$I = d + g(op_1, op_2, \dots, op_K) + e \quad (1)$$

L'erreur peut provenir du bruit de la chaîne d'acquisition, de la non prise en compte d'autres paramètres dans la fonction  $g$  tels que le type de carburant, ou même du fait que l'indicateur n'est pas totalement explicable par les conditions opératoires. L'information recherchée étant la valeur de  $d$ , il est proposé d'estimer  $g(op_k) + e$  pour retrouver l'information pertinente portée par  $d$  en calculant pour chaque indicateur la différence, appelée résidu, entre la mesure et l'estimation de l'indicateur noté  $\hat{I}$  [4] :

$$r = I - \hat{I} \quad (2)$$

Pour un démarrage sain, il n'y a pas d'effet de dégradation donc la valeur du résidu n'est influencée que par l'imprécision du modèle d'estimation. Le processus de détection se basera donc sur les valeurs des résidus observés et des résidus obtenus sur la base d'apprentissage.

Les performances de la prédiction ne sont connues que pour des conditions opératoires similaires à celles de l'ensemble d'apprentissage, définissant un domaine de validité. En dehors de ce domaine la probabilité de fausse alarme n'est plus maîtrisée ce qui est intolérable pour les compagnies aériennes. Par conséquent la prédiction des indicateurs et la détection d'une éventuelle dégradation ne peut se faire que dans le domaine de validité.

Le système de monitoring proposé consiste alors à déterminer si les conditions opératoires du démarrage du moteur sont incluses dans le domaine couvert par la base de données d'apprentissage. Si c'est le cas, les 8 indicateurs sont estimés et l'algorithme utilise les 8 résidus calculés pour prendre une décision sur une éventuelle dégradation. Sinon, aucun avis n'est donné et les données sont traitées en accord avec une stratégie d'apprentissage

permettant une future mise à jour de l'espace couvert par l'algorithme. La figure 3 illustre ce principe général :

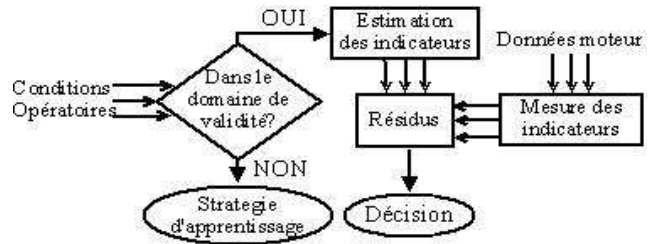


FIG. 3: Solution générale pour le Monitoring.

Le domaine de variation des conditions opératoires où l'algorithme permet de prendre une décision dépend donc de l'ensemble d'apprentissage. Il peut être intéressant d'élargir ce domaine tout en garantissant la maîtrise de l'erreur d'estimation. L'étude de l'évolution de l'erreur d'estimation lorsque les conditions s'éloignent de la zone couverte par les données d'apprentissage, présentée en section 5, repose sur les modèles d'estimation présentés en section 4.

## 4. Modèles d'estimation

Cette partie présente la méthode de régression étudiée, le critère de performance pour construire les modèles de prédiction des indicateurs ainsi que les performances obtenues pour chacun d'entre eux.

### 4.1 Principes de la méthode de régressions

Les méthodes à vecteurs support ont émergé des travaux de Vapnik et Chervonenkis dans les années 70. Il s'agit d'une méthode non linéaire et non paramétrique. L'idée est de déployer les données dans un espace de dimension élevée à l'aide d'une fonction  $\Phi$  de manière à pouvoir trouver une régression linéaire dans ce nouvel espace. La formulation du problème est :

$$\text{Minimiser } \|w\|^2 + C \sum_{i=1}^N (\xi_i^2 + \xi_i^{*2}) \quad (3)$$

$$\text{avec } \begin{cases} (\langle w, \Phi(x_i) \rangle + b) - y_i \leq \varepsilon + \xi_i \\ y_i - (\langle w, \Phi(x_i) \rangle + b) \leq \varepsilon + \xi_i^* \end{cases}$$

Cela consiste à déterminer l'hyperplan de la régression de sorte que les valeurs estimées  $\langle w, \Phi(x_i) \rangle + b$  ne s'éloignent pas de la valeur cible  $y_i$  au delà d'une tolérance  $\varepsilon$  prédéfinie. Le modèle peut sortir de cet intervalle de tolérance avec pour conséquence une valeur  $\xi_i$  ou  $\xi_i^*$  non nulle. Le paramètre  $C$  permet de gérer le compromis entre la complexité du modèle et le respect de cette contrainte.

Il a été montré que ce problème peut se mettre sous la forme d'un problème d'optimisation d'une fonction quadratique qui ne dépend que des produits scalaires entre les observations. Sous les conditions de Mercer, ces produits scalaires peuvent directement être calculés à



l'aide de fonctions noyaux et ne nécessitent pas l'expression analytique de la fonction de transformation  $\Phi$ .

Un modèle de régression par vecteurs support nécessite donc de déterminer les paramètres  $\varepsilon$  et  $C$  et de choisir la fonction noyau appropriée. Notons que cette fonction dépend généralement d'un paramètre  $h$ .

Le passage de l'espace d'origine à l'hyperplan est réalisé à l'aide d'un noyau. Par exemple, pour un noyau polynomial, il s'agit de l'ordre du modèle, pour un noyau gaussien, il s'agit de l'écart type. [6].

## 4.2 Le critère de performance

Un modèle d'estimation a été déterminé pour chaque indicateur. Il est obtenu à l'aide d'une méthode de régression à vecteurs support dont les paramètres sont optimisés pour minimiser un critère de performance en l'absence de défaillance. Ce critère est l'erreur quadratique moyenne (EQM) entre la mesure de l'indicateur  $I$  et son estimation calculée sur les  $N$  données de l'ensemble considéré :

$$EQM = \frac{1}{N} \sum_{n=1}^N (I_n - \hat{I}_n)^2 = \frac{1}{N} \sum_{n=1}^N (r_n)^2 \quad (4)$$

Le critère de performance à minimiser étant l'erreur quadratique, les données ont été normalisées de manière à ne pas optimiser les modèles sur les variables qui ont la plus grande variance.

Ensuite la base de données a été divisée en 2 parties égales, l'une servant à apprendre un modèle et l'autre servant à tester sa capacité de généralisation. Le modèle doit être assez complexe pour s'adapter au mieux aux données tout en conservant le plus possible une bonne capacité de généralisation. En effet, l'augmentation de la complexité du modèle sur la base d'apprentissage conduit à diminuer l'EQM puisque le modèle s'ajuste de mieux en mieux aux données. Mais ce n'est pas forcément le cas sur les données qui servent au test. Dès que le modèle sur-apprend les données, il généralise moins bien et l'erreur d'estimation sur des points qui n'ont pas servi à l'apprentissage augmente comme l'illustre la figure 4 :



FIG. 4: Evolution de l'EQM selon l'ordre du modèle sur la base d'apprentissage (rond) et de test (triangle)

Les modèles sont donc construits sur la base d'apprentissage de manière à minimiser l'EQM sur cette base. Le modèle retenu est celui qui a l'EQM la plus petite sur la base de test. Pour l'exemple illustré par la figure 4,

l'ordre du modèle le mieux adapté est le 2<sup>nd</sup>. La recherche du meilleur paramétrage de  $\varepsilon$ ,  $C$ , le type de noyau et la variable associée  $h$  pour la régression par support vecteur a donc été effectuée.

## 4.3 Erreur d'estimation des modèles

Ce paragraphe va présenter, par indicateur, la moyenne et l'écart type des erreurs quadratiques obtenus sur la base de test pour 4 paramétrages différents. Les valeurs de  $C$  et  $\varepsilon$  sont les mêmes et valent respectivement 10 et 0,05. Le choix du paramétrage et des résultats complémentaires sont dans [7].

TAB. 3 : EQM par indicateur selon le paramétrage.

Type de noyau →	Polynomial		Gaussien	
Indicateur ↓	h= 1	h= 2	h= 4	h= 10
1	1,04	1,07	1,06	1,05
2	0,36	0,36	0,34	0,34
3	<i>0,14</i>	<i>0,07</i>	0,08	<i>0,10</i>
4	0,12	0,11	0,12	0,13
5	<i>0,30</i>	0,11	0,11	<i>0,22</i>
6	<i>0,30</i>	<i>0,07</i>	<i>0,07</i>	<i>0,19</i>
7	<i>0,32</i>	0,16	0,16	<i>0,27</i>
8	<i>0,54</i>	0,32	0,32	<i>0,49</i>

TAB. 4 : Ecart type des erreurs quadratiques par indicateur selon le paramétrage.

Type de noyau →	Polynomial		Gaussien	
Indicateur ↓	h= 1	h= 2	h= 4	h= 10
1	2,56	2,67	2,66	2,56
2	0,54	0,57	0,53	0,52
3	<i>0,23</i>	0,12	0,14	<i>0,18</i>
4	0,31	0,29	0,30	0,31
5	<i>0,42</i>	0,22	0,19	<i>0,29</i>
6	<i>0,46</i>	0,13	0,10	<i>0,24</i>
7	<i>0,64</i>	0,34	0,38	<i>0,61</i>
8	<i>1,03</i>	0,65	0,65	<i>1,00</i>

Comme le montre les tableaux 3 et 4 tous ces paramétrages ne fournissent pas des performances équivalentes pour un même indicateur. Les paramétrages les moins performants sont mis en italiques et ne seront plus étudiés pour la suite de l'étude.

## 5. Evolution de l'erreur

### 5.1 Démarche

La régression est une méthode qui permet de trouver la fonction qui respecte le mieux les données apprises. La relation ainsi obtenue est valable à l'intérieur de l'espace d'apprentissage. En dehors de cet espace, il n'existe pas de relation mathématique entre l'erreur commise par le modèle de régression et la distance à l'espace d'apprentissage. Comme l'influence des conditions opératoires évolue lentement, l'hypothèse que l'augmentation de l'erreur restera modérée dans une région restreinte est faite. Il est donc proposé ici d'étudier

l'accroissement de l'erreur d'estimation. Or, aucune donnée supplémentaire en dehors de la zone d'apprentissage n'est disponible pour réaliser cette étude. L'approche retenue consiste donc à volontairement restreindre le domaine d'apprentissage à une fraction de l'ensemble des données disponibles et de calculer l'EQM commise sur cet ensemble. La fraction choisie est un demi de manière à avoir suffisamment de points à l'extérieur du domaine. Ensuite les points à l'extérieur sont regroupés selon leur distance au domaine et l'EQM de chaque groupe est calculée. Cette étude donnera donc une information sur l'augmentation de l'erreur en dehors de la région d'apprentissage.

## 5.2 Définition du domaine d'apprentissage restreint

### 5.2.1 Définition de l'espace d'apprentissage restreint

Dans l'espace des conditions opératoires normalisées, la méthode des One class SVM [8] a été utilisée pour définir l'enveloppe de la région la plus dense contenant la moitié des essais disponibles. La figure 5 illustre cet espace en projetant les K variables représentatives des conditions opératoires dans le plan des 2 premiers vecteurs propres.

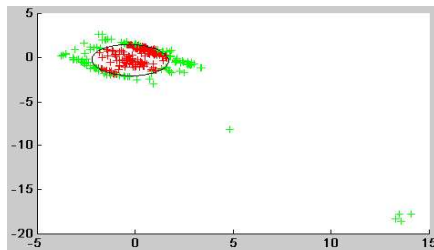


FIG 5 : Projection de la base de données sur ses 2 vecteurs propres.

Les quelques points excentrés en bas à droite de la figure représentent des démarrages pour lesquels la valeur d'une variable caractéristique des conditions opératoires est très différente des valeurs obtenues pour la plupart des démarrages. Il s'agit de la pression d'alimentation d'air qui est envoyée au démarreur pendant le démarrage du moteur. Elle est de l'ordre de 3,5 bars pour les points excentrés alors qu'elle est comprise entre 2,1 et 2,3 bars pour la plupart des démarrages. Ces points singuliers n'ont pas été utilisés pour définir l'écart type qui a servi à la normalisation de l'ensemble de la base de données.

La figure 6 constitue un zoom sur la partie la plus dense des points :

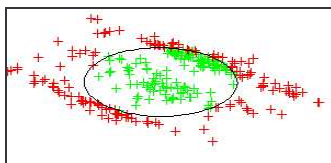


FIG. 6 : Zoom sur les points sélectionnés par les One SVM ayant servi à l'apprentissage.

### 5.2.2 Performance des modèles sur le domaine restreint

Le but de l'étude étant d'estimer l'accroissement de l'erreur quadratique moyenne, il est nécessaire d'avoir une référence, qui peut être obtenue en déterminant la moyenne, notée  $EQM_0$ , et l'écart type des écarts quadratiques à l'intérieur de l'espace d'apprentissage.

Pour les points situés en dehors de la zone d'apprentissage, l'EQM calculée sera issue exclusivement des données test, c'est à dire de points qui n'ont pas servi à l'apprentissage. Il est essentiel pour la cohérence des résultats de calculer l'EQM de l'espace d'apprentissage avec des données test également. Pour cela, la méthode de validation croisée a été utilisée. La base d'apprentissage définie au §5.1.1 a été séparée en 10 ensembles, 9 sont utilisés pour l'apprentissage du modèle et 1 pour le test. L'opération est répétée 10 fois en changeant à chaque fois l'ensemble utilisé pour le test. L'EQM sur la base d'apprentissage est donc la moyenne de tous les écarts quadratiques obtenus sur les 10 ensembles de test. De même, l'écart type des écarts quadratiques est calculé uniquement avec les résultats obtenus sur les ensembles de test. Cette manipulation est effectuée pour les 4 paramétrages adaptés à l'estimation des indicateurs avec la régression par vecteurs support.

TAB. 5 :  $EQM_0$  obtenues par 10-validations croisées sur l'espace d'apprentissage restreint.

Type de noyau →	Polynomiale		Gaussien	
Indicateur ↓	h= 1	h= 2	h= 4	h=10
1	0.82	0.90	0.83	0.82
2	0.33	0.36	0.33	0.33
3		0,05	0,06	
4	0.14	0.12	0.13	0.15
5		0,13	0.11	
6		0,07	0,06	
7		0,10	0,12	
8		0,29	0,29	

TAB. 6 : Ecart type des écarts quadratiques obtenus sur l'espace d'apprentissage restreint.

Type de noyau →	Polynomiale		Gaussien	
Indicateur ↓	h= 1	h= 2	h= 4	h=10
1	1,37	1,56	1,38	1,37
2	0,47	0,51	0,46	0,47
3		0,08	0,10	
4	0,35	0,29	0,34	0,37
5		0,20	0,17	
6		0,14	0,10	
7		0,30	0,35	
8		0,71	0,77	

## 5.3 Définition de groupes selon la distance à la zone d'apprentissage

Une distance de chaque point à l'espace d'apprentissage est définie dans le §5.3.1 puis des groupes de points éloignés sensiblement de la même distance sont identifiés

dans le §5.3.2. Cela permet de calculer l'EQM sur chacun des groupes et étudier son évolution en fonction de l'éloignement du groupe par rapport à l'ensemble d'apprentissage.

Le modèle retenu pour calculer l'écart quadratique sur les points en dehors de l'espace défini dans le §5.1.1 est le modèle construit à l'aide de l'ensemble des données disponibles dans le domaine d'apprentissage restreint.

### 5.3.1 Distance d'un point à l'ensemble d'apprentissage

La fonction définie par les One-class SVM dans le §5.1.1 a été utilisée pour déterminer la distance à l'ensemble défini comme espace d'apprentissage. Cette fonction  $f$  est telle que  $f(y) > 0$  si  $y$  appartient au domaine d'apprentissage et  $f(y) < 0$  sinon. De plus il existe une relation d'ordre entre l'éloignement d'un point  $y$  à la frontière de l'espace et la valeur absolue de la fonction  $f(y)$ . Ainsi il est possible de classer les points qui sont en dehors de l'espace d'apprentissage du plus proche au plus éloigné, en ordonnant la valeur absolue de leur image par  $f$ . La figure 7 en donne une représentation. Les quelques points excentrés observés sur la figure 5 correspondent aux points qui ont une valeur  $f(y)$  importante sur la figure 7.

### 5.3.2 Formation des groupes

Des groupes ont été définis arbitrairement en fixant des seuils sur la fonction  $f$ . La figure 7 illustre leur définition :

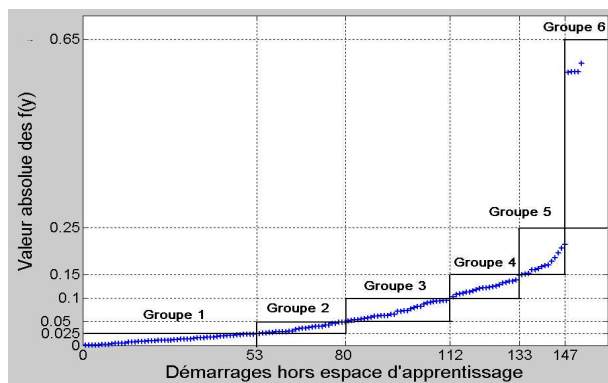


FIG. 7: Mesure de la distance à l'espace d'apprentissage des points extérieurs et définition des groupes

Le tableau 7 donne les bornes de chaque groupe ainsi que la moyenne des valeurs absolues des images par  $f$  des points qui composent le groupe.

TAB. 7 : Définition et caractéristiques des groupes

	Borne sup.	Taille du groupe	Moyenne de la valeur absolue des images des points du groupe
Groupe 1	0.025	53	0.012
Groupe 2	0.05	27	0.036
Groupe 3	0.10	32	0.073
Groupe 4	0.15	21	0.123
Groupe 5	0.25	14	0.173
Groupe 6	0.65	5	0.584

## 5.4 Erreur selon les groupes

L'EQM obtenue sur chaque groupe notée  $EQM(G_i)$  où  $i$  est le numéro du groupe a été calculée selon les modalités présentées en (4).

Il y a donc 6 valeurs d' $EQM(G_i)$  correspondant aux 6 groupes pour chacun des 8 indicateurs et pour chaque paramétrage utilisé pour les régressions, à savoir 4. Seuls les résultats les plus significatifs sont présentés.

La figure 7 montre que l'augmentation de la distance entre les groupes n'est pas constante. De manière à conserver cet écart, il a été choisi d'utiliser en abscisse, la moyenne des images des points qui composent les groupes dans les graphiques ci dessous. Cette moyenne est donnée dans le tableau 7. Les  $EQM_0$  données dans le tableau 5 seront représentées au point d'abscisse 0. Les résultats pour chaque indicateur vont maintenant être présentés et analysés.

Pour chacun des indicateurs, la méthode de régression associée à son paramétrage le plus performant pour chaque type de noyau a été retenue. Les indicateurs dont les résultats sont représentés sont synthétisés dans le tableau 8, avec la configuration sélectionnée pour la régression basée sur les vecteurs support.

L'indicateur 1 est estimé avec une très mauvaise précision. Il n'est donc pas intéressant de dégrader sa précision pour élargir le domaine de validité du modèle. Cependant, il est dépendant des mêmes dégradations que l'indicateur 2, son absence n'empêche donc pas la localisation.

TAB 8 : Configuration sélectionnée pour la régression, pour les résultats représentés

Indicateur	Noyau		Figure n°
	Poly	Gaussien	
2	h= 1	h= 10	8
3	h= 2	h= 4	9
4	h= 1	h= 4	10
5	h= 2	h= 4	11
6	h= 2	h= 4	12
7	h= 2	h= 4	13
8	h= 2	h= 4	14

Excepté pour les indicateurs 4, 7 et 8, l'EQM du groupe 6 pour les 2 types de noyau est beaucoup plus grande que celle des autres groupes. Pour une meilleure lecture des graphiques, il a été choisi de ne pas les faire apparaître. Pour les indicateurs 4, 7 et 8 seul le noyau gaussien fournit des EQM du même ordre de grandeur que les autres pour le groupe 6, respectivement 0,39 ; 1,04 ; 1,23. D'une manière générale, le noyau polynomial estime très mal les points du groupe 6 comparée au noyau gaussien. La moyenne des EQM des 8 indicateurs du groupe 6 pour un noyau polynomial à l'ordre 2 est de 50000 alors qu'elle est de 36 pour un noyau gaussien quand  $h$  est égal à 4.

Pour les autres groupes, les résultats sont présentés sur les figures de 8 à 14 ci dessous. Les barres verticales représentent les écarts types des écarts quadratiques obtenus pour chaque groupe. Un petit décalage sur

l'abscisse des groupes entre le noyau polynomial et gaussien a été effectué pour améliorer la lisibilité.

Les résultats pour les noyaux gaussiens sont représentés par des triangles et pour les noyaux polynomiaux par des ronds.

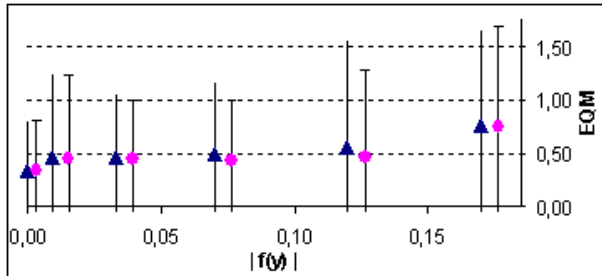


FIG. 8 : Accroissement de l'erreur pour l'indicateur 2

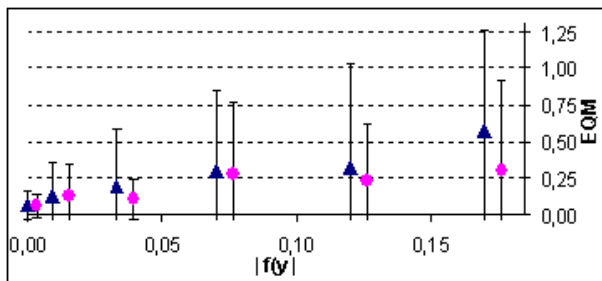


FIG. 9 : Accroissement de l'erreur pour l'indicateur 3

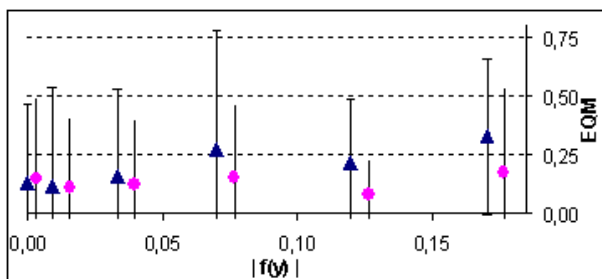


FIG. 10 : Accroissement de l'erreur pour l'indicateur 4

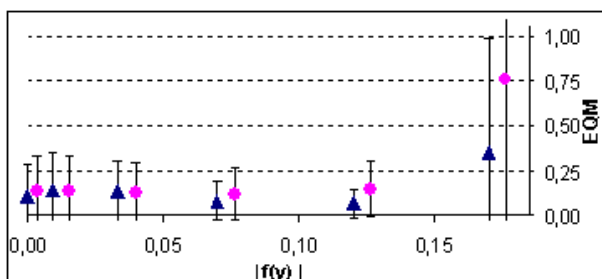


FIG. 11 : Accroissement de l'erreur pour l'indicateur 5

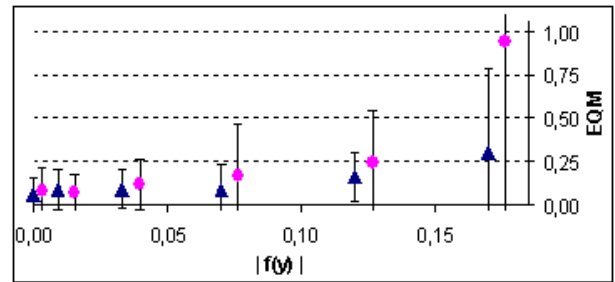


FIG. 12 : Accroissement de l'erreur pour l'indicateur 6

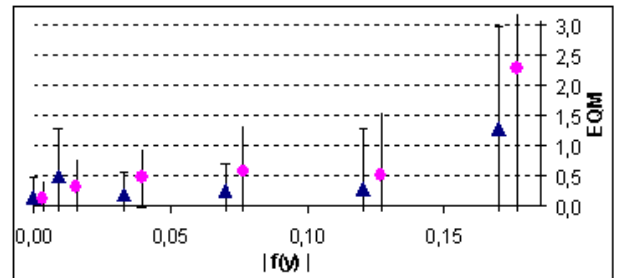


FIG. 13 : Accroissement de l'erreur pour l'indicateur 7

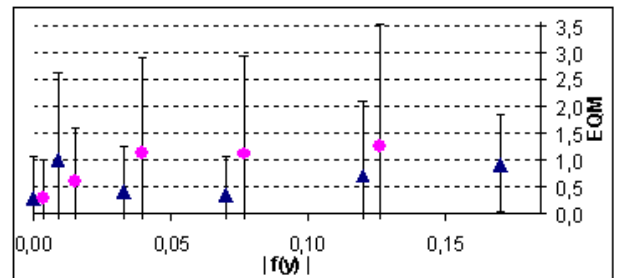


FIG. 14 : Accroissement de l'erreur pour l'indicateur 8

Pour l'indicateur 2, les résultats obtenus avec les deux types de noyau sont identiques. Cet indicateur est l'un des moins bien estimés sur la base d'apprentissage avec l'indicateur 8 (cf. tableaux 3 et 5). Cependant l'erreur d'estimation augmente mais stagne ensuite autour de 0,5 jusqu'au groupe 4. Il est également important de noter qu'il s'agit du seul indicateur où l'EQM du groupe 6 est plus faible avec un noyau polynomial (22 contre 185).

Pour l'indicateur 3, l'augmentation de l'erreur paraît linéaire. Cet indicateur est l'un des mieux prédit ce qui laisse une perspective d'élargissement du domaine. Bien que le noyau polynomial soit plus précis, l'erreur du groupe 6 est très élevée comparée au noyau gaussien (28 mille contre 70). Donc le noyau gaussien est plus performant pour estimer des points très éloignés de la base d'apprentissage.

L'indicateur 4 semble ne pas être très sensible à l'élargissement du domaine. Comme pour l'indicateur précédent, jusqu'au groupe 5 la régression par noyau polynomial est meilleure mais l'EQM du groupe 6 est bien plus faible pour un noyau gaussien (4,4 contre 0.39).

L'indicateur 5 est très bien estimé sur sa base d'apprentissage et l'erreur n'augmente pas sur les premiers



groupes. Ensuite le noyau gaussien donne de meilleurs résultats.

L'indicateur 6 représente une grandeur physique proche de l'indicateur 5 : L'un est le gradient de température maximal, l'autre la moyenne (cf. §2.3). Les observations sont donc équivalentes à l'indicateur 5.

Les indicateurs 7 et 8 représentent tous les 2 des durées spécifiques dans la phase de démarrage du moteur. C'est pourquoi des similitudes sont observées dans les résultats. Les EQM obtenues avec le noyau polynomial augmentent assez vite mais se stabilisent jusqu'au groupe 4. Ensuite au groupe 5, toujours pour un noyau polynomial, l'EQM devient très aléatoire et très élevée. Il a d'ailleurs été choisi de ne pas la faire apparaître sur la figure 14 pour conserver une bonne lecture des premiers points. Pour ces deux indicateurs le noyau gaussien est plus performant car les EQM sont plus faibles et moins dispersées. Il est cependant étrange de voir le 1<sup>er</sup> groupe moins bien estimé que les 3 suivants. Cette particularité doit trouver une explication dans la physique du démarrage d'un moteur d'avion car elle est observée pour les 2 indicateurs.

Cette étude a montré que l'erreur d'estimation restait modérée au moins jusqu'au groupe 4. Elle a également fait ressortir les différences de performance entre les deux types de noyau lorsque les données d'entrée des modèles s'éloignent légèrement du domaine d'apprentissage. Enfin elle a montré que, excepté pour l'indicateur 2, les noyaux gaussiens commettaient toujours moins d'erreur que les noyaux polynomiaux lorsque les données d'entrée s'éloignaient d'une distance importante. Cette caractéristique sera à prendre en compte dans le choix du paramétrage définitif des modèles en fonction des possibilités d'élargissement du domaine de validité de la méthode de monitoring. Ces possibilités d'élargissement vont dépendre des résultats de l'étude en cours sur la capacité à détecter les dégradations qui peuvent empêcher les prochains démarrages.

## 6. Conclusion

Un dispositif de surveillance de la capacité à démarrer d'un moteur d'avion a été décrit. L'analyse des modes de défaillance de la séquence de démarrage a permis la définition d'indicateurs représentatifs de l'état de santé des composants qui peuvent perturber le démarrage. Cependant, ces indicateurs sont sensibles aux conditions opératoires. La méthode de surveillance proposée, pour réduire l'effet des conditions opératoires, et donc améliorer le rapport signal/ bruit, est basée sur l'analyse des résidus. Les résidus sont la différence entre l'indicateur et son estimation par un modèle de régression qui a pour données d'entrée les paramètres caractéristiques des conditions opératoires. Cependant les modèles ne sont pas parfaits et leur pouvoir prédictifs varient selon les indicateurs. De plus la performance obtenue pour chaque indicateur n'est valable que pour des estimations qui ont lieu à l'intérieur du domaine d'apprentissage. Cet espace étant restreint par rapport à l'ensemble des conditions opératoires possibles

en exploitation, il a été proposé d'étudier l'évolution de l'erreur de prédiction lorsque les données d'entrée des modèles s'éloignent de l'espace d'apprentissage. Il a été observé que l'erreur quadratique moyenne, qui est le critère de performance défini pour les modèles, n'augmente pas à la même vitesse pour tous les indicateurs. Cependant l'accroissement de l'erreur reste modéré pour des petites distances au domaine d'apprentissage. Lorsque cette distance augmente, la moyenne ainsi que l'écart type des erreurs quadratiques d'estimation peut fortement s'élever, surtout pour les modèles basés sur des noyaux polynomiaux. Or, tolérer une erreur des modèles plus grande conduit à accroître l'écart type des distributions des résidus autant en fonctionnement normal qu'en fonctionnement dégradé. L'élargissement de la région de validité des modèles diminue donc les performances de détection. Une étude menée actuellement évalue les performances de la détection grâce à des données issues de démarrages en état dégradé. Elle montre que les distributions de l'état sain et de l'état dégradé sont suffisamment éloignées pour accepter une perte de précision des modèles. Il est donc possible d'élargir la région de validité tout en conservant les taux de fausses alarmes et de bonnes détections satisfaisant les exigences fixées par l'exploitant. L'importance de l'élargissement toléré dépendra des résultats de l'étude en cours.

## Références

- [1] M. Ashby et al. Intelligent Engine Systems. NASA CR -- 2005-213965.
- [2] K.R. Leamy et al. System and Method for Monitoring Engine Starting Systems. General Electric Compagny Patent, US6470258B1.
- [3] A. Ausloos, JR. Massé et al. Engine Start Capability, Snecma Patent pending. FR0950839. 2009.
- [4] J. Lacaille. Context Removal and Normalization. Snecma Patent. FR0858608. 2008.
- [5] Civil engine start data from test bench, 2006.
- [6] A J. Smola Regression estimation with support vector learning machines, 1996.
- [7] A. Ausloos, E Grall, P. Beuseroy et al. Estimation of Monitoring Indicators Using Regression Methods - Application to Turbofan Starting Phase. ESREL conference 2009.
- [8] DMJ Tax et RPW Duin. Support Vector Domain Description. Pattern Recognition Letters, vol. 20 pp 1191-1199, 1999.