

**Un modèle de prédiction et de classification des visites moteurs basé sur les  
machines à vecteur de support**

Par

Eric Gagnon

Mémoire présenté au Département d'informatique en vue de l'obtention du grade de  
maître ès sciences (M.Sc.)

FACULTÉ DES SCIENCES  
UNIVERSITÉ DE SHERBROOKE

Sherbrooke, Québec, Canada, janvier 2010



Library and Archives  
Canada

Bibliothèque et  
Archives Canada

Published Heritage  
Branch

Direction du  
Patrimoine de l'édition

395 Wellington Street  
Ottawa ON K1A 0N4  
Canada

395, rue Wellington  
Ottawa ON K1A 0N4  
Canada

*Your file* *Votre référence*  
ISBN: 978-0-494-63001-3  
*Our file* *Notre référence*  
ISBN: 978-0-494-63001-3

**NOTICE:**

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

**AVIS:**

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

---

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

■+■  
**Canada**

Le 15 janvier 2010

*le jury a accepté le mémoire de Monsieur Éric Gagnon  
dans sa version finale.*

Membres du jury

Professeur Kacem Zeroual  
Directeur de recherche  
Département d'informatique

Professeur Abdelhamid Benchakroun  
Membre  
Département d'informatique

Professeur Shengrui Wang  
Président rapporteur  
Département d'informatique

## Sommaire

En tant que contribution pour l'amélioration du service clients aux détenteurs d'avions, ce modèle de prédiction et de classification des visites moteurs à été inspiré des machines à vecteur de support (SVM) afin d'unifier les méthodes de prédiction des visites en usines. Ainsi, un tout nouveau modèle de classification de haut niveau pour la prédiction des visites de maintenance des moteurs d'avion en atelier se présente, permettant également l'identification des types de maintenances effectuées.

Les SVM classiques permettent de résoudre des problèmes de classification binaire, mais comme le problème abordé n'est pas de type binaire, certaines modifications sont nécessaires sur les SVM afin de les rendre plus flexibles et efficaces pour résoudre des problèmes de classes multiples.

Les résultats de l'expérimentation effectuée sur le modèle de prédiction et de classification des visites moteurs démontrent l'efficacité du modèle, ainsi que sa performance, sur des modèles multi-classes. Ces derniers sont beaucoup plus appropriés que d'autres résultats obtenus en utilisant les arbres de décision, ou encore l'algorithme des K plus proches voisins.

L'un des plus grands avantages de cette méthode, comparativement aux méthodes traditionnelles de classification et de prédiction, est qu'elle permet de saisir les relations présentes dans la structure arborescente et d'ainsi déduire les relations entre les classes dans les nœuds de l'arbre. Ceci constitue une percée importante dans l'avancement des techniques d'apprentissage automatique qui utilisent un historique de données pour classer ou prédire les valeurs futures qui risquent d'être rencontrées.

Il s'agit également d'un point tournant dans le passage d'un travail laborieux effectué manuellement par des experts à un algorithme performant qui déduit lui-même les règles intrinsèques enfouies dans un entrepôt de données.

## Remerciements

J'aimerais tout d'abord remercier le professeur Kacem Zéroual pour sa direction et son support constant tout au long de ce travail de recherche, sans qui, rien de tout cela n'aurait été possible. Merci pour sa persévérance, son souci du travail bien fait et surtout pour son immense coopération. Merci également à Shengrui Wang et Abdelhamid Benchakroun. Je tiens également à remercier Taoufik Guernine, pour son apport et ses connaissances des SVMs. Sans oublier ses nombreuses heures à analyser nos résultats.

Je remercie ma grand-mère, Fernande Berriault, pour son immense support tout au long de mes nombreuses années d'études, pour ses encouragements, sa patience dans les moments plus difficiles et surtout pour n'avoir jamais cessé de croire en mes capacités. Merci tout aussi à Lise Gagnon, Alain Aubé, Claudia Gagnon, Mathias B. Aubé et sa petite famille.

Je remercie Micheal Meraw pour m'avoir donné la chance d'explorer et de travailler dans le domaine de l'aéronautique chez Pratt & Whitney Canada. Merci à tous les experts qui m'ont guidé pendant cette année et demi passée à amasser, analyser et comprendre de nombreuses données. Merci à Sébastien Giguère, Jean Manibal, Stéphane Hébert, Maureen Fraser, Sharon Murphy, Charlene Larrissey, Daniel Gagnon, Rodrigo Moavro, Brian Webster, Margaret McWilliams, Guy Malette, Louis-Philippe Dubois, Jean-François Desrosiers, Mauro Gargano, Anna Cerone, Richard Dubé, Giovanni Mulas, Todd Bailey, Ken Ritchie, Jean Heydra, François Massicotte, Denis Jacob, Tom Milligan, Micheal Parent, Lucia Appleyard, Laszlo Ambrus, Mireille Boily, Nathalie Ouellet, Pierre Paquette, Jim Stothers, Duncan Murdoch, Micheal Ianniello, Pierre Heron et Laurent Azancot.

Je remercie le personnel du Département d'informatique de l'Université de Sherbrooke pour leur disponibilité et leur bon travail dans le département. Tout spécialement Marc Frappier, Chantal Proulx, Jean Goulet et Gabriel Girard. En terminant, je souhaiterais remercier du fond du cœur mes amis qui ont également su me supporter tout au long de cette aventure par leur présence et leur soutien; Serge Ménard, Carlos Colette, David Beaudry, Guillaume Lavoie, Louis-Marie Baptiste, Christophe Côté, Simon Gilbert, Simon Daniel Brisebois et Micheal Koutchikian. Un merci tout spécial pour Ara Merzikian.

# Table des Matières

<b>SOMMAIRE</b> .....	<b>II</b>
<b>REMERCIEMENTS</b> .....	<b>III</b>
<b>TABLE DES MATIÈRES</b> .....	<b>IV</b>
<b>LISTE DES ABRÉVIATIONS</b> .....	<b>VII</b>
<b>LISTE DES TABLEAUX</b> .....	<b>VIII</b>
<b>LISTE DES FIGURES</b> .....	<b>X</b>
<b>LISTE DES FIGURES</b> .....	<b>X</b>
<b>CHAPITRE 1 – INTRODUCTION</b> .....	<b>1</b>
1.1 <i>POURQUOI LA CLASSIFICATION ET PRÉDICTION DE MAINTENANCES DE MOTEURS?</i> .....	1
1.2 <i>LES TYPES DE MAINTENANCE DE MOTEURS D'AÉRONEFS</i> .....	3
1.3 <i>LES APPROCHES DE CLASSIFICATION ET DE PRÉDICTION</i> .....	5
1.4 <i>CYCLE DE VIE ADOPTÉ POUR ANALYSER LES DONNÉES</i> .....	8
<b>CHAPITRE 2 – CONCEPTS FONDAMENTAUX DU FORAGE DE DONNÉES POUR LA CLASSIFICATION ET LA PRÉDICTION</b> .....	<b>11</b>
2.1 <i>QU'EST-CE QUE LE FORAGE DE DONNÉES?</i> .....	11
2.1.1 <i>La classification</i> .....	13
2.1.2 <i>La prédiction</i> .....	14
2.2 <i>QUELLES SONT LES BASES DU FORAGE DE DONNÉES?</i> .....	14
2.2.1 <i>L'apprentissage inductif</i> .....	14
2.2.1.1 <i>Apprentissage supervisé</i> .....	14
2.2.1.2 <i>Apprentissage non-supervisé</i> .....	15
2.2.2 <i>Les statistiques</i> .....	16
2.2.3 <i>L'apprentissage machine</i> .....	17
2.3 <i>LES MODÈLES DE FORAGE DE DONNÉES</i> .....	17
2.3.1 <i>Modèle de vérification</i> .....	17
2.3.2 <i>Modèle de découverte</i> .....	18
2.4 <i>LES ENTREPÔTS DE DONNÉES</i> .....	18
2.5 <i>LES PROBLÈMES DU FORAGE DE DONNÉES</i> .....	19
2.6 <i>NOTIONS IMPORTANTES SUR LES SVM</i> .....	20
2.6.1 <i>Cas binaire et ternaire – les explications théoriques de l'approche choisie</i> .....	22
2.6.1.1 <i>Cas de trois classes</i> .....	22
<b>CHAPITRE 3 – TRAVAUX CONNEXES</b> .....	<b>25</b>
3.1 <i>LA LOGIQUE FLOUE</i> .....	25
3.2 <i>LES ARBRES DE DÉCISION</i> .....	27
3.3 <i>LES KNN (KPPV: K PLUS PROCHES VOISINS)</i> .....	29
3.4 <i>LES SVM</i> .....	31
<b>CHAPITRE 4 – APPROCHE DE CLASSIFICATION ET DE PRÉDICTION DE MAINTENANCE DE MOTEURS</b> .....	<b>36</b>
4.1 <i>INTRANTS</i> .....	38
4.1.1 <i>Sélection des données pertinentes</i> .....	39
4.1.1.1 <i>Voir ce qui se fait manuellement</i> .....	39
4.1.1.2 <i>Les maintenances de base et les bulletins de service recommandés</i> .....	40
4.1.1.3 <i>Inclure l'historique d'événements passés</i> .....	40
4.1.1.4 <i>Inclure les événements fortuits (ce qu'il peut arriver)</i> .....	40

4.1.1.5 Inclure les facteurs macro-économiques et environnementaux influençant le modèle de moteur .....	41
4.1.2 Conversion des valeurs catégorielles .....	42
4.1.3 Normalisation des données.....	43
4.2 CREATION DU MODELE DE PRÉDICTION.....	44
4.2.1 Création des jeux d'observation.....	45
4.2.2 Appliquer le SVM sur les jeux d'observation.....	47
4.2.3 Création de l'arbre de décision hiérarchique.....	52
4.2.4 Répéter les étapes précédentes et raffiner le modèle.....	52
4.2.5 Utiliser le modèle créé avec le jeu de validation pour prédire les résultats du jeu de test.....	55
4.3 EXTRANTS .....	55
<b>CHAPITRE 5 – ÉTUDE EXPÉRIMENTALE.....</b>	<b>57</b>
5.1 UN EXEMPLE CONCRET D'APPLICATION .....	57
5.2 COMPARAISONS.....	63
5.3 ANALYSES DE PERFORMANCE .....	67
<b>CHAPITRE 6 – CONCLUSION.....</b>	<b>71</b>
6.1 CONCLUSION ET TRAVAUX FUTURS .....	71
<b>ANNEXE A - LES DIFFÉRENTES SOURCES DE DONNÉES DISPONIBLES ET LA DESCRIPTION DU CONTENU DE CHACUNE D'ELLES.....</b>	<b>74</b>
A.2 MAINTENANCE MANUALS .....	75
A.3 SHOP VISIT HISTORICAL DATA (TRAC, TRS).....	76
A.4 WARRANTY DATA .....	76
A.5 PART SALES DATA .....	76
A.6 MODÈLES DE PRÉVISION DE CHAQUE DÉPARTEMENT .....	76
A.6.2 Exemple de notes relatives aux rencontres avec les experts: .....	84
A.7 NOTES DES ADMINISTRATEURS DE SYSTÈME.....	85
A.8.1 Les facteurs macro-économiques .....	88
A.8.2 les facteurs environnementaux.....	89
<b>ANNEXE B – RÉPONSES AUX QUESTIONS POSÉES DANS LA SÉRIE D'ÉTAPES DU PLAN DE TRAITEMENT DES ENTRÉES.....</b>	<b>90</b>
ÉTAPE B.1 – CE QUI EST FAIT (CE QUI EXISTE DÉJÀ).....	90
ÉTAPE B.2 – LA BASE (CE QUE L'ON RECOMMANDE).....	100
ÉTAPE B.3 – L'HISTORIQUE (QUE S'EST-IL PASSÉ AVANT?) .....	101
ÉTAPE B.4 – LE POSSIBLE (QUE PEUT-IL ARRIVER?).....	119
ÉTAPE B.5 – LA VISIO ÉCONOMIQUE (UN LIEN MACRO?).....	119
<b>ANNEXE C - LA DESCRIPTION DÉTAILLÉE DES NOMBREUSES CARACTÉRISTIQUES RETENUES EN ENTRÉE.....</b>	<b>126</b>
<b>ANNEXE D - LA CONVERSION DES VALEURS CATEGORIELLE EN VALEURS NUMERIQUES (« RANKING »).....</b>	<b>138</b>
<b>ANNEXE E – JEUX D'OBSERVATION SUR LE SVM.....</b>	<b>144</b>
E.1 PREMIER JEU D'ESSAI INCLUANT TOUTES LES CARACTÉRISTIQUES.....	144
E.2 DEUXIÈME JEU D'ESSAI INCLUANT CERTAINES DES CARACTÉRISTIQUES CHOISIES.....	145
E.3 TROISIÈME JEU D'ESSAI INCLUANT CERTAINES DES CARACTÉRISTIQUES CHOISIES.....	147
E.4 QUATRIÈME JEU D'ESSAI INCLUANT LA PLUPART DES CARACTÉRISTIQUES CHOISIES .....	150
E.5 CINQUIÈME JEU D'ESSAI INCLUANT CERTAINES DES CARACTERISTIQUES CHOISIES.....	151
E.6 SIXIÈME JEU D'ESSAI INCLUANT CERTAINES DES CARACTÉRISTIQUES CHOISIES .....	152
E.7 SEPTIÈME JEU D'ESSAI INCLUANT CERTAINES DES CARACTÉRISTIQUES CHOISIES .....	152
E.8 HUITIÈME JEU D'ESSAI INCLUANT CERTAINES DES CARACTÉRISTIQUES CHOISIES .....	154
E.9 NEUVIÈME JEU D'ESSAI INCLUANT CERTAINES DES CARACTÉRISTIQUES CHOISIES.....	155
E.10 DIXIÈME JEU D'ESSAI INCLUANT CERTAINES DES CARACTÉRISTIQUES CHOISIES.....	156
E.11 LES EXEMPLES RELATIFS AU TEMPS SUR SPSS 14.0 .....	157

<b>ANNEXE F - UN EXEMPLE SUR UN MODÈLE SPÉCIFIQUE. ....</b>	<b>161</b>
F.1. SERVICE BULLETINS .....	161
F.2. MAINTENANCE MANUALS .....	163
F.3. SHOP VISIT HISTORICAL DATA (TRAC, TRS).....	164
F.4. WARRANTY DATA .....	169
F.5. PART SALES DATA.....	169
F.6. MODÈLES DE PRÉVISION DE CHAQUE DÉPARTEMENT .....	169
F.7. NOTES DES ADMINISTRATEURS DE SYSTÈME.....	171
F.8. LES FACTEURS MACRO-ÉCONOMIQUES .....	171
<b>ANNEXE G – TECHNIQUE STATISTIQUE POUR L'ESTIMATION DE DISTRIBUTIONS.....</b>	<b>173</b>
<b>ANNEXE H – NOTES DE LA RENCONTRE AVEC LE GROUPE SAS ET LA SOLUTION PROPOSÉE (ANGLAIS) .....</b>	<b>175</b>
<b>BIBLIOGRAPHIE .....</b>	<b>180</b>



## Liste des abréviations

AD / DT	Arbre de Décision / Decision Tree
AI / IA	Artificial Intelligence / Intelligence Artificielle
AMU	Average Monthly Usage / Usage mensuel moyen
APRE	l'Agent de Prédiction de Réaction Émotionnelle
AVG	Average / Moyenne
BPO	Basic Planned Overhaul (O/H)
BPR	Basic Planned Removal
BUO	Basic Unplanned Overhaul (O/H)
BUR	Basic Unplanned Removal
COP	Crude Oil Prices / Prix du pétrole brut
CRF	Conditional Random Fields
DRF	Discriminative Random Fields
FD / DM	Forage de Données / Data Mining
IFSD	In-Flight Shut Down / Arrêt du moteur en vol
JIT	Just In Time
KDD	Knowledge Discovery
KNN / Kppv	K Nearest Neighbors / K plus proches voisins
MRT	Monthly Running Time / Usage mensuel moyen
NB	Number / Nombre
NUO	Non-basic Unplanned Overhaul (O/H)
NUR	Non-basic Unplanned Removal
NPPS	Nuclear Power Plants Safety management alorithm
NPO	Non-basic Planned Overhaul (O/H)
NPR	Non-basic Planned Removal
USCP	US Corporate Profits / Profits corporatifs aux É.-U.
USIR	US Interest rates / Taux d'intérêts des États-Unis
TOC	Theory Of Constraints
SV	Support Vectors / Vecteurs de Support
SVM	Support Vector Machine
WGDP	World's Gross Domestic product / Produit domestique brut mondial

## Liste des Tableaux

TABLE 1 - CLASSIFICATION D'UN ENSEMBLE DE DONNÉES IRIS : CAS TROIS CLASSES .....	22
TABLE 2 - RÉSULTATS DE LA CLASSIFICATION AUTOMATIQUE, EN POURCENTAGE DE BONNE PRÉDICTION ....	33
TABLE 3 - LISTE DES CARACTÉRISTIQUES SÉLECTIONNÉES .....	49
TABLE 4 - PARAMÈTRES D'EXÉCUTION EXEMPLE 1 .....	59
TABLE 5 - RÉSULTATS DE L'EXEMPLE 1 .....	60
TABLE 6 - PARAMÈTRES D'EXÉCUTION EXEMPLE 2 .....	60
TABLE 7- RÉSULTATS DE L'EXEMPLE 2 .....	60
TABLE 8 - PARAMÈTRES D'EXÉCUTION EXEMPLE 3 .....	62
TABLE 9 - RÉSULTATS DE L'EXEMPLE 3 .....	62
TABLE 10 - RÉSULTATS EN AUGMENTANT LE NOMBRE DE CLASSES .....	63
TABLE 11 - MOYENNE ET ÉCART-TYPE DES CSN .....	64
TABLE 12 - COMPARATIF DES MOYENNES ET ÉCARTS-TYPES SUR 5 CARACTÉRISTIQUES (CLASSE 1) .....	66
TABLE 13 -COMPARATIF DES MOYENNES ET ÉCARTS-TYPES SUR LA CARACTÉRISTIQUE LA PLUS INFLUENTE (CLASSE 2) .....	67
TABLE 14 - « CALCULATIONS FOR THE YEARS PRESENTED IN THE WORKBOOK » .....	78
TABLE 15 - « OH COMPLETED BY CSC BERLIN IN 2006 AND BELGIAN POLICE LISTING » .....	80
TABLE 16 - ADACXX206_060601 ZONE 1 .....	83
TABLE 17 - ADACXX206_060601 ZONE 2 .....	83
TABLE 18 - DISTRIBUTION DES MAINTENANCES « MOTEUR ENLEVÉ » SUR 5 ANS, PAR MS EXCEL .....	100
TABLE 19 - SOMMAIRE DES ÉVÉNEMENTS REGROUPÉS .....	102
TABLE 20 - SOMMAIRE DES ÉVÉNEMENTS POUR TOUTE LA FLOTTE DES MODÈLES 300 .....	102
TABLE 21 - TAUX PUBLIÉS DE BUR ET IFSD SUR UN 306B .....	118
TABLE 22 - RÉPARTITION DES PAYS D'APPARTENANCE POUR LES 306B .....	127
TABLE 23 - RÉPARTITION DES CLASSES D'OPÉRATEURS POUR LES 306B .....	129
TABLE 24 - RÉPARTITION DES NOMS D'OPÉRATEURS POUR LE 306B .....	130
TABLE 25 - RÉPARTITION DES NOMS COURTS DES DÉTENTEURS POUR LE 306B .....	131
TABLE 26 - RÉPARTITION DES PAYS D'ÉVÉNEMENTS POUR LE 306B .....	132
TABLE 27 - RÉPARTITION DES ÉTATS/PROVINCES D'ÉVÉNEMENTS POUR LE 306B .....	132
TABLE 28 - RÉPARTITION DES ÉVÉNEMENTS DU 306B .....	133
TABLE 29 - STATISTIQUES SUR L'ÂGE DES 306B .....	133
TABLE 30 - CONVERSION DU NUMÉRO DE SÉRIE .....	138
TABLE 31 - CONVERSION DU PAYS D'APPARTENANCE .....	139
TABLE 32 - RÉSULTATS DE LA CONVERSION DES PAYS D'APPARTENANCE .....	139
TABLE 33 - CONVERSION DU NUMÉRO DE SÉRIE DE L'AÉRONEF .....	139
TABLE 34 - CONVERSION DE LA CLASSE DE L'OPÉRATEUR .....	140
TABLE 35 - RÉSULTATS DE LA CONVERSION DES CLASSES D'OPÉRATEURS .....	140
TABLE 36 - SCORES DES PAYS RICHES ET NON DÉLINQUANTS .....	141
TABLE 37 - PAYS .....	142
TABLE 38 - ÉTAT/PROVINCE .....	142
TABLE 39 - CLASSIFICATION D'ÉVÉNEMENT SUR 3 CARAVTÈRES .....	142
TABLE 40 - SCORES DES TYPES D'ÉVÉNEMENTS .....	143
TABLE 41 - PARAMÈTRES DU SVM .....	144
TABLE 42 - RÉSULTATS OBTENUS SUR LE 2 <sup>IÈME</sup> JEU D'ESSAI .....	146
TABLE 43 - RÉSULTATS OBTENUS SUR LE 3 <sup>IÈME</sup> JEU D'ESSAI .....	147
TABLE 44 - RÉSULTATS OBTENUS SUR LE 4 <sup>IÈME</sup> JEU D'ESSAI .....	151
TABLE 45 - RÉSULTATS OBTENUS SUR LE 7 <sup>IÈME</sup> JEU D'ESSAI .....	153
TABLE 46 - « CLUSTER MEMBERSHIP » EXEMPLE 1 RELATIF AU TEMPS .....	157
TABLE 47 - « CLUSTER MEMBERSHIP » EXEMPLE 2 RELATIF AU TEMPS .....	158
TABLE 48 - « CLUSTER MEMBERSHIP » EXEMPLE 3 RELATIF AU TEMPS .....	158
TABLE 49 - « CLUSTER MEMBERSHIP » EXEMPLE 4 RELATIF AU TEMPS .....	159
TABLE 50 - « CLUSTER MEMBERSHIP » EXEMPLE 5 RELATIF AU TEMPS .....	159

TABLE 51 - FAMILLES DE MOTEURS .....	163
TABLE 52 - MANUEL DE MAINTENANCE AVEC CHACUNE DES CARACTÉRISTIQUES DU 306B .....	163
TABLE 53 - HISTORIQUE D'UN MOTEUR DE TYPE 306B.....	165
TABLE 54 - DÉTAILS À CONSIDÉRER.....	167

# Liste des figures

FIGURE 1 - LES TYPES DE MAINTENANCES (« <i>WATOG INT'L CODING</i> ») .....	4
FIGURE 2 - ALGORITHME EM.....	16
FIGURE 3 - PLUSIEURS HYPERPLANS SÉPARENT DEUX CLASSES DE DONNÉES.....	21
FIGURE 4- ARBRE DE DÉCISION CONÇU AVEC ID3.....	28
FIGURE 5 - ALGORITHME POUR LE CLASSEMENT D'IMAGES.....	29
FIGURE 6 - APPROCHE DE CLASSIFICATION ET DE PRÉDICTION DE MAINTENANCES DE MOTEURS.....	37
FIGURE 7 - PROCESSUS DE SÉLECTION DES DONNÉES .....	42
FIGURE 8 - LE JEU D'ENTRAÎNEMENT.....	46
FIGURE 9 - LE JEU DE VALIDATION.....	46
FIGURE 10 - LE JEU DE TEST.....	47
FIGURE 11 - LE CHOIX DE LA FONCTION DE « <i>KERNEL</i> ».....	51
FIGURE 12 - ARBRE DE DÉCISION HIÉRARCHIQUE.....	52
FIGURE 13 - ARBRE DE DÉCISION HIÉRARCHIQUE.....	53
FIGURE 14 - DONNÉES CONVERTIES.....	57
FIGURE 15 - DONNÉES NORMALISÉES.....	58
FIGURE 16 - RÉSULTATS DU SVM POUR LE CHOIX D'UNE FONCTION « <i>KERNEL</i> ».....	58
FIGURE 17 - RÉSULTATS DU JEU DE VALIDATION .....	59
FIGURE 18 - ARBRE HIÉRARCHIQUE .....	62
FIGURE 19 - PARAMÈTRES TRIÉS EN NOMBRE CROISSANT DE DIMENSIONS (NB DE CARACTÉRISTIQUES UTILISÉES).....	69
FIGURE 20 - PARAMÈTRES TRIÉS EN NOMBRE CROISSANT D'ITÉRATIONS.....	69
FIGURE 21 - COURBE DE TEMPS SUR LES ESSAIS SVM TRIÉS PAR NOMBRE CROISSANT D'ENREGISTREMENTS.....	70
FIGURE 22 - HIÉRARCHIE DES ÉLÉMENTS LIÉS AU MOTEUR.....	91
FIGURE 23 - PRÉVISIONS DÉPARTEMENTALE, COMPARANT LE 306B AU MODÈLE LE PLUS RÉCENT, LE 600.....	94
FIGURE 24 - TABLEAU DE PRÉVISIONS.....	95
FIGURE 25 - GRAPHIQUE DE PRÉVISIONS .....	95
FIGURE 26 - LCFs UTILISÉS DANS LA PRÉDICTION .....	96
FIGURE 27 - LISTE DES OPÉRATEURS UTILISÉE DANS LA PRÉDICTION .....	97
FIGURE 28 - UN EXEMPLE DE « <i>WORK SCOPE</i> ».....	98
FIGURE 29 - UN EXEMPLE DE « <i>CAMPAIGN REVIEWS</i> » .....	98
FIGURE 30- RÉSULTATS DE L'ALGORITHME SUR MS EXCEL.....	99
FIGURE 31 - DÉCOMPTE DES TYPES D'ÉVÉNEMENTS .....	101
FIGURE 32 - HISTORIQUE DES ÉVÉNEMENTS SUR LES MODÈLES 300.....	102
FIGURE 33 - LES BPOS SUR LES MODÈLES 300.....	103
FIGURE 34 - LES BPOS SUR LES MODÈLES 306B .....	104
FIGURE 35 - LES BPRS SUR LES MODÈLES 300 .....	104
FIGURE 36 - LES BPRS SUR LES MODÈLES 306B .....	105
FIGURE 37 - LES BUOS SUR LES MODÈLES 300 .....	105
FIGURE 38 - LES BURs SUR LES MODÈLES 300 .....	106
FIGURE 39 - LES BURs SUR LES MODÈLES 306B.....	106
FIGURE 40 - LES NUOS SUR LES MODÈLES 300.....	107
FIGURE 41 - LES NURS SUR LES MODÈLES 300.....	107
FIGURE 42 - LES NURS SUR LES MODÈLES 306B .....	108
FIGURE 43 - ÉVÉNEMENTS SUR LE MOTEUR 00001 DES 306B.....	108
FIGURE 44 - ÉVÉNEMENTS SUR LE MOTEUR 00002 DES 306B.....	109
FIGURE 45 - LE TSN COMPARÉ AUX CSN POUR UN MODÈLE 306B.....	110
FIGURE 46 - LE TSN COMPARÉ AUX CSN POUR UN AUTRE MODÈLE 306B .....	110
FIGURE 47 - DISTANCE MARQUÉE ENTRE LE TSN ET LES CSN D'UN 306B.....	111
FIGURE 48 - DISTANCE ENTRE LE TSN ET LES CSN D'UN 306B .....	112
FIGURE 49 - DISTANCE ENTRE LE TSN ET LES CSN D'UN AUTRE 306B.....	112
FIGURE 50 - PRÉVISION D'ÉVÉNEMENT POSSIBLE SUR LE MOTEUR 000014.....	113
FIGURE 51 - PRÉVISION D'ÉVÉNEMENT POSSIBLE SUR LE MOTEUR 000014 #2.....	114

FIGURE 52 - LES TEMPS PAR RAPPORT AUX CYCLES .....	115
FIGURE 53 - DISTRIBUTION DES TEMPS ET CYCLES.....	115
FIGURE 54 - DIFFÉRENCES ENTRE CHAQUE ÉVÉNEMENT EN NOMBRE DE JOURS ET DE TEMPS/CYCLES .....	116
FIGURE 55 - CORRÉLATION ENTRE LES DIFFÉRENCES DES JOURS ET LES ÉVÉNEMENTS .....	117
FIGURE 56 - « <i>CRUDE OIL</i> » VS NOMBRE DE MOTEURS MIS EN CIRCULATION .....	120
FIGURE 57 - « <i>CRUDE OIL</i> » VS USAGE MENSUEL EN NOMBRE D'HEURES DE VOL .....	120
FIGURE 58 - LE NOMBRE DE MISES EN SERVICE VS LES USCP .....	121
FIGURE 59 - USAGE MENSUEL EN NOMBRE D'HEURES DE VOL VS USCP.....	122
FIGURE 60 - LE NOMBRE DE MISES EN SERVICE VS WGDP.....	122
FIGURE 61 - L'USAGE MENSUEL EN NOMBRE D'HEURES DE VOL VS WGDP .....	123
FIGURE 62 - NOMBRE DE MISES EN SERVICE VS S&P500 .....	124
FIGURE 63 - L'USAGE MENSUEL EN NOMBRE D'HEURES DE VOL VS S&P500 .....	124
FIGURE 64 - LE NOMBRE DE MISES EN SERVICES VS USIR .....	125
FIGURE 65 - L'USAGE MENSUEL EN NOMBRE D'HEURES DE VOL VS USIR.....	125
FIGURE 66 - FONCTION DE RETRAITE DES MOTEURS.....	134
FIGURE 67 - LISTE EXPORTABLE DE TRS-TRAC (SAS).....	168
FIGURE 68 - MOYENNES DES TAUX .....	168
FIGURE 69 - TAUX EN LIGNE .....	169
FIGURE 70 - « <i>ENGINE CONFIGURATION NEW</i> » .....	170
FIGURE 71 - FORECAST VERSION X 2006-05-03 .....	171

## **Chapitre 1 – Introduction**

### ***1.1 Pourquoi la classification et prédiction de maintenances de moteurs?***

Une compagnie dans le domaine de l'aéronautique vend, répare et maintient des moteurs d'aéronefs dans les quatre coins du monde. D'un peu partout, surviennent des événements qui sont rapportés à l'entreprise chaque fois qu'une communication a lieu avec un client. Lorsqu'un problème survient et qu'une action doit être entreprise sur un moteur, un rapport d'événement est enregistré dans un système de gestion d'événements. De plus, lorsque tout autre type d'interaction pouvant changer les données à propos des modèles de moteur en service, ou même à propos de ceux qui seront mis en service, a lieu, un rapport d'événement est également enregistré dans le système. Une quantité immense de données est accumulée et beaucoup d'efforts doivent être déployés pour nettoyer, analyser et extraire des connaissances de cette mer d'information.

Les connaissances enfouies dans ces données sont cruciales pour maintenir un service de qualité et accroître le niveau de satisfaction des clients. Pour cette entreprise, les réponses à deux questions doivent être connues afin de s'élever au rang de leader mondial: quand est-ce que les moteurs devront entrer en usine et pour quel type de bris ou maintenance le feront-ils? Sachant cela, la compagnie peut prédire ce qui sera à accomplir dans les prochains mois et ainsi s'assurer d'avoir les

ressources nécessaires et disponibles pour retourner les avions en service à l'intérieur d'un délai de vingt-quatre heures. Comme les données ne sont pas véritablement contenues dans un entrepôt de données, mais dans plusieurs systèmes, il n'est pas possible de forer d'entrepôt afin d'obtenir les connaissances recherchées. De plus, très peu de développements sont effectués en ce sens et l'entreprise n'utilise pas de forage de données.

Actuellement, pour effectuer la prédiction des visites de moteurs en usine, les experts du domaine doivent s'en remettre à leur expérience personnelle et à l'analyse des événements survenus dans le passé pour tenter de déduire une tendance, des cycles ou encore des séries qui se cachent dans les données accumulées sur les modèles de moteurs. Chaque département est en charge des prédictions pour un modèle en particulier, car tous les modèles n'ont pas les mêmes caractéristiques de fabrication et d'utilisation. Il y a donc beaucoup de travail manuel, d'analyses statistiques et même parfois quelques fonctions de hasard qui sont utilisées pour créer des algorithmes de prédictions sur un modèle de moteur. Certains algorithmes performant mieux que d'autres et on constate ainsi une grande inégalité dans la fiabilité des prédictions.

Le principal objectif de ce travail est donc de présenter une technique hybride, permettant de créer un modèle de prédiction unifié qui soit utilisable par tous les départements, tout en permettant d'améliorer la façon de faire actuelle face aux problèmes de grande taille, surtout lorsque nous observons la présence de plusieurs classes. Ainsi, un modèle de prédiction uniforme, applicable à tous les modèles de moteurs, s'avèrerait être un outil indispensable à l'entreprise pour se hisser au rang du plus grand producteur de moteurs d'avions mondial.

Nous faisons face, ici, à un problème à deux volets. Lorsque nous abordons la maintenance des moteurs d'avions, nous devons considérer deux aspects afin de satisfaire les besoins de l'industrie. Pour répondre adéquatement à la demande des clients, une compagnie en aéronautique qui construit des moteurs d'avions doit prévoir quel type de maintenance elle devra faire sur ses moteurs et aussi quand

elle devra le faire. Le type de maintenance attendu est important, mais de pouvoir savoir quand cette maintenance aura lieu est d'autant plus important. En sachant quand auront lieu les maintenances, il est possible de prévoir les pièces et l'espace nécessaire dans l'atelier pour effectuer les maintenances, sans que le client ne souffre des délais de livraison des pièces, ou encore, des délais occasionnés lorsque la capacité maximale des ateliers est atteinte. Il s'agit ici de l'information clé contenue dans les données historiques pour atteindre la pleine satisfaction des clients.

C'est dans cette optique que le projet a pris forme. Beaucoup de recherches ont été faites pour comprendre le domaine d'application et beaucoup de discussions ont eu lieu avec les experts du domaine pour tenter de déterminer quelles données utiliser, et surtout, quelles sont les données qui risquent d'avoir un impact pour classer ou prédire les futures visites en atelier.

## **1.2 Les types de maintenance de moteurs d'aéronefs**

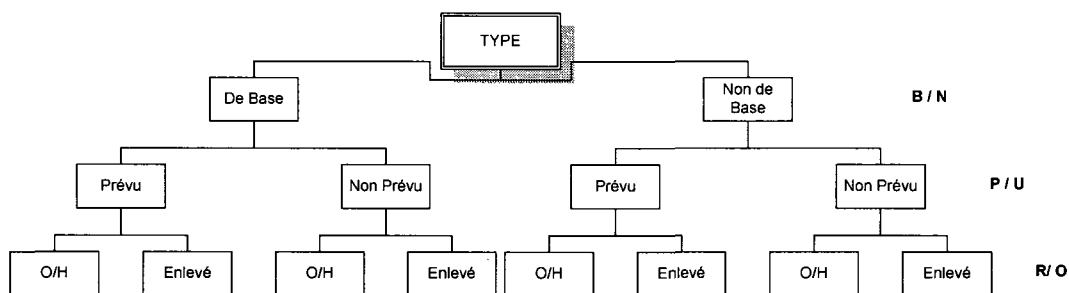
Afin de comprendre la notion de type de maintenance, nous devons nous référer à la nomenclature « WATOG », signifiant « *World Airlines Technical Operations Glossary* », qui se veut la référence internationale dans la description d'événements reliés au domaine de l'aéronautique.

Il existe plusieurs types de maintenances possibles sur un moteur d'aéronef, ainsi la norme « WATOG » classe ces types à l'aide de 5 lettres. Chaque lettre se voit issue d'un ensemble contenant plusieurs choix. La première lettre représente la nature de l'événement. S'agit-il d'un événement de base (B) ou non (N)? La seconde lettre représente la planification de l'événement. S'agit-il d'un événement prévu (P) ou imprévu (U)? La troisième lettre représente l'état du moteur. A-t-on enlevé le moteur (R) ou a-t-on démonté le moteur en pièces (O)? Les deux dernières lettres nous



indiquent si les rapports d'événements ont été effectués et ils ne nous sont d'aucun intérêt dans notre analyse.

On peut représenter graphiquement, à l'aide d'un arbre, l'ensemble des combinaisons possibles qui ont été considérées dans notre projet:



**Figure 1 - Les types de maintenances (« WATOG Int'l coding »)**

Donc, nous pouvons avoir un type de maintenance qui soit BPO. Cela signifie qu'il s'agit d'une maintenance de base, prévue, et que l'on doit défaire le moteur en pièces (O/H signifie « *Overhaul* »).

Nous pouvons également avoir un type de maintenance qui soit NUR. Cela signifie qu'il s'agit d'une maintenance non de base, non prévue, et qu'on a du enlever le moteur de l'aéronef (« *removal* »).

Évidemment, les maintenances de type non prévues sont beaucoup plus difficiles à prévoir et il est alors difficile d'avoir toutes les pièces en stock à temps pour faire la maintenance. Il y a beaucoup d'autres codes possibles, mais ici, nous nous concentrons sur les suivants : BPR, BPO, NPR, NPO, BUR, BUO, NUR et NUO. Il semble bien, jusqu'à maintenant, que ces cas soient linéairement séparables.

Lorsque nous nous attardons à savoir *quand* aura lieu cette maintenance, alors là ça se complique. Il semble bien que ce cas ne soit pas linéairement séparable.

### **1.3 Les approches de classification et de prédiction**

Comme nous l'avons vu, il existe plusieurs types de maintenances. Certaines sont à délais fixes (par exemple, à toutes les 3000 heures de vol, une pièce connue doit être remplacée, ou inspectée) et d'autres sont fortuites (comme par exemple, l'ingestion d'un oiseau dans le moteur pendant un vol). Le problème consiste alors à classer les types de maintenances selon leurs ressemblances et déterminer des classes de maintenances dans lesquelles nous pouvons approximer le nombre d'heures de vol qui sera atteint lorsque l'aéronef devra entrer en usine. Comme la majorité des approches en classification et prédiction fonctionnent sur des problèmes d'ordre binaire (contenant 2 classes), alors il nous faut trouver une méthode permettant la séparation de plusieurs classes (multi-classes).

Il existe plusieurs approches pour résoudre ce type de problèmes, chacune présentant des avantages et des inconvénients. L'approche d'un système expert peut s'avérer très efficace pour uniformiser les techniques de prédiction utilisées par les nombreux départements de l'entreprise. Comme tous les experts ont été rencontrés, il est possible de déduire les règles communes rencontrées et ainsi implanter les composantes les plus communes et de proposer, par la suite, des options pour chaque type de moteur rencontré. Le principal avantage de cette approche est que le système est conçu spécifiquement pour les besoins actuels de prédiction et répond exactement aux exigences cognitives recherchées. Cependant, puisque l'industrie ne cesse de progresser, plusieurs moteurs s'ajoutent à la liste d'années en années. Ainsi, le système se doit d'être extensible et flexible pour permettre l'ajout de nouveaux modèles, de nouvelles règles de décisions et il se peut bien que la maintenance d'un

tel système devienne rapidement un énorme casse-tête, affectant de la sorte la stabilité et la performance du système.

Dans l'approche de la recherche opérationnelle, plusieurs outils sont à notre disposition pour résoudre des problèmes d'optimisation linéaires, des problèmes mixtes, ou encore, ceux qui se présentent sous une forme quadratique. Cette approche nécessite une fonction à minimiser ou maximiser. Nous pouvons alors minimiser les délais d'attente de maintenance en s'assurant d'effectuer une optimisation sujette à plusieurs contraintes (ici les différentes caractéristiques des nombreux moteurs), en limitant l'espace de solutions par des bornes inférieures et supérieures. L'avantage de cette approche est qu'elle propose la meilleure solution (un optimum), cependant, toutes les contraintes et variables doivent être modélisées, et cela permet difficilement l'ajout ultérieur de nouvelles contraintes sans les connaissances d'un expert ou celles d'un usager qui a participé au développement ou à l'implantation du système.

Une autre approche [1] consiste à appliquer le forage de données. Généralement, nous pouvons utiliser cette approche pour la classification d'éléments, l'association d'éléments, la prédiction de classification ou d'association, la planification de déplacements (ex: robotique), ou encore, la segmentation d'ensemble d'éléments. Plusieurs travaux [2,3,4,5,6,7,8] abordent les problèmes de classification et de prédiction dans le forage de données. L'avantage de ces techniques est qu'elles peuvent être effectuées sur l'ensemble des données et qu'elles ne nécessitent pas beaucoup de modélisation. Il faut, par contre, que les algorithmes présents permettent d'effectuer des classifications ou prédictions sur plus de deux classes, ce qui est rarement le cas pour les algorithmes existants.

De nos jours, les grands projets incluant le forage de données représentent, pour les entreprises, un des leviers les plus performants et surtout les mieux adaptés pour la prise de décision. En effet, le forage de données représente de nos jours un outil opérationnel très important, qui permet d'extraire des connaissances à partir

d'une ou de plusieurs bases de données touchant pratiquement une multitude de domaines. La technique du forage de données [1] devient très utile et applicable dans tous les domaines que ce soit dans l'industrie, le domaine de la médecine, de la bioinformatique...

Dans le cadre de ce projet, nous avons opté pour l'approche du forage de données pour résoudre le problème de prédiction de la maintenance des moteurs d'aéronefs.

Dans l'optique de prédire des événements futurs comme pour notre problème de maintenance de moteurs d'aéronefs, le forage de données s'avère très utile vu la quantité astronomique de données conservées sur chaque modèle de moteur. Les travaux récents de classification et de prédiction portent principalement sur la réduction des problèmes de décisions. Les algorithmes d'arbres de décision [1] sont fréquemment utilisés dans la détection ou la prédiction de fraude [9], on les utilise aussi dans la prédiction des fonctions de gènes [10]. H Willenbrock et al. [11] utilisent les KNN pour la prédiction des réactions aux traitements de la leucémie. Dans plusieurs travaux de classification, les techniques du SVM [12,13,14] ont prouvé leur performance et leur efficacité. Plusieurs approches ont tenté de réduire le nombre de vecteurs de support (SV) à utiliser.

Dans ce travail de recherche, qui est approprié pour un grand volume de données d'entraînement, nous proposons une approche qui consiste à approximer les problèmes SVM de grande taille par la décomposition du problème original en des sous problèmes linéaires. Cette approche est particulièrement opérationnelle dans les cas où le problème à traiter se prête bien à une décomposition en sous problèmes. Nous mettons en œuvre aussi bien les méthodes pour les sous-problèmes linéaires que les méthodes pour les sous problèmes non linéaires servant à décomposer le problème original.

Étant donné que les classificateurs SVM sont binaires, nous proposons un modèle hiérarchique de classification SVM qui supporte les multi-classes. De plus, la structure d'arbre de la classification SVM permet d'identifier les relations qui existent entre les classes dans les nœuds de l'arbre.

## **1.4 Cycle de vie adopté pour analyser les données**

Nous nous sommes basés sur la méthodologie CRISP-DM [1], qui a été utilisée tout au long des travaux sur notre modèle de prédiction et de classification des visites moteurs, afin d'effectuer les recherches dans un cadre structuré et bien défini. Cette méthodologie présente le cycle de vie pour un projet de forage de données et elle se divise en six étapes:

1. **La phase de compréhension métier:** prendre connaissance avec le domaine d'application, se familiariser avec le vocabulaire et le sens donné aux termes utilisés, afin de sélectionner ou segmenter les données qui seront utilisées;

2. **La phase de compréhension des données:** faire le ménage des données recueillies, les compléter, vérifier qu'il n'y ait pas d'erreurs ou de redondances. On appelle aussi cette étape le « prétraitement »;

3. **La phase de préparation des données:** préparer les données pour l'étude statistique. Souvent ici, on va normaliser les données. À l'aide de quelques calculs statistiques, on va souvent découvrir sur quelles variables on fera le forage. On appelle aussi cette étape la « transformation »;

4. **La phase de modélisation:** prendre les données normalisées et appliquer les algorithmes de forage. Il existe plusieurs algorithmes et le choix de l'algorithme est très important. Tout dépendamment de la tâche que l'on souhaite accomplir, mais aussi du domaine d'application de cette tâche et des résultats attendus. Il s'agit de l'étape de forage de données, proprement dite;

5. **La phase d'évaluation:** interpréter les résultats et valider l'étude, afin de découvrir si les connaissances trouvées répondent à ce que l'on voulait savoir et si les résultats sont valables. Ainsi, on doit valider si les connaissances recueillies peuvent servir à la prise de décision ou non;

6. **La phase de déploiement:** déployer le projet dans l'entreprise pour rendre les connaissances accessibles aux décideurs.

Toutes les étapes ont été réalisées, de façon itérative, jusqu'à trouver un modèle qui satisfasse les conditions initiales proposées, sauf la phase finale de déploiement.

Les résultats se veulent convaincants, quoi qu'il reste encore beaucoup de travail à faire pour adapter l'algorithme aux nombreux domaines de la vie courante, mais il s'agit bien d'un pas dans une toute nouvelle direction, permettant maintenant de supporter les problèmes multi-classes à l'aide des SVMs pour la prédiction de la maintenance des moteurs d'aéronefs.

Tout au long ce document, deux termes techniques reviendront assez souvent, à savoir : « Classification » et « *Support Vector Machine* », leur définition préalable permettra sans doute de mieux situer le sujet traité. Notre objectif consiste à les étudier en détail puisqu'ils sont des concepts fondamentaux et spécifiques à notre champ d'étude.

Nous commençons par présenter dans le chapitre 2, les bases du forage de données. Dans le chapitre 3, nous présentons les différents travaux connexes. Suivront les aspects théoriques du SVM, en démontrant les notions de base du SVM ainsi que le problème d'optimisation et la construction de l'hyperplan. Dans le chapitre 4, nous présentons notre solution au problème de maintenance de moteurs d'aéronefs. Dans la section 5, nous illustrons notre étude théorique par un cas

pratique sur un échantillon de nos données : un cas de classification binaire et un autre tertiaire. Nous concluons, dans la section 6, en discutant les grands axes à suivre pour une amélioration de l'implantation de notre modèle.

## **Chapitre 2 – Concepts fondamentaux du forage de données pour la classification et la prédiction**

### ***2.1 Qu'est-ce que le forage de données?***

Depuis que les coûts d'entreposage de données ont diminués, les compagnies ont commencé à accumuler des données. Ces compagnies utilisaient l'analyse statistique traditionnelle pour répondre aux besoins des décideurs, mais avec autant de données, la tâche se complexifiait. Elles ont alors introduit de nouvelles méthodes d'apprentissage machine, mais les besoins en capacité de traitement de l'information ont crû eux aussi. C'est ainsi qu'elles ont du faire appel au forage de données.

Avec autant d'information, le problème est maintenant de trouver quoi faire de cette information. Il est bien connu que l'information se situe au cœur même des opérations d'affaires et que les décideurs doivent y avoir accès rapidement lorsqu'il est question de prendre des décisions, de planifier des projets à l'avance ou tenter de dénicher ce qui pourrait les amener au devant de leurs compétiteurs.

Le forage de données « *Data Mining* » [1] est un processus par lequel on tente d'extraire des connaissances, par des méthodes statistiques, mathématiques et informatiques (surtout des algorithmes d'I.A.), sur de larges quantités d'informations. Habituellement, on tente d'extraire des connaissances intrinsèques



aux données qui exprimeraient les relations existantes entre les éléments présents dans la base de données. Un autre terme est également utilisé: la découverte de connaissances dans les bases de données (KDD; « *Knowledge Discovery in Databases* ») [15,16,17].

Dans les étapes du processus de forage de données [1,15], décrites dans le chapitre 1.4, le cycle de vie d'un projet de forage de données est généralement composé des étapes:

1. La phase de compréhension métier;
2. La phase de compréhension des données;
3. La phase de préparation des données;
4. La phase de modélisation;
5. La phase d'évaluation.

Selon les données qui seront traitées, il existe plusieurs types de DM [1]. Le « *Data Mining* » de base, qui s'exécute soit sur: des bases de données, des entrepôts de données (« *Data Warehouse* ») [15] ou encore des magasins de données (« *Data Marts* ») [18]. Le « *Web Mining* » [19,20,21], qui s'exécute directement sur le web, sur les pages web, les sites, les engins de recherche. Finalement, le « *Text Mining* » [22], qui s'exécute sur du texte (généralement non formalisé) qui n'est pas organisé.

En d'autres mots, le forage de données (« *Data Mining* ») utilise des algorithmes d'intelligence artificielle ainsi que des outils statistiques pour effectuer des tâches de segmentation, de prédiction, de classification et d'associations afin d'aider l'industrie à prendre de meilleures décisions [23].

Il n'y a pas de limites aux domaines où le « *Data Mining* » peut s'appliquer. Du moment que nous avons une large quantité de données (historiques ou non), des relevés de transactions, des formulaires remplis, etc. Par exemple, dans le domaine

commercial, on analyse principalement les habitudes de consommation des clients, dans le but de trouver des catégories de clients, des associations de produits, des prédictions d'achats ou de rétention. Dans le domaine médical, on peut prévoir quel type de médicament donner à un nouveau patient, répertorier les diagnostics et les classer. Dans le domaine de l'astronomie, on tente de répertorier les nouvelles découvertes, comme par exemple, tenter de classer les nouvelles galaxies par rapport à celles connues et prévoir où on pourrait en trouver des similaires dans l'univers, en considérant les caractéristiques de celles déjà répertoriées. Même jusqu'au domaine Musical, aidant dans la création de compositeurs automatiques, d'association d'émotions aux extraits musicaux, pour classer également des œuvres musicales ou même suggérer des œuvres sur une base d'œuvres connues. Il y a beaucoup d'autres domaines possibles...

Parmi les techniques suggérées, celles se rapprochant le plus de notre domaine d'application, à savoir prédire les visites des moteurs d'aéronefs en atelier, notons la classification et la prédiction (prévision/estimation).

### **2.1.1 La classification**

On définit la classification comme étant le regroupement d'enregistrements ou d'observations similaires. On peut employer les mêmes méthodes (dites supervisées, où l'algorithme doit apprendre par des exemples pré-classifiés) que celles utilisées pour prédire ou estimer. On peut également, utiliser des méthodes dites non supervisées, telles que la classification hiérarchique des k moyennes ou les réseaux de kohonen. Cette technique est surtout utilisée sur les données d'entrée pour une autre méthode (mixte), par exemple : les réseaux de neurones. La classification est différente de la segmentation, dans la mesure où il n'y a pas de variable cible [1].

## 2.1.2 La prédiction

Lorsque l'on effectue des prédictions, les résultats portent sur le futur. Les mêmes méthodes que pour la classification et l'estimation sont utilisées. Il s'agit de méthodes d'apprentissage dites **supervisées** [1].

## 2.2 Quelles sont les bases du forage de données?

Le forage de données a été créé par la conjonction de quelques champs d'intérêts tels que l'apprentissage inductif, l'apprentissage machine et les statistiques.

### 2.2.1 L'apprentissage inductif

L'apprentissage inductif consiste à regrouper des objets similaires dans des classes et formuler des règles de classifications pour des nouveaux objets. Il s'agit d'un modèle d'apprentissage permettant de créer un processus qui analyse les données (par une vue sur une base de données) afin de trouver des « *patterns* ». Il se divise en deux principales stratégies: l'apprentissage supervisé et l'apprentissage non-supervisé.

#### 2.2.1.1 Apprentissage supervisé

Dans l'apprentissage supervisé, l'algorithme apprend par des exemples pré-classifiés, ou toutes les règles sont définies, afin de pouvoir classer de nouveaux exemples par la suite. L'algorithme doit trouver les propriétés distinctes d'une classe, lesquelles sont ensuite utilisées pour classer de nouveaux objets. L'analyse discriminante linéaire, la méthode des K plus proches voisins ou les SVM sont des exemples typiques d'algorithmes d'apprentissage supervisé.

Considérant les hypothèses et formules de l'analyse discriminante linéaire [26]. Nous disposons d'un échantillon de  $n$  observations réparties dans  $K$  groupes

d'effectifs  $n_K$ . Notons  $Y$  la variable à prédire, elle prend ses valeurs dans  $\{Y_1, \dots, Y_K\}$ . Nous disposons de  $J$  variables prédictives  $X = (X_1, \dots, X_J)$ . Nous notons  $\mu_K$  les centres de gravité des nuages de points conditionnels,  $W_K$  leurs matrice de variance-covariance. En utilisant la règle bayésienne, l'objectif est de produire une règle d'affectation  $F : X \Rightarrow \{Y_1, \dots, Y_K\}$  qui permet de prédire, pour une observation  $\omega$  donnée, sa valeur associée de  $Y$  à partir des valeurs prises par  $X$ . La règle bayésienne consiste en fait à produire une estimation de la probabilité à posteriori d'affectation. Toute la problématique de l'analyse discriminante revient alors à proposer une estimation de la quantité  $P(X|Y = Y_k)$ .

### 2.2.1.2 Apprentissage non-supervisé

Dans l'apprentissage non-supervisé, l'algorithme doit lui-même définir les classes en regroupant les objets similaires, mais aucune définition ne lui est donnée. Il doit reconnaître les « *patterns* » par observation et découverte. L'algorithme doit donc fournir une description des classes rencontrées dans l'environnement présenté (l'ensemble des données d'apprentissage). Les réseaux de neurones, le « *clustering* » et les modèles de mélanges de gaussiennes sont des algorithmes d'apprentissage non supervisés. Le **partitionnement de données** (« *data clustering* » en anglais) est une méthode statistique d'analyse des données qui a pour but de regrouper un ensemble de données en différents paquets homogènes. L'algorithme EM (espérance-maximisation) en est un bon exemple.

Considérons l'algorithme EM [27] et étant donné un échantillon  $X = (x_1, \dots, x_n)$  d'individus suivant une loi  $f(x_i, \theta)$  paramétrée par  $\theta$ , on cherche à déterminer le paramètre  $\theta$  maximisant la log-vraisemblance donnée par l'équation (1).

$$L(\mathbf{x}; \theta) = \sum_{i=1}^n \log f(\mathbf{x}_i, \theta). \quad (1)$$

On peut donc définir l'algorithme EM de la manière suivante:

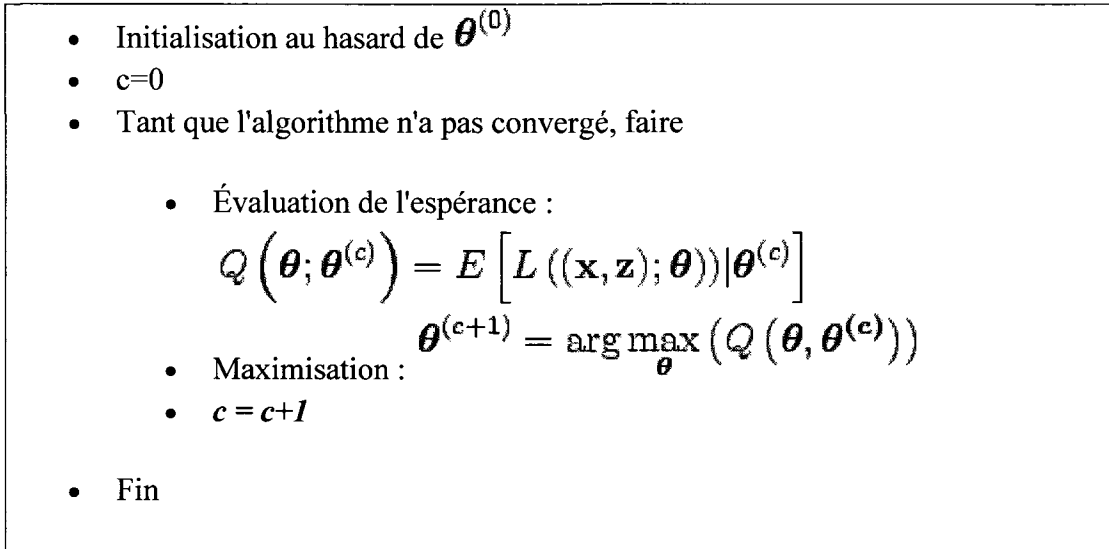


Figure 2 - Algorithme EM

En pratique, pour s'affranchir du caractère local du maximum atteint, on fait tourner l'algorithme EM un grand nombre de fois à partir de valeurs initiales différentes de manière à avoir de plus grandes chances d'atteindre le maximum global de vraisemblance.

### 2.2.2 Les statistiques

Les théories statistiques reposent sur de bonnes bases, mais elles peuvent rapidement devenir difficilement interprétables ou même accablantes. Afin de résoudre notre problème en utilisant les méthodes statistiques, il nous faut faire l'estimation de la distribution de nos moteurs et ensuite estimer les futures distributions possibles pour les prochains moteurs. On peut consulter l'Annexe G, à la p.173 pour approfondir les calculs d'estimation de distributions.

Le forage de données, quant à lui, permet d'inclure les connaissances de l'expert du domaine ainsi que des techniques avancées pour réduire la complexité et augmenter la compréhension du problème. Il existe aussi certains systèmes d'analyse statistique, dont SAS ou SPSS, qui peuvent s'avérer utiles, mais ils créent,

pour la plupart, des modèles linéaires. On peut voir comment le groupe SAS solutionnerait le problème en Annexe H, à la page 175.

### **2.2.3 L'apprentissage machine**

L'apprentissage machine se veut une automatisation du processus d'apprentissage. Il peut y avoir plusieurs types d'apprentissages: L'algorithme apprend par des exemples [30], de l'apprentissage renforcé [18] ou encore avec supervision [1]. Il tente de découvrir les règles de généralisation et pouvoir ensuite les appliquer sur de nouveaux cas. Dans l'apprentissage machine, les données doivent être les plus parfaites possibles et elles sont contenues dans un petit ensemble. En forage de données, par contre, on doit très souvent travailler avec de grandes quantités de données réelles où il peut y avoir des données bruitées, manquantes ou aberrantes.

### **2.3 Les modèles de forage de données**

Il existe, selon IBM [18], deux types de modèles ou modes pour effectuer du forage de données: Le modèle de vérification et le modèle de découverte. Le choix du modèle appartient à l'utilisateur, dépendamment de ce qu'il tente de découvrir.

#### **2.3.1 Modèle de vérification**

Dans un modèle de vérification, l'utilisateur tente de valider une hypothèse sur les données, afin de confirmer ou infirmer l'hypothèse. Toute l'opération peut être raffinée et réitérée, jusqu'à ce que le jeu d'enregistrements retournés soit le plus représentatif possible.

### **2.3.2 Modèle de découverte**

Dans un modèle de découverte, le système doit découvrir automatiquement les tendances et les généralisations sans intervention ou guidance. Le système doit tenter de trouver le plus grand nombre de faits possibles en un temps restreint.

### **2.4 Les entrepôts de données**

Les entrepôts de données [15] fournissent des données déjà totalisées et pré-analysées, contrairement aux bases de données relationnelles, qui elles, rencontrent les besoins transactionnels de la compagnie sans toutefois résumer l'information qu'elles emmagasinent. Les entrepôts de données intègrent souvent des données qui proviennent de plusieurs bases de données, elles sont orientées par sujets et non par application. Ces entrepôts conservent souvent d'anciennes données qui ne seront plus jamais modifiées et qui servent de base de comparaison à travers le temps. Aucune mise à jour n'est effectuée sur les données emmagasinées, mais on ajoute à période fixe de nouvelles données récentes.

On utilise des méta-données pour que l'utilisateur s'y retrouve. Ces méta-données doivent contenir la structure des données, les algorithmes de totalisation et les associations de champs référencés. Il faut s'assurer de retirer les duplicatas et de standardiser les formats provenant de différentes sources.

Il est possible d'utiliser un entrepôt plus petit, comme un magasin de données, contenant des bases de données multidimensionnelles pouvant accélérer le temps de traitement des requêtes les plus fréquemment utilisées. Le Dr. Kamran Parasaye [24] propose une façon plus incrémentale: effectuer un pré-forage pour déterminer les formats des données, créer un mini prototype d'entrepôt de données, réviser la stratégie si nécessaire et implanter l'entrepôt final à partir de l'expérience précédente.

Les compagnies se sont tellement précipitées dans l'utilisation de logiciels de forage de données, qu'ils ont du faire face à plusieurs problèmes de précision, de fiabilité et de disponibilité de systèmes. Ainsi, C. Erickson [25] propose une liste de critères pour éviter ce genre de problèmes:

- Mesurer la performance de chargement en centaines de millions de lignes par heure;
- La procédure de chargement doit se faire en une étape, demandant la qualité de données la plus haute possible;
- Les requêtes doivent être performantes pour les opérations clés de l'entreprise et la performance des requêtes ne doit pas être liée à la taille de la base de données;
- Le système doit être utilisable par plus d'utilisateurs que l'élite seule, donc il faut que le serveur inclue des outils d'extraction de données de diverses sources;
- Le système doit permettre une priorisation des requêtes pour surveiller les surcharges et de plus, le support aux bases de données multidimensionnelles doit être présent et efficace;
- Le système doit également inclure un jeu d'opérations analytiques avancées.

## **2.5 Les problèmes du forage de données**

Certains problèmes ont été mentionnés précédemment, comme les données bruitées, manquantes ou aberrantes. Les données peuvent aussi être inadéquates ou non pertinentes pour résoudre le problème. La quantité de données utilisées peut



également devenir un problème, surtout lorsqu'on travaille avec de très larges bases de données.

De plus, il faut s'assurer que les attributs sélectionnés sont des données essentielles à la compréhension du domaine, sinon il ne sera peut-être pas possible de découvrir de nouvelles connaissances. Les bases de données accumulées contiennent souvent plusieurs erreurs. Cette contamination risque de se transmettre dans le modèle et de biaiser les résultats.

Il faut retirer les enregistrements problématiques, ignorer les champs manquants ou encore associer des valeurs réelles aux valeurs manquantes ou aberrantes. Tout cela, sans affecter la qualité ou la quantité de données nécessaires au bon fonctionnement de l'algorithme. On peut également effectuer une analyse statistique pour découvrir les valeurs extrêmes, nulles ou manquantes.

## ***2.6 Notions importantes sur les SVM***

Dans la pratique, le SVM est l'une des méthodes les plus performantes et les plus faciles à utiliser surtout pour le processus de classification, selon Hsu et al. [28]. Le SVM est conçu pour effectuer une séparation binaire, mais il est également possible de l'utiliser pour séparer trois classes d'éléments.

La technique du SVM est basée sur deux principes. Le premier principe vise à minimiser l'erreur empirique de classification et le deuxième vise à maximiser la marge entre les frontières des deux classes séparées.

À l'aide de cette technique, nous tentons de trouver un hyperplan qui permet de regrouper les données similaires et de séparer les données hétérogènes. Cela nous conduit à déterminer les fonctions discriminantes qui permettent de résoudre le problème.

La figure 3 ci-dessous illustre bien le principe du SVM.

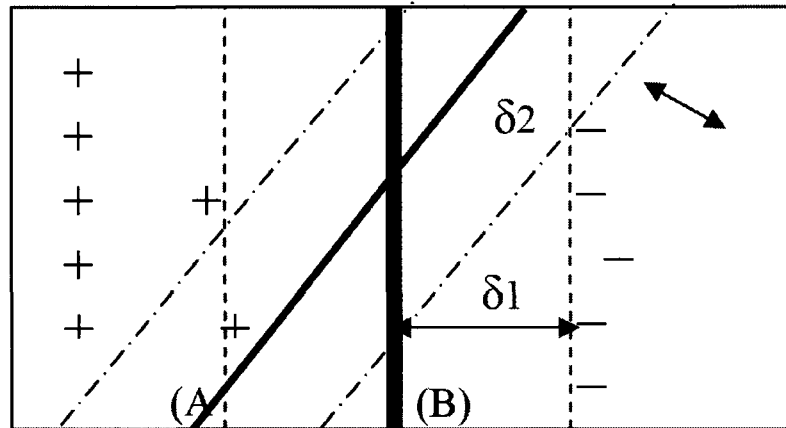


Figure 3 - Plusieurs hyperplans séparent deux classes de données

La figure 3 ci-dessus, montre qu'il existe une infinité de séparateurs de classes, avec l'idée de choisir le meilleur hyperplan qui assure une bonne séparation. Une bonne séparation signifie la recherche d'un hyperplan qui maximise la marge entre les deux classes.

Le SVM permet de résoudre des problèmes linéaires et d'ordre binaires. Un problème linéairement séparable signifie qu'il existe un séparateur linéaire capable de discriminer les données en deux ou plusieurs classes. La marge entre ces deux classes doit être maximale. Il s'agit donc de trouver les paramètres qui satisfassent cette condition.

Lorsqu'on est en présence de problèmes linéairement séparables, il est facile de déterminer la fonction de discrimination. Toutefois, une question importante doit être adressée : comment la technique du SVM se comporte-t-elle devant un problème non linéairement séparable? C'est à ce niveau où les fonctions de « *Kernel* » interviennent. On doit alors effectuer la technique de « *mapping* » sur un

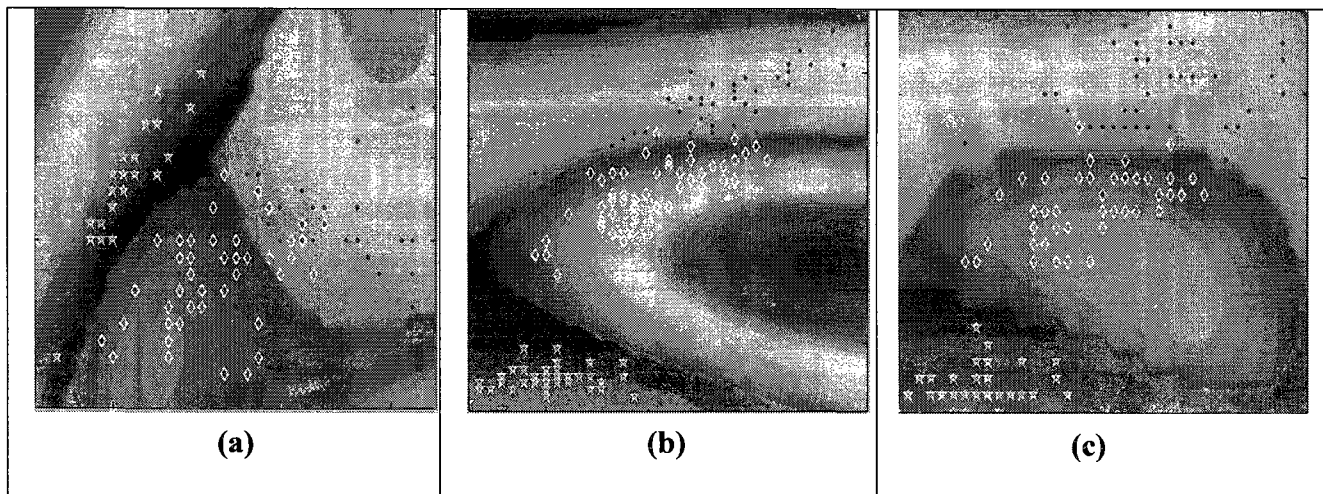
autre plan. La technique de « *mapping* » rend la séparation plus facile dans un espace de grande dimension.

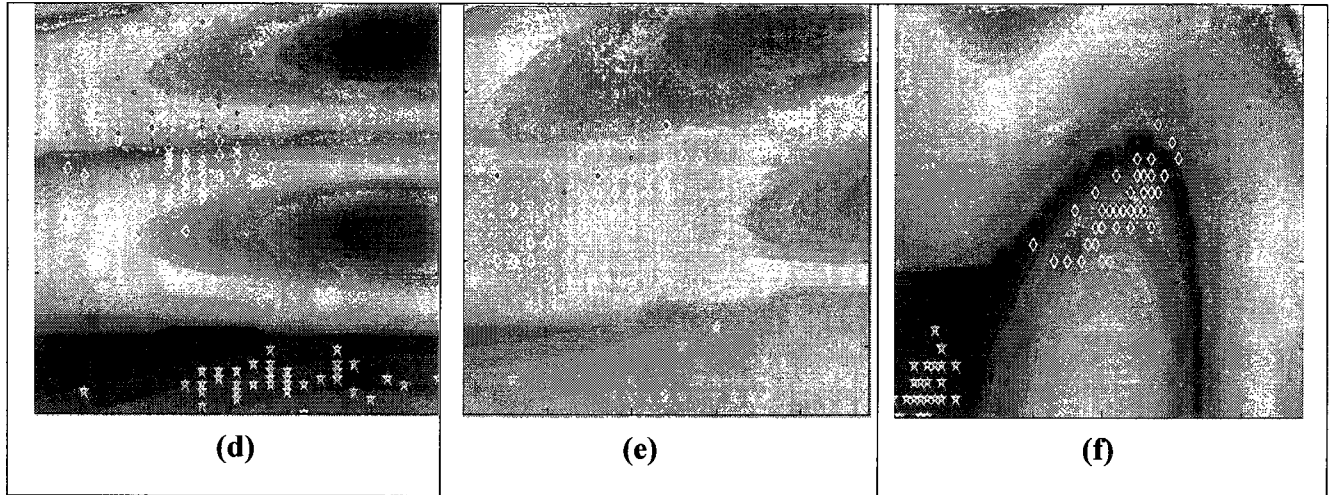
## 2.6.1 Cas binaire et ternaire – les explications théoriques de l'approche choisie

### 2.6.1.1 Cas de trois classes

La technique SVM ne se limite pas seulement à la classification binaire. Elle traite également les cas de classification multi-classes. Dans cet exemple, nous allons aborder un cas où le nombre des classes est égal à trois. L'ensemble de données utilisé est IRIS [31]. Notre but est de classifier trois classes de fleurs (Setosa, Versicolor et Virginica). La séparation se fait dans un espace à deux dimensions. Nous avons combiné les variables deux à deux pour déterminer les variables dominantes. On appelle variables dominantes celles qui sont le plus facilement séparables dans les ensembles, celles qui seront au sommet de la hiérarchie des classes. Les résultats obtenus sont représentés dans la table 1 ci-dessous.

**Table 1 - classification d'un ensemble de données IRIS : Cas trois classes**





On a obtenu trois classes séparées. Dans la table 1 (f), les classes sont bien séparées. Donc les deux variables combinées influent d'une façon remarquable dans la séparation des données.

Le cas qui a été décrit dans cette section est similaire à notre problème de classification de moteurs d'aéronefs. Dans notre projet, nous souhaitons classifier les types de visites de maintenances des moteurs en atelier. En se basant sur l'historique de notre base de données, trois types de maintenance sont déterminés. Une maintenance planifiée, une maintenance non planifiée et une maintenance concurrente. Nous essayons ultérieurement de les classifier en appliquant la technique du SVM.

L'un des problèmes majeurs de la technique du SVM est bien la lenteur du temps d'apprentissage lors la résolution du problème quadratique. Plusieurs recherches ont été réalisées par Hsu et al., puis Xiaoujuan et al. [28,29] dans le but de réduire le temps d'entraînement surtout dans les bases de données de grandes dimensions. Intuitivement, Le temps d'apprentissage augmente avec la complexité du problème. Les recherches ont démontré que la complexité de l'algorithme du SVM varie :

$$n^2 \leq \text{Complexité} \leq n^3$$

Où  $n$  représente le nombre d'entrées.

Cela nécessite un temps énorme pour parcourir tous les enregistrements. L'une des méthodes utilisée pour résoudre ce problème est basée sur la réduction des données d'apprentissage. L'idée de cette technique est de trouver une fonction de décision qui n'exploite qu'un sous ensemble de l'ensemble d'apprentissage au lieu de tout l'ensemble.

## **Chapitre 3 – Travaux connexes**

Dans ce chapitre, nous explorerons les approches disponibles pour résoudre les problèmes de classification et de prédiction en forage de données lorsque nous faisons face à un nombre élevé de données. Les techniques de forage de données ont été retenues, au détriment des techniques de recherche opérationnelle ou des systèmes experts, qui se veulent beaucoup moins flexibles et maintenables. En forage de données, les problèmes de classification et de prédiction utilisent plusieurs méthodes ou algorithmes pour résoudre les problèmes confrontés, nous avons retenus ceux qui s'apparentent le plus à notre description de problème. Nous regarderons les solutions proposées à l'aide de la logique floue [32,36,37,38], des arbres de décision [1,39,40], de la méthode des K plus proches voisins (KNN) [1,3,42,43,56,57], ainsi que les Machines à Vecteurs de Support (SVM) [3,5,6,7,12,13,14].

### ***3.1 La logique floue***

La logique floue est l'une des techniques utilisées surtout en intelligence artificielle, dans les domaines de l'automatisme, de la robotique, comme la gestion des feux de circulation par exemple, du contrôle aérien, de la médecine, de la météorologie [51], de la climatologie [52] et de la sismologie [53]. Elle a été formalisée par Lofti Zadeh [50]. Elle se veut une généralisation de la logique

booléenne classique, amenant une toute nouvelle zone de valeurs d'amplitude tenant compte d'ensembles définis de façons imprécises. Elle s'appuie sur la théorie mathématique des ensembles flous, induite par Zadeh, qui a su montrer que sa théorie des sous-ensembles flous se réduisait effectivement à la théorie des sous-ensembles classiques pour la prise en compte d'ensembles flous [32].

Idri et Abran [5] ont utilisé la logique floue pour modifier les modèles d'estimation d'efforts, surtout le modèle COCOMO'81, qui utilisait la logique traditionnelle. Ils utilisent la logique floue pour qualifier des valeurs linguistiques telles que « trop cher », « cher », et ainsi de suite. Ils définissent ainsi quatre classes contenant 15 facteurs d'évaluation similaires et utilisent COCOMO'81 pour résoudre l'équation. Les ensembles flous ont été représentés par des fonctions d'appartenance trapézoïdales [54]. Finalement, en utilisant la logique floue sur le modèle linguistique COCOMO'81, l'algorithme devient moins sensible aux variations dans les entrées et cette stabilité est très appréciable dans ce type de problèmes, puisque l'algorithme originel ne tolère pas les variations.

Dans un autre cadre, Barra et al. [37] proposent un algorithme de segmentation multi-spectrale d'images du Soleil acquises en extrême ultraviolet (EUV), utilisant un algorithme de classification flou spatialement contraint. L'utilisation de la logique floue permet de prendre en compte les imprécisions et les incertitudes inhérentes à la définition des différentes régions d'intérêt dans l'image. Ainsi, la méthode obtenue à l'aide de la logique floue permet de discriminer de manière satisfaisante les différentes régions d'intérêt de l'atmosphère solaire, à partir d'images EUV. Les méthodes issues de la reconnaissance des formes, appliquées à des données multidimensionnelles ou multi-spectrales, ont démontré leur potentiel en segmentation d'images. Dans l'algorithme présenté et les contraintes rédigées en logique floue, l'idée sous-jacente est la minimisation de la variance totale inter-classe. Les chercheurs en tirent des séries temporelles qui correspondent aux observations télescopiques. La méthode proposée ici peut alors

devenir un composant essentiel dans une méthode de reconstruction du spectre solaire.

Afin de gérer les réserves en eau, Faye et al. [38] se sont penchés sur une approche globale de gestion à long terme pour le stockage, le transfert et la distribution d'eau. L'objectif principal est de minimiser l'écart entre l'offre et la demande. Ils proposent une approche contenant une procédure d'adaptation des pondérations du critère fondée sur la logique floue. Il s'agit donc d'un problème d'optimisation, comme on voit dans l'approche de la recherche opérationnelle. Leur technique aboutit à une solution pratique adaptative pour la modélisation du problème d'optimisation, puisqu'elle tient compte des contraintes opérationnelles du système et des objectifs en termes d'optimisation. L'approche a été appliquée à un système de complexité moyenne, afin de vérifier s'il pouvait constituer un outil d'aide à la décision efficace et aussi pour observer la lourdeur de la mise en œuvre.

### ***3.2 Les arbres de décision***

Ce qui nourrit à la fois la Statistique et l'Informatique en ce domaine, ce sont les méthodes de classification hiérarchique, ainsi que la personnalisation de la nomenclature binomiale.

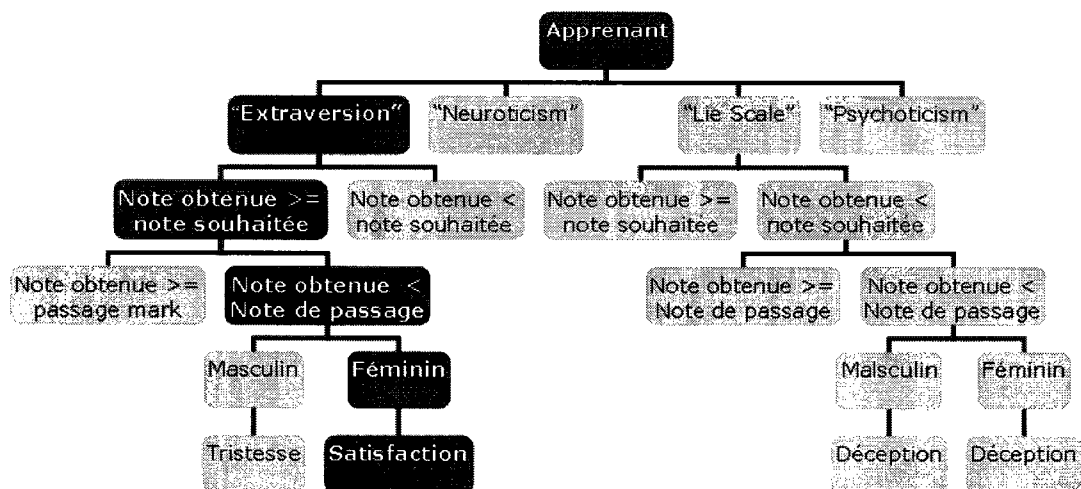
Ce qui nous intéresse dans les méthodes de classification hiérarchique, c'est la capacité d'automatiser la répartition des éléments d'un ensemble en groupes. Cette répartition en groupes doit se faire de sorte que les groupes soient le plus homogènes possibles. En bio-statistique, on représente la classification hiérarchique sous la forme d'un arbre binaire appelé le dendrogramme [55].

Hoare [39] tente de démontrer que CHAID (un des principaux algorithmes utilisés par les arbres de décision) peut s'avérer très utile dans la résolution de problèmes non-linéaires, ou de problèmes comprenant des jeux de données



complexes, afin de découvrir des « *patterns* » significatifs dans des données commerciales ou scientifiques. Il a utilisé CHAID sur des données financières réelles de la Nouvelle-Zélande afin de découvrir dans un bassin de clients lesquels pourraient être classés comme de mauvais clients. Malheureusement, les données ne contenaient pas d'information pouvant servir à déduire les règles définissant les mauvais clients. Même si les mesures du taux d'erreurs sont pertinentes, certaines données sur les banqueroutes viennent fausser les résultats.

Chaffar et al. [40] ont tenté de détecter la réaction émotionnelle d'un apprenant pour une situation donnée. Il s'agit là d'un élément essentiel dans les environnements d'apprentissage à distance. Des travaux existent dans ce domaine, présentés par Chaffar et al. [40], en utilisant une technique d'apprentissage machine, à savoir l'algorithme ID3. Le système « APRE » a été mis en ligne et a été évalué par cent trente-neuf participants. Les chercheurs définissent l'APRE comme l'Agent de Prédiction de Réaction Émotionnelle. Cet agent est conçu pour prédire la réaction émotionnelle de l'apprenant suite à l'obtention de sa note dans un test d'évaluation en ligne. La réaction émotionnelle fournie par chacun des participants à la fin de l'évaluation a été comparée à celle prédite par notre système. Ici, ID3 a très bien performé.



### 3.3 Les KNN (Kppv: K plus Proches Voisins)

La technique des K plus proches voisins (« *K-Nearest Neighbors* »; KNN) [1,42,43,56,57] est une technique qui classe chaque enregistrement d'un ensemble de données en se basant sur l'historique de classification des k-1 autres enregistrements. L'élément sera classé dans la classe contenant le maximum d'enregistrements semblables. L'algorithme se base sur une fonction de distance euclidienne, ou encore une fonction cosinus afin de jumeler des paires d'observations.

Il faut d'abord examiner comment l'algorithme se comporte sur un jeu de données test (ou d'apprentissage), pour vérifier que les futures données (inconnues) soient bien classées. On applique souvent l'algorithme sur un ensemble de validation, où la variable cible est cachée temporairement au modèle. Le modèle est reconnu pour son efficacité quant aux classes multi-modales puisqu'il se base sur un voisinage restreint d'enregistrements.

On utilise, par exemple, la méthode des K plus proches voisins pour retrouver dans une banque d'images celles qui sont similaires. Quand une image requête  $R_I$  est présentée au système, l'algorithme suivant est appliqué :

- 1- Calculer les p caractéristiques de bas niveau de l'image requête  $R_I$ , qui correspondent à l'espace de représentation  $R_p$ ;
- 2- Calculer les distances entre le nouveau point et tous les points existants.
- 3- Insérer le nouveau point dans le graphe topologique, en vérifiant la condition suivante:  $d(\alpha, \beta) \leq \text{Max}(d(\alpha, \gamma), d(\beta, \gamma)), \forall \gamma \in \Omega \setminus \gamma \neq \alpha, \beta$

Figure 5 - Algorithme pour le classement d'images

Il est à noter qu'afin de comprendre l'équation de la troisième étape dans la figure 5, présentée ci-haut, nous devons préciser les variables suivantes: soit un ensemble  $\Omega$  composé de  $n$  images. Chaque image est représentée par un vecteur de caractéristiques de dimension  $p$ , qui est un point dans l'espace  $\mathbf{R}^p$ , tel qu'exprimé dans l'équation (2) ci-dessous:

$$X(i) = (X_1(i), \dots, X_j(i), \dots, X_p(i)) \in \mathbf{R}^p. \quad (2)$$

Les points  $\alpha$  et  $\beta$  sont reliés par une arête s'ils vérifient la condition proposée dans un graphe des voisins relatifs. La condition vérifie si les deux points sont reliés par une arête. L'algorithme performe assez bien, mais il est surpassé par l'algorithme du graphe topologique [41] utilisé dans leur système appelé « *Topological Image Query (TIQ)* ».

Dans le domaine de la prédiction de structure des protéines, Maupetit [42] présente les méthodes de type plus proches voisins utilisées sur des structures de protéines disponibles pour inférer la structure secondaire par comparaison de fragments [57]. Dans le même domaine, et plus récemment, de nouvelles méthodes basées sur les SVM (« *Support Vector Machine* ») ont été proposées par Hu et al. [58]. Leur étude comparative de plusieurs méthodes de prédiction vente bien les mérite des méthodes hybrides, car dans la prédiction des structure de protéines ni le SVM, ni les Kppv ne sont utilisés tels quels, mais les techniques hybrides présentées qui les utilisent fonctionnent bien.

Boufadel et al. [43] ont mené des expériences sur 1027 dialogues Personne-Machine [59] dans le domaine des agences de voyage. Ils testent ces combinaisons avec trois algorithmes de classification : « *Support Vector Machine* » (SVM), kppv (kNN) et un arbre de décision (DT). Ils effectuent la Détection des dialogues problématiques, suivant un pré-classement positif ou négatif de dialogues de l'historique d'appels. Les résultats pour le modèle kppv sont sensiblement les

mêmes pour toutes les combinaisons d'attributs, alors qu'une petite amélioration est observée pour l'arbre de décision (DT) et le SVM. Enfin, bien qu'ils n'aient pas amélioré le résultat de la classification obtenu avec les mesures d'efficacité, ils ont obtenu un meilleur taux de classification de 91,8% avec le modèle kNN.

### 3.4 Les SVM

L'approche du SVM permet de résoudre des problèmes linéaires et d'ordre binaire. Un problème linéairement séparable signifie qu'il existe un séparateur linéaire capable de discriminer les données en deux ou plusieurs classes. La marge entre ces deux classes doit être maximale. Il s'agit donc de trouver les paramètres qui satisfassent cette condition.

Vino et al. [44] utilisent des algorithmes de classification automatique pour détecter le contenu raciste sur l'Internet. Chaque document  $W$  est représenté par un vecteur  $[d]$  de  $R^n$  où chaque coordonnée  $d_w$  est calculée par rapport à la fréquence  $Occ(w,d)$  du terme  $w$  dans  $d_w$  selon l'équation (3) :

$$d_w = TFIDF(w,d) = \log(1 + Occ(w,d)) * \log\left(\frac{N}{N(w)}\right) \quad (3)$$

Où  $N$  est le nombre de documents du corpus et  $N(w)$  est le nombre de documents dans lequel  $w$  apparaît au moins une fois. Pour effectuer la normalisation  $\underline{d_w}$ , on divise chaque coordonnée  $d_w$  par la norme euclidienne du vecteur, comme dans l'équation (4) :

$$\underline{d_w} = \frac{d_w}{\sqrt{\sum_w d_w^2}} \quad (4)$$

Ces valeurs sont ensuite traitées par les algorithmes de classification suivants : Rocchio [60], kppv [1,3,42,43,56,57] et SVM [3,5,6,7,12,13,14]. Dans

leurs travaux, ils ont comparé l'algorithme des Kppv et le SVM. Il semble que le SVM performe mieux dans la plupart des situations.

En Chine, la classification des questions chinoises joue un rôle important pour des systèmes servant à répondre aux questions dont les réponses sont enfouies dans des documents. Bo et al. [49] utilisent le SVM pour sa capacité supérieure à reconnaître les « *patterns* ». Plus spécifiquement, ils utilisent la version du SVM appelée « *One-against-All* » [61]. Les résultats démontrent que l'algorithme performe de façon excellente, sur des problèmes binaires. Ils précisent qu'il n'est plus nécessaire de connaître la forme de la distribution des exemples fournis et qu'il s'agit d'un bon algorithme pour résoudre ce type de problèmes. Ce groupe de recherche ajoute qu'il est possible de transformer un problème multi-classes en un problème binaire.

Shina et al. [45] ont tenté de déduire, d'un large ensemble de données marketing, lesquels seraient les clients les plus susceptibles d'adhérer à une offre d'achat de produit, en utilisant le SVM. L'idée est d'utiliser l'algorithme NPPS (« *Nuclear Power Plants System* ») [62], qui permet d'aller chercher seulement les « *patterns* » qui se situent autour de la frontière de décision entre les classes, puisqu'ils sont susceptibles de contenir plus d'information. Ces derniers sont ensuite utilisés pour entraîner un SVM. Ils ont quand même tenté l'expérience avec la méthode des K plus proches voisins, mais l'algorithme était beaucoup trop lent à s'exécuter.

Shina et al. [45] introduisent l'échantillonnage informatif ainsi que différents coûts pour chaque classe et réussissent à résoudre le problème avec de très bons résultats. Deux de leurs algorithmes, qui sont des variations du SVM, s'affrontent au niveau de la performance: S-SVM [45] et R\*SVM [45].

Afin d'effectuer de la classification spatiale, Lee et al. [46] ont utilisé les SVM qu'ils ont modifiés. Le SVM effectue des modèles explicites de corrélations

entre les données multidimensionnelles, sous le principe que les données répondent au critère « iid » [66] i.e. Les données sont indépendantes et distribuées identiquement. Lorsque l'on fait face à des exceptions sur cette règle (lorsqu'on a des données spatialement dépendantes par exemple), alors l'algorithme réagit plutôt mal et ne donne pas de résultats valables. L'algorithme développé, le SVRF [46], est inspiré de l'approche du SVM et du CRF (« *Conditional Random Fields* ») [46]. L'algorithme performe très bien et dépasse l'algorithme auquel ils le comparent (le DRF - « *Discriminative Random Fields* »). En fait, le CRF permet de trouver les voisins adjacents immédiats, lorsqu'on étend l'algorithme avec le DRF, alors on ajoute une deuxième dimension au voisinage. En combinaison avec le SVM, ceci permet de modeler les dépendances spatiales non triviales bidimensionnelles.

Pour Sibon et al. [47], la classification SVM semble donner de meilleurs résultats encore une fois, comparativement aux arbres de décision. On s'attarde à la prédiction de la capacité d'un système à répondre à une question factuelle. Ces prédictions permettent de déterminer si l'on doit initier un dialogue afin de préciser ou si l'on doit reformuler la question posée par l'utilisateur. En comparant les deux algorithmes sur des types de questions avec les classes correspondantes aux résultats pondérés, c'est ce qu'ils ont obtenus:

« <i>Feature</i> »	SVM	Arbres de Décision
toutes	68,5%	62,4%
questions	69%	64,1%
documents	53,8%	49,9%
passages	52,3%	50,1%
aucune	51,6%	

**Table 2 - Résultats de la classification automatique, en pourcentage de bonne prédiction**

Comme peu de travaux présentés en classification et en prédiction abordent les questions de maintenance de moteurs, les travaux rencontrés présentent plusieurs autres techniques, comme celle du système expert. Jafar et al. [48] se sont penchés sur le problème de maintenance d'aéronefs utilisés dans la marine américaine. Ils ont utilisé la théorie des contraintes (TOC) [63], le « *Just-in-time* » (JIT) [64] et le Six-Sigma (règles de l'amélioration continue dans la gestion des opérations liées aux maintenances d'aéronefs) [65]. Le résultat est un processus du nom de « *AIRspeed* », qui parvient à faire diminuer les coûts d'opération et de maintenance. Le but étant de faire passer la durée du cycle de maintenance de 468 heures à 233 heures. Beaucoup d'efforts furent déployées pour créer le processus, mais ici on ne s'en tient qu'à un seul type d'aéronefs, ceux utilisés dans la marine américaine. Il est donc clair que l'approche du système expert s'avérait une solution adéquate, puisqu'ils ne comptaient pas plusieurs types de modèles d'aéronefs différents, contrairement à notre travail, qui lui, doit supporter plusieurs types de modèles ayant tous leur mission propre (Corporatif, militaire, régional, etc).

Nous avons utilisé la technique des arbres de décision dans la première itération du cycle de vie de notre projet. Cependant, la technique à elle seule ne permettait pas une classification adéquate des événements passés pour pouvoir prédire les événements futurs. Ce fut la première avenue explorée, mais les résultats n'ont pas été satisfaisants.

Nous avons ensuite utilisé la technique des K plus proches voisins dans la seconde itération du cycle de vie de notre projet. Cependant, comme pour les arbres de décision, la technique à elle seule ne permettait pas une classification adéquate des événements passés pour pouvoir prédire les événements futurs. Ce fut la seconde avenue explorée, mais les résultats n'ont pas été satisfaisants.

Nous nous sommes donc inspiré de l'approche présentée pour la classification spatiale (présenté ci-haut dans ce chapitre) afin tenter de résoudre notre problème de maintenance de moteurs d'aéronefs. Étant donné que le SVM

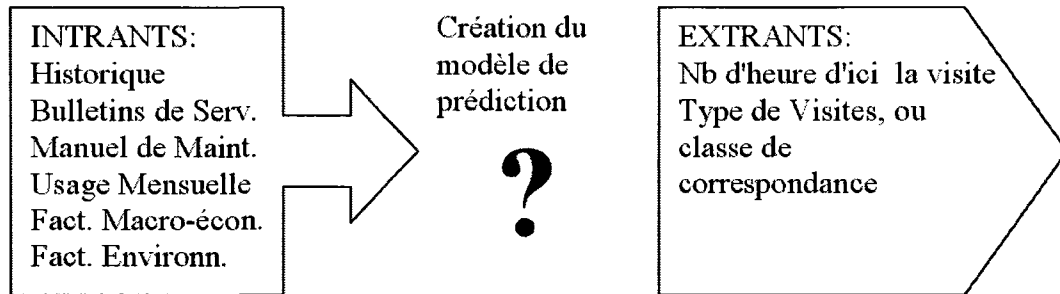
semble obtenir beaucoup de succès, même lorsqu'il est jumelé à d'autres techniques afin de former une nouvelle technique hybride, c'est cette technique que nous avons priorisée. Notre orientation dans le projet, vu les travaux dans le domaine, et vu notre problème à résoudre, s'est dirigée vers une approche mixte qui intégrerait les arbres de décision et l'approche du SVM. Cette orientation provient du fait qu'aucune des techniques présentées, à elle seule, ne parvient à solutionner le problème posé avec un taux d'erreurs qui soit inférieur à cinq pourcents.



## **Chapitre 4 – Approche de classification et de prédiction de maintenance de moteurs**

Notre approche pour résoudre le problème de prédiction des visites de moteurs en atelier se décompose en différentes composantes fonctionnelles, qui sont au nombre de trois, chacune ayant sa responsabilité dans l'accomplissement du travail demandé. Une première composante contient la liste des entrées nécessaires à la résolution du problème. On la nommera « Intrants ». Une seconde composante contient la liste des sorties attendues (le résultat), donnant ainsi une solution au problème proposé. On la nommera « Extrants ». Finalement, la tierce composante détient l'information sur la résolution du problème: les traitements à effectuer, tels que les différents calculs et algorithmes utilisés pour résoudre le problème. On la nommera « Création du modèle de prédiction ».

Dans la Figure 6, nous présentons le processus exprimant l'interaction des trois composantes dans notre approche de résolution du problème proposé:



**Figure 6 - Approche de classification et de prédiction de maintenances de moteurs**

La figure 6 donne une vue globale du processus de prédiction, nous permettant de comprendre comment le futur système fonctionnera. Bien entendu, il nous faudra ensuite bien définir chacun des éléments qui figurent dans chacune des composantes. Ces composantes permettent la compréhension de la pertinence des éléments sur la réponse finale. Considérons la série d'étapes incluses dans ce processus de prédiction:

## 1. Intrants

### a. Sélection des sources de données pertinentes

- i. Voir ce qui se fait manuellement;
- ii. Les maintenances de base et les bulletins de service recommandés;
- iii. Inclure l'historique d'événements passés;
- iv. Inclure les événements fortuits (ce qu'il peut arriver);
- v. Inclure les facteurs macro-économiques et environnementaux influençant le modèle de moteur.

### b. Conversion des valeurs catégorielles;

### c. Normalisation des données

## 2. Création du modèle de prédiction

- a. Création des jeux d'observation: Grouper les intrants des jeux d'entraînement, de validation et de test en sous-groupes de 3 classes distinctes;

- b. Appliquer le SVM sur les jeux d'observation: Soumettre chacun des groupes au SVM et déterminer la séparabilité des classes;
- c. Création de l'arbre de décision hiérarchique, ayant pour nœuds les groupes soumis au SVM, jusqu'à l'obtention de classes distinctes.
- d. Répéter les points b et c sur le modèle créé avec le jeu d'entraînement, mais en utilisant le jeu de validation.
- e. Utiliser le modèle créé avec le jeu de validation pour prédire les résultats du jeu de test.

### 3. Extrants

- a. La classe d'événement prévu (Type de maintenance et nombre d'heures de vol).

Nous allons maintenant considérer comment cette approche est réalisée.

## **4.1 Intrants**

Les intrants sont en fait les données qui seront considérées par le système afin de résoudre le problème proposé. Chacune des sources de données a été identifiée et un échantillon suffisant de données représentatives a été extrait de chacune d'elles. Comme il s'agit d'une entreprise, chaque source de données est administrée et maintenue par un administrateur de système. Cet administrateur a dans son département des personnes clés, qui connaissent à fond le système. Il a fallu s'entretenir avec ces personnes, afin de comprendre les données et leur impact sur la prédiction d'événements (on considère une visite en usine comme un événement). Neuf sources de données ont été analysées pour savoir si elles contenaient de l'information pertinente. Une liste complète des sources est disponible en annexe A p.74.

## **4.1.1 Sélection des données pertinentes**

L'application des étapes suivantes nous a permis de répondre à plusieurs questions concernant les intrants du système et ainsi pouvoir démystifier les éléments ayant une importance dans la résolution du problème. Pour voir comment nous avons répondu aux différentes questions de chacune des étapes présentées, il existe une version détaillée du plan en annexe B p.90.

### **4.1.1.1 Voir ce qui se fait manuellement**

Tous les modèles de moteurs n'étant pas identiques, chaque famille de moteurs possède ses propres spécifications. Certains modèles ont plus d'âge que d'autres. Certains modèles viennent d'entrer en service. Il faut considérer tous ces facteurs lors de l'analyse de chaque modèle. Pour un modèle plus âgé, les intervalles de maintenances sont plus stables, alors que pour un modèle plus jeune, ces intervalles sont beaucoup moins stables. Ceci est considéré dans les prédictions manuelles qu'effectue l'entreprise.

De plus, chaque opérateur (qu'il soit un opérateur privé, ou encore un opérateur pour la flotte d'une compagnie aérienne) est responsable du suivi sur la maintenance. Certains respectent les exigences et recommandations de l'entreprise, alors que d'autres, considérés comme des « délinquants », ne suivent pas ces dernières recommandations. On dit qu'un opérateur utilise la maintenance qualifiée de « *Soft-Time* » lorsqu'il adhère au programme de maintenance proposé et « *hard-Time* » lorsqu'il est considéré délinquant. Ceci est également considéré dans les prédictions manuelles qu'effectue l'entreprise.

Nous avons modélisé ces deux dernières contraintes, préalablement effectuées manuellement par les experts, en ajoutant l'âge du moteur ainsi que son usage mensuel dans l'historique d'événements.

#### **4.1.1.2 Les maintenances de base et les bulletins de service recommandés**

Avec l'aide des experts du domaine, nous avons convenus que les informations sur les maintenances de base ainsi que les bulletins de services étaient intrinsèquement incluses dans l'historique d'événements, puisqu'ils sont considérées et enregistrées comme les événements fortuits.

#### **4.1.1.3 Inclure l'historique d'événements passés**

Nous avons inclus l'historique d'événements provenant de la source de données « *Shop Visit Historical* » dans les intrants.

**Nom:** Shop Visit Historical Data (TRAC, TRS)

**Format:** Base de données Oracle reliée à un système SAS 9

**Contenu:** Enregistrements contenant les événements passés (maintenances), ayant pour caractéristiques (champs) le type de maintenance, la description de la maintenance, le moteur impliqué et le nombre d'heures de vol (cycles de vol).

#### **4.1.1.4 Inclure les événements fortuits (ce qu'il peut arriver)**

Ce type d'événement est également inclus dans l'historique d'événements. Mais, advenant une piètre qualité des données ou un volume insuffisant, certains intrants peuvent être exclus ou ignorés. D'ailleurs, certaines caractéristiques (champs) de ces intrants peuvent être manquantes ou non-existantes. Il est important de savoir que la qualité des données utilisées dans le cadre de ce projet ne dépend pas seulement du « bon vouloir » de la compagnie, mais aussi de celui de tous les opérateurs et de tous les Centres d' « *Overhauls* » Désignés.

#### **4.1.1.5 Inclure les facteurs macro-économiques et environnementaux influençant le modèle de moteur**

Les facteurs macro-économiques et environnementaux ont été ajoutés à l'historique pour tous les enregistrements. Ainsi on peut savoir dans quel type d'environnement ce moteur était-il opéré et quels étaient les valeurs des principaux indices économiques lors de l'événement.

Une fois toutes ces questions répondues, nous avons retenu un total de 31 caractéristiques, pour chaque enregistrement de l'historique d'événements. La caractéristique qui suit a été retenue pour la classification de l'historique d'événements utilisé avec le SVM:

##### **Days between events**

Source: Champ calculé

Utilité: Ce champ calcule le délai entre les événements d'un même moteur en nombre de jours. En classant les événements en ordre croissant de dates, on peut identifier ce délai. Il suffit, pour un même numéro de série, de soustraire la date du dernier événement de celle de l'événement courant.

Importance : Ce champ est l'un des 2 champs cibles. Nous souhaitons prédire quand aura lieu le prochain événement pour un moteur précis.

Il existe une description détaillée des caractéristiques retenues en entrée en annexe C p.126. On peut représenter le processus de sélection comme suit:

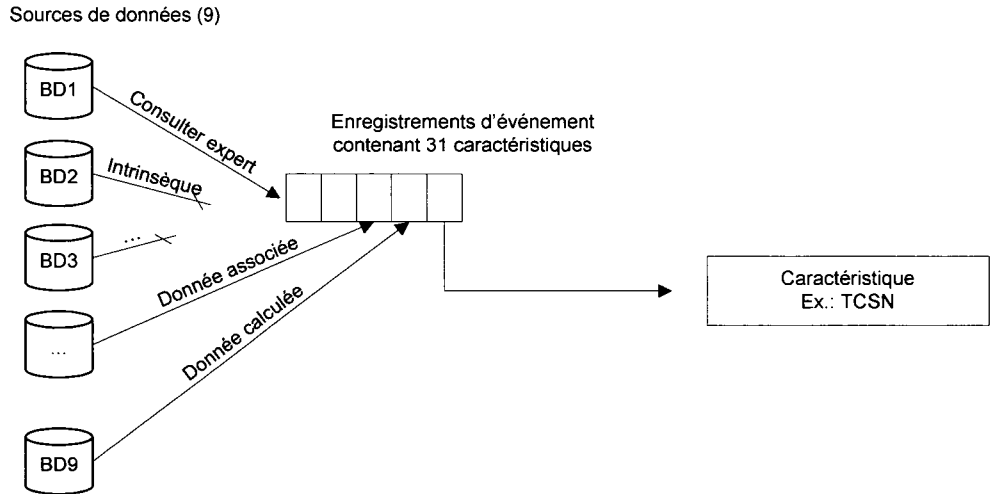


Figure 7 - Processus de sélection des données

#### 4.1.2 Conversion des valeurs catégorielles

Il nous a fallu par la suite effectuer une conversion des valeurs catégorielles, parmi les données considérées pertinentes, en valeurs numériques (le « *ranking* »). En voici un exemple sur le champ (la caractéristique) Event Classification 3

Nom	Event Classification
Ranking	3
Valeur originelle	BUR

Ceci est l'un des champs clés. Il s'agit d'une partie de la prédiction (Target) à fournir. Il indique le type d'événement qui s'est produit pour que le moteur ait besoin d'une visite en usine. La première lettre signifie s'il s'agit d'un événement « *B-Basic* » ou « *N-Non Basic* ». Les événements de base (basic) sont moins graves que les événements qui ne sont pas de base (« *Non Basic* »). La deuxième lettre signifie si l'événement est prévu (« *Planned* ») ou non-prévu (« *Unplanned* »). Habituellement, les événements non-prévus sont plus graves que les prévus. Finalement, la troisième lettre nous renseigne sur ce qui a été fait comme maintenance sur ce moteur en relation avec cet événement. Par exemple, un « *O* » signifie un « *overhaul* » (on doit défaire le moteur en pièces), alors qu'un « *R* »

signifie « *Removal* » (on a du enlever le moteur). Voici les scores (« *ranks* ») obtenus, du moins critique au plus critique :

BPR	4
BPO	5
NUR	6
BUR	8
BUO	9

Pour une description détaillée de la conversion des valeurs, nous vous convions à consulter l'annexe D, p.138.

### 4.1.3 Normalisation des données

Nous avons ensuite effectué une normalisation sur les données pertinentes sélectionnées (converties ou non), afin de les ramener sur l'intervalle [0,1]. Dans un classeur Excel où les données ont été converties en nombres, nous effectuons ensuite une normalisation « min-max ». La formule utilisée est la suivante :

$$\frac{1 \text{ Caractéristique de (Données historiques + moyennes d'utilisation mensuelle + Facteurs environnementaux + facteur Macro-économiques)} - \text{Valeur Minimum du champ}}{(\text{Valeur Maximale du champ} - \text{Valeur Minimale du champ})}$$

Ce qui en revient à faire pour chaque caractéristique de l'enregistrement:

$$= \frac{(\text{Valeur du champ} - \text{Valeur Min pour ce champ})}{(\text{Valeur Max pour ce champ} - \text{Valeur Min pour ce champ})} \quad (5)$$

En appliquant la formule de l'équation (5), nous obtenons des valeurs contenues entre 0 et 1. En fait, la valeur la plus faible devient = 0 et la valeur maximale devient = 1. Aucun ratio n'est perdu. Cependant, s'il y a des valeurs manquantes, il se peut que la normalisation ne soit pas efficace et donne des valeurs



négligentes. C'est ce qui s'est produit avec le numéro de série de l'aéronef une fois normalisé. Il y avait des valeurs de l'ordre de -27 dans les données normalisées. Aussi, cette caractéristique a dû être retirée des jeux d'essais, bien avant de créer les jeux d'entraînements.

Une fois toutes ces transformations réalisées, nous obtenons une vision plus claire des intrants et des caractéristiques utiles dans la résolution du problème. Nous pouvons maintenant sélectionner l'approche de résolution du problème de maintenance des moteurs d'aéronefs qui se veut la plus efficace pour l'obtention des résultats attendus.

## **4.2 Création du modèle de prédiction**

Parmi les différentes techniques proposées au chapitre précédent, nous avons retenu le SVM pour sa capacité à aborder des problèmes complexes et de grande taille, mais aussi pour sa facilité d'implantation. Nous avons tenté également, préalablement de résoudre le problème à l'aide des KNN, des arbres de décision, sans atteindre les résultats escomptés.

Notre approche, se veut une approche hybride, puisque nous avons utilisé conjointement un modèle d'arbre décisionnel jumelé au SVM. Nous obtenons ainsi la flexibilité nécessaire dans le traitement d'une grande quantité de données, tout en s'assurant que les résultats obtenus s'approchent indéniablement des résultats escomptés.

Globalement, la réalisation du modèle de prédiction se déroule comme suit: on sépare l'historique de données normalisées en trois groupes. Chaque groupe est exporté dans un fichier indépendant comprenant l'en-tête nécessaire à l'exécution du SVM et la valeur de la classe cible, sauf pour le jeu de test, qui lui ne contient pas la valeur de la classe cible. Chaque fichier est soumis au SVM. Avec le jeu

d'entraînement, on vérifie la fonction « *Kernel* » à utiliser et on applique cette fonction sur le jeu de validation. Si l'algorithme réussit avec un taux d'erreurs satisfaisant (inférieur à cinq pourcents), alors on l'exécute sur le jeu de test. La classe ayant le taux d'erreurs de classification le plus faible se verra ajoutée comme première feuille de notre arbre décisionnel. Un nœud feuille est un nœud sur lequel on ne peut appliquer d'autres scissions. On ré-exécute ensuite la séparation des classes restantes à l'aide du SVM et à chaque fois on ajoute le nœud dans l'arbre décisionnel, jusqu'à l'obtention de toutes les feuilles.

#### **4.2.1 Création des jeux d'observation**

L'ensemble des données normalisées est divisé en deux parties: 80% pour former un jeu d'entraînement et le 20% restant pour former un jeu de validation. Il faut compter aussi sur un jeu de test ayant une taille similaire au jeu de validation, mais qui contient des enregistrements différents des deux précédents. Ces jeux doivent être formatés pour le SVM et sauvegardés en format texte.

Le jeu d'entraînement est conçu tel que nous le montre la Figure 8 ci-dessous:

```

Train21 - Notepad
File Edit Format View Help
@parameters

#training set
@examples
format xy
0.245706274 0.678486998 0.460146937 0.403735632 0 1
0.421661409 0.718676123 0.606251733 0.525359195 0 0
0.127584998 0.678486998 0.307457721 0.269396552 0 1
0.046267087 0.14893617 0.166343222 0.193103448 0 1 0
0.049071153 0.14893617 0.215137233 0.244396552 0 1 0
0.123729408 0.14893617 0.432284447 0.429382184 0.96
0.146862951 0.14893617 0.482741891 0.51408046 0.8 1
0.119873817 0.14893617 0.328597172 0.321551724 0 1 0
0.156677182 0.14893617 0.415927363 0.431178161 0.8 1
0.029092184 0.14893617 0.415927363 0.431178161 0.8 1
0.406940063 0.678486998 0.569725534 0.565014368 0.4
0.121976866 1 0.223662323 0.241954023 0 0 0.8 1 0 0.
0.127584998 0 0.267881896 0.414511494 0 0.666666667
0.208201893 1 0.335597449 0.347557471 0 1 0.8 0.9347
0.184016824 1 0.57256723 0.553591954 1 0 0.8 0.93472
0.037153873 0.378250591 0.092874965 0.165804598 0 0

```

Caractéristique

```

lisp
21737 2
26 2
4504 2
34 2
27 2
9693028 2
8371 2
625 2

543554 2

```

Classe cible

Figure 8 - Le jeu d'entraînement

Le jeu de validation ne diffère pas beaucoup du jeu d'entraînement. Son en-tête est différente, mais les données sont similaires, il s'agit de d'autres cas non rencontrés encore dans le jeu d'entraînement. La figure 9 nous expose un exemple de jeu de validation:

```

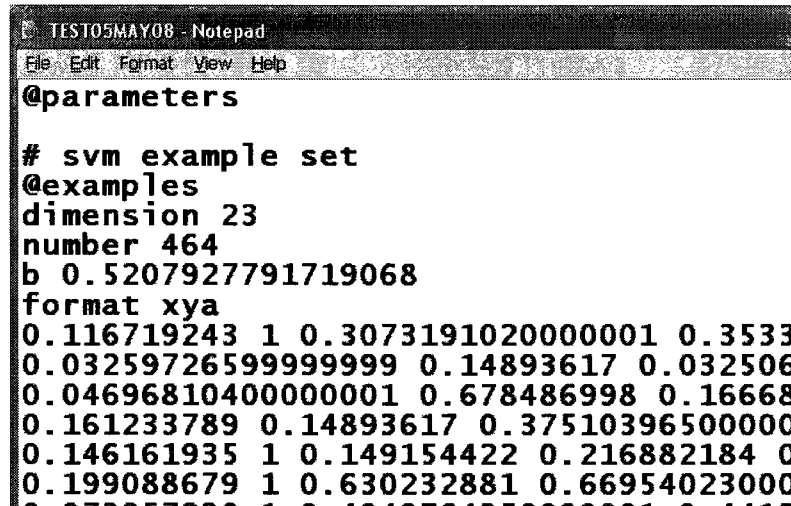
Validate23 - Notepad
File Edit Format View Help
@parameters

# svm example set
@examples
dimension 23
number 381
b -0.8722590209257255
format xya
0.132141605 1 0.170640421 0.162643678 0.48
0.09533824000000003 0.14893617 0.109301358
0.126533474 1 0.3788466870000001 0.36034482
0.031195233 1 0.3249930690000001 0.36400862
0.412197687 0.14893617 0.160451899 0.149137
0.116719243 1 0.3073191020000001 0.35330459
0.032597266 0.14893617 0.032506237999999999
0.199088679 1 0.630232881 0.6695402300000000
0.073957238 1 0.4049764350000001 0.44130747
0.139852787 1 0.3121014690000001 0.33879310

```

Figure 9 - Le jeu de validation

Dans le jeu de test, l'en-tête ne diffère pas du jeu de validation, mais les enregistrements n'ont pas de classe cible attribuée. La Figure 10 nous présente un exemple de jeu de test:



```
TEST05MAY08 - Notepad
File Edit Format View Help
@parameters
# svm example set
@examples
dimension 23
number 464
b 0.5207927791719068
format xya
0.116719243 1 0.3073191020000001 0.3533
0.03259726599999999 0.14893617 0.032506
0.04696810400000001 0.678486998 0.16668
0.161233789 0.14893617 0.37510396500000
0.146161935 1 0.149154422 0.216882184 0
0.199088679 1 0.630232881 0.66954023000
```

Figure 10 - Le jeu de test

#### 4.2.2 Appliquer le SVM sur les jeux d'observation

Nous avons donc créé des jeux d'observation, afin de tenter de séparer les enregistrements en des groupes uniformes, à l'aide du SVM. Un jeu d'observation est constitué de 3 sous-ensembles: un jeu d'apprentissage, un jeu de validation et un jeu de test (Prédiction de nouvelles valeurs). Dans les jeux d'apprentissage et de validation, la valeur cible est connue, alors que dans le jeu de test, l'algorithme doit lui-même déterminer la classe d'appartenance.

Nous avons utilisé le SVM spécialement pour représenter nos données dans un espace de deux dimensions et pour démontrer l'utilité des fonctions de « *Kernel* » dans le choix de la meilleure fonction. Pour évaluer la performance du SVM, une mesure de l'exactitude a été effectuée: la classe associée aux enregistrements du jeu de test doit être prédite avec une marge d'erreur de moins de cinq pourcents.

Les étapes de notre méthode de prédiction sont les suivantes :

- 1- Définir l'ensemble d'apprentissage ;
- 2- Définir l'ensemble de validation : fichier validation.txt;
- 3- Affecter une classe «  $y$  » à chaque enregistrement dans les jeux d'entraînement et de validation:  $y = 1$  ou bien  $y = -1$  lorsque l'on a deux classes ou bien  $y = \{1, 2, 3 \dots n\}$  lorsque l'on a  $n$  classes;
- 4- Appliquer le SVM sur l'ensemble de données d'apprentissage (les jeux d'entraînement et de validation);
- 5- Choisir la bonne fonction « *Kernel* » (le bon séparateur parmi les fonctions proposées par le SVM);
- 6- Appliquer le SVM pour calculer l'exactitude, sur le jeu de test, afin de voir s'il a pu prédire avec un minimum de cinq pourcents d'erreurs.
- 7- Répéter sur les classes non pures pour scinder en sous groupes plus uniformes.

Dans chacun des tests effectués, nous avons fait varier le nombre de caractéristiques (dimensions) pour déterminer quel modèle serait le plus efficace pour bien classifier les événements sur les moteurs. Ceci nous a permis d'isoler les caractéristiques les plus pertinentes, réduisant le plus possible les erreurs de classification.

Nous allons regarder le cœur de la méthode avec notre troisième jeu d'essai incluant certaines des caractéristiques choisies. Comme les résultats précédents étaient

satisfaisants, pour séparer les événements de type Prévus, comparativement aux types imprévus, nous avons tenté d'effectuer une classification/prédiction sur un plus grand ensemble de données. Nous avons utilisé les mêmes caractéristiques et la même configuration SVM que dans le deuxième jeu d'essai:

Chaque enregistrement contient les caractéristiques suivantes:

**Table 3 - Liste des caractéristiques sélectionnées**

CARACTÉRISTIQUE	DESCRIPTION
Days between events	Le nombre de jours écoulés depuis le dernier événement
Political Belonging Normalized	Le pays d'appartenance normalisé
Engine Time Since New	Le nombre d'heures de vol depuis la mise en service du moteur
Engine Cycles Since New	Le nombre de cycles depuis la mise en service du moteur. 1 cycles = 1 décollage + 1 atterrissage.
Time Between Overhaul	Le nombre d'heures de vol entre les O/H
Opr clas normalized	La classe de l'opérateur normalisée
Event C 3 Rating	Le « <i>ranking</i> » de l'événement (le type de maintenance)
Engine age in days	L'âge du moteur en nombre jours
Sand	Le taux de sable moyen des régions où ce moteur est utilisé
Salt	Le taux de sel moyen des régions où ce moteur est utilisé
Volcano	Le taux de volcans moyen des régions où ce moteur est utilisé
Pollution	Le taux de pollution moyen des régions où ce moteur est utilisé
Elevation	L'élévation moyenne en mètres des

CARACTÉRISTIQUE	DESCRIPTION
	régions où ce moteur est utilisé.
Crude Oil Price USD	Le prix du baril de pétrole en dollars US au moment de l'événement
S&P500	La valeur du S&P500 au moment de l'événement
Corp Profits US	La valeur des profits corporatifs aux États-Unis au moment de l'événement
Interest Rates US	La valeur des taux d'intérêts aux États-Unis au moment de l'événement
W GDP	La valeur de l'indice de croissance mondiale au moment de l'événement
USDCAD=X	La valeur tu taux de change américain en dollars canadiens au moment de l'événement
Nb of Engines	Le nombre de moteurs en service par cet opérateur
MRT	L'usage mensuel du moteur en nombre d'heures de vol
TRT	L'usage total du moteur depuis sa mise en service en nombre d'heures
AMU	L'utilisation moyenne mensuelle du moteur en nombre d'heures
Target	La classe cible: 1=Planned 2=Basic Unplanned 3=Non-Basic Unplanned

Considérons ci-dessous les paramètres du SVM, en utilisant la fonction « Kernel » DOT standard, sur 100 itérations (« runs »):

**Jeu:** # svm example set  
**Validation/Test:** @examples  
**Nb de caractéristiques:** dimension 23  
**Nb d'enregistrements::** number 771

**Valeur de bêta:** b -0.1664645680990583

Afin de savoir quel calibrage effectuer, il faut exécuter le SVM sur le jeu d'entraînement. La fonction contenant le moins d'erreurs de classification sera en tête de liste lorsque l'on trie le fichier en ordre décroissant (le fichier créé par le SVM contient la mention « -opt » ajouté au nom original). La Figure 11 nous montre un fichier « -opt » trié en ordre décroissant:

	A	B	C	D	E	F	G	H
1	MSE	C	epsilon	nu	kernel			
2	6.13E+20	1.00E+06			0.4	type dot		
3	1.28E+16	100000			1	type polynom degree 3		
4	1.28E+16	100000			1	type polynom degree 3		
5	1.46E+13	1.00E+06	0.001			type neural	a 0.9	b 0.8
6	1.38E+13	1.00E+06	1.00E-05			type neural	a 1	b 0.6
7	1.28E+13	0.01			1	type polynom degree 6		
8	8.92E+12	1			0.4	type polynom degree 6		
9	7.50E+11	100000			1	type anova	gamma 0.3	degree 2
10	7.47E+11	10			0.6	type polynom degree 5		
11	8.21E+10	1.00E+06			1	type anova	gamma 0.3	degree 1
12	8.03E+08	10000	0.1			type neural	a 0.3	b 1

**Figure 11 - Le choix de la fonction de « Kernel »**

Il est important de noter ces valeurs avant d'exécuter le SVM sur le jeu de validation et de test. Il faut considérer le taux d'erreurs minimum (le MSE le plus petit), et s'assurer de noter les valeurs pour le C, epsilon, nu, et bien sur la fonction « Kernel » choisie ainsi que les paramètres a et b lorsqu'ils sont présents.

Cette fois-ci, les résultats étaient tout aussi impressionnants, sur 129 enregistrements à prédire, seuls ceux de la classe 3 ont obtenu des erreurs, mais si on observe bien, le SVM leur a quand même assigné à tous la même classe (ici, la valeur 1.5, au lieu de 3). Les autres jeux d'essai vous sont présentés en annexe E, p.144. À première vue, lorsque nous conservons toutes les caractéristiques, nous obtenons un taux d'erreur particulièrement élevé. Cependant, plus nous retirons de



caractéristiques qui ne semblent pas fournir d'information pertinente, le taux d'erreur semble baisser dramatiquement.

### 4.2.3 Création de l'arbre de décision hiérarchique

Nous aurions, graphiquement, la situation suivante, où l'on peut bien voir l'arborescence dans l'utilisation des SVM pour résoudre le problème multi-classes:

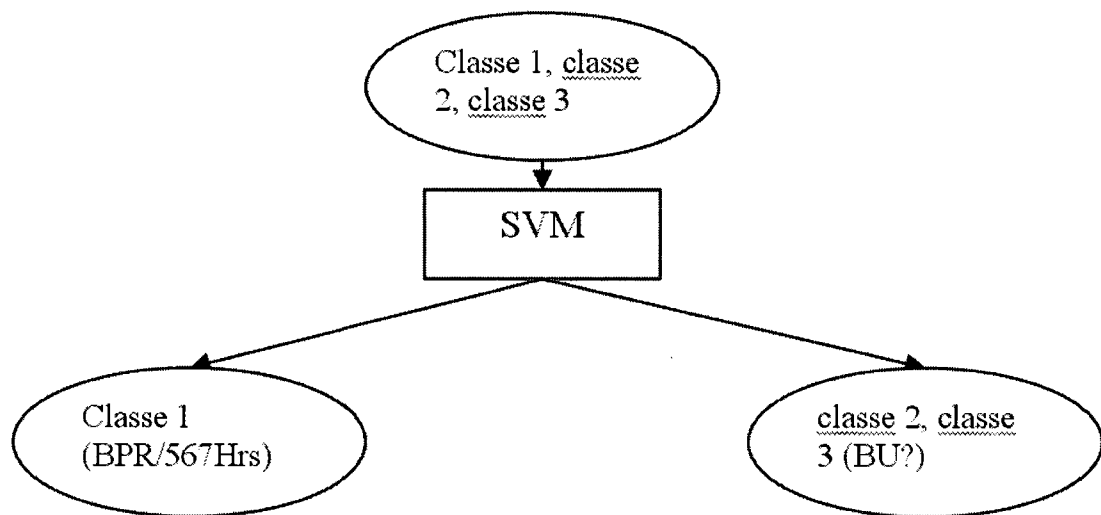
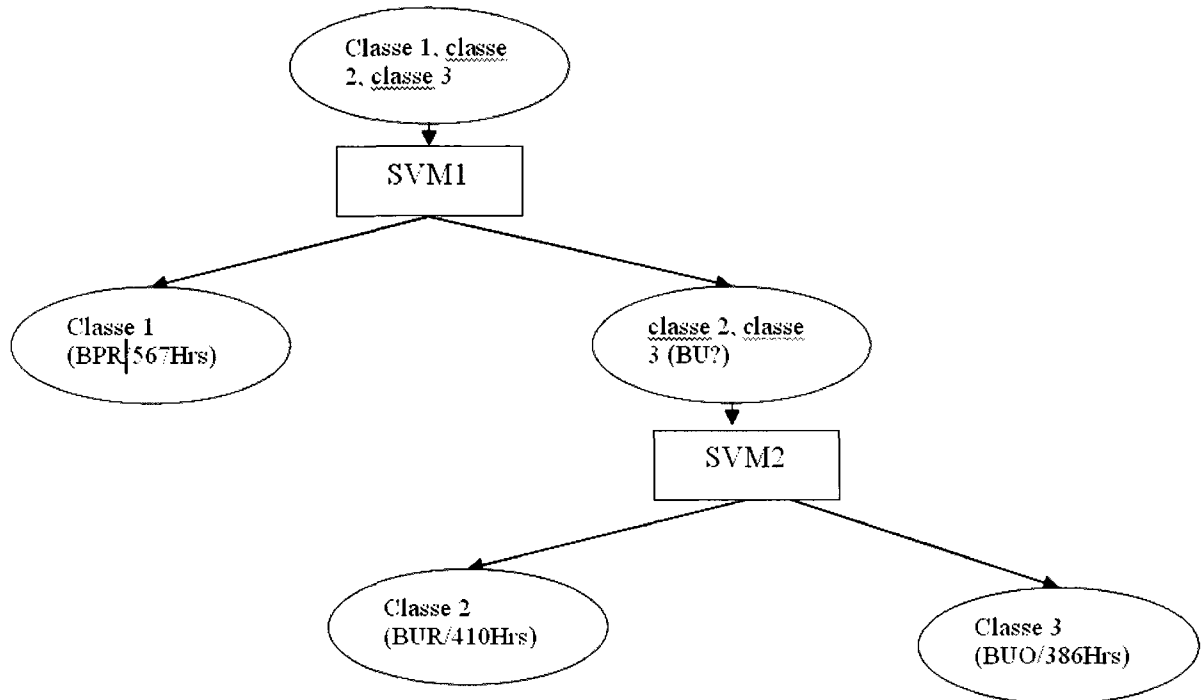


Figure 12 - Arbre de décision hiérarchique

### 4.2.4 Répéter les étapes précédentes et raffiner le modèle

Afin de compléter la classification, nous devrions réutiliser le SVM sur l'ensemble des 2 classes isolées dans le nœud de droite pour obtenir la dernière scission. Nous ajoutons ensuite un nouveau nœud à l'arbre de décision préalablement construit:



**Figure 13 - Arbre de décision hiérarchique**

Si nous regardons le premier exemple dans les jeux d'essais, nous avons utilisé les 31 caractéristiques et les résultats n'étaient pas concluants. Tous les exemples dans lesquels nous avons laissé la totalité des caractéristiques n'ont pas donné de bons résultats. Il est donc clair que le choix des caractéristiques joue un rôle plus qu'important dans l'atteinte de résultats pertinents. Si l'on regarde le quatrième jeu, encore une fois, avec presque toutes les caractéristiques, il a obtenu beaucoup d'erreurs de prédiction avec la même variable cible que les exemples 2, 3 ou 8, qui eux, utilisent un jeu réduit de caractéristiques et pour lesquels les erreurs de prédiction sont d'un ordre beaucoup plus faible.

Dans le dixième exemple, nous avons utilisé 26 caractéristiques. Le taux d'erreur était beaucoup plus faible que nos jeux d'essais ayant plus de caractéristiques. La même chose semble se produire dans le neuvième jeu d'essai, où nous avons obtenus un taux d'erreur relativement bas, mais non satisfaisant avec 25 caractéristiques. Dans les exemples 2, 3 et 8, nous utilisons un jeu réduit de

caractéristiques. En utilisant 23 caractéristiques, nous obtenons de bons résultats pour séparer les événements prévus de ceux qui sont imprévus. Dans les jeux 5, 6, 7, 9 et 10, nous tentons de séparer l'ensemble sur le nombre de jours entre les événements. Bien que nous soupçonnions ce cas d'être non linéairement séparable, nous avons tenté l'expérience quand même. En retirant des caractéristiques, d'un jeu à l'autre, l'erreur de prédiction diminue.

Ceci nous a permis de sélectionner les caractéristiques les plus influentes pour l'atteinte des résultats attendus. Ainsi la classification se fait sans erreur et le modèle est robuste pour l'association de classes pour de nouveaux enregistrements (surtout si l'on utilise les caractéristiques des jeux d'essais 2, 3 et 8).

#### **4.2.5 Utiliser le modèle créé avec le jeu de validation pour prédire les résultats du jeu de test**

Nous utilisons le modèle de prédiction (l'arbre de décision hiérarchique) créé à l'aide du jeu de validation pour prédire la valeur de la classe associée à chaque enregistrement du jeu de test. En regardant l'arborescence, nous savons sur quelles classes effectuer la première scission: il faut séparer la classe 1 (les BPRs ayant en moyenne 567 heures de vol depuis leur dernier événement). Si le taux d'erreurs est satisfaisant (donc, sous le minimum de cinq pourcents en comparant la valeur de la prédiction donnée avec celle attendue), on ré-exécute le SVM sur les deux classes restantes pour séparer les BURs ayant en moyenne 410 de vol depuis leur dernier événement des BUOs ayant en moyenne 386 de vol depuis leur dernier événement. Si l'algorithme obtient un taux d'erreurs inférieur à cinq pourcents pour séparer toutes les classes, alors il peut être utilisé comme modèle de prédiction pour les futures visites en usines. Il est possible de passer à la phase d'implantation. Évidemment, si les données concernant les moteurs changent dans les sources de données (un nouveau moteur s'ajoute, par exemple), alors il faut repartir des sources et créer un nouveau modèle, sinon les prévisions de l'ancien modèle s'effectueront avec les anciennes données sur lesquelles le modèle a été créé.

### **4.3 Extrants**

La réponse attendue du système se veut assez simple. L'entreprise a besoin de savoir, pour un moteur précis, dans combien de temps (en termes de nombre d'heures de vol) aura lieu la prochaine visite en usine et quelle sera la maintenance à effectuer (selon les types prédéfinis: BUR, NUR, etc). Le résultat peut par contre s'exprimer de plusieurs façons, en utilisant une technique hybride basée sur le SVM, alors nous obtiendrons une classe en sortie, regroupant tous les moteurs susceptibles de réagir de la même manière (donc ayant sensiblement les mêmes maintenances aux mêmes nombres d'heures).

Chaque feuille (classe) de l'arbre décisionnel hiérarchique contient deux informations: le type de maintenance et quand elle aura lieu. Il s'agit de notre sortie finale. Le type est connu directement, puisqu'il est unique à tous les éléments de la classe. Le nombre d'heures de vol moyen est ensuite calculable en utilisant les caractéristiques « *Days between events* » des membres de cette classe. Nous obtenons ainsi, par exemple, la prévision qu'un BPR risque de se produire (lorsqu'un enregistrement est classé dans la classe 1 de la figure 12 par exemple) et comme la moyenne du nombre de jours entre les événements de ce type est de 567 heures, cet événement risque de se produire dans 567 heures.

## Chapitre 5 – Étude expérimentale

### 5.1 Un exemple concret d'application

Suivons un exemple depuis le début, sur un modèle spécifique. Nous devons suivre les étapes définies dans les intrants et s'assurer de choisir les bonnes sources d'information. La définition de chacune des étapes avec les données réelles est présentée en annexe F, p.161.

Une fois toutes les données connues, il faut effectuer la conversion des valeurs catégorielles en valeurs numériques. Nous obtenons un tableau similaire à celui-ci:

Numéro	Normalized SN	Days between events	Political Bel	Engine Time Since New	Engine TSN Recorded Date	Engine Cycles Since New	Time Between O	Time Since Overhaul
1	211	346	0.4380	4434	25/04/04	4918	12500	0
2	211	106	0.0780	469	15/10/01	441		
3	211	147	0.3020	2405	15/05/03	3024	6000	0
4	208	473	0.0780	5412	06/11/04	5152	6000	2184
5	211	430	0.4380	2152	19/12/02	3019	6000	0
6	210	581	0.4380	9093	16/06/06	9320	6000	0
7	210	224	0.4380	5843	12/11/04	6143	6000	0

Figure 14 - Données converties

Ensuite, il faut normaliser les données, à l'aide de la normalisation min-max, afin d'obtenir un tableau dont les valeurs se situent dans l'intervalle [0,1], comme dans l'exemple suivant:

Numéro	Normalized SN	Days between	Political Belc	Engine Time	Engine TSN	Engine Cycle	Time Between	Time Since	Cycles Since	Normalized	A Opr	clas non
0	0.823529412	0.11671924	1	0.3073191	0.60606061	0.3533046	1	0	0	0.67256637	1	
0.0012987	0.823529412	0.03259727	0.14893617	0.03250624	0.28822314	0.03168103	0	0	0	0.59292035	1	
0.00649351	0.819607843	0.19908868	1	0.63023288	0.87534435	0.66954023	0.48	0	0	0.15044248	0	
0.00779221	0.819607843	0.07395724	1	0.40497643	0.67527548	0.44130747	0.48	0	0	0.05309735	1	
0.00909091	0.819607843	0.13985279	1	0.31210147	0.5981405	0.3387931	0.48	0	0	0.6460177	1	
0.01038981	0.823529412	0.06168945	1	0.36027169	0.6707989	0.4200431	1	0	0	0.46017699	1	
0.01168831	0.819607843	0.06694707	0.14893617	0.06986415	0.3219697	0.08666667	0	0	0	0.34513274	1	
0.01428571	0.811764706	0.15878023	0.14893617	0.22373163	0.51033058	0.23505747	0.48	0	0	0.76106195	1	
0.01558442	0.811764706	0.09533824	0.14893617	0.10930136	0.34986226	0.11436782	0	0	0	0.33628319	1	
0.01818182	0.819607843	0.13214161	1	0.17064042	0.45626722	0.16264368	0.48	0	0	0.62831858	1	

Figure 15 - Données normalisées

On doit préparer ensuite 3 jeux d'essais: un jeu d'entraînement, un jeu de validation et un jeu de test. En commençant par le jeu d'entraînement, nous pouvons déterminer quelle sera la fonction à utiliser. Habituellement, nous choisissons celle qui indique le plus bas taux d'erreur de classification.

Regardons les résultats du SVM sur notre ensemble d'entraînement:

MSE	C	epsilon	nu	kernel	gamma	degree
2.19E-05	10	0.001		type anova	gamma 0.6	degree 3
6.66E-09	10000	1.00E-06		type radial	gamma 100	
1.53E-08	100000	1.00E-05		type radial	gamma 0.5	
2.01E-08	100	1.00E-06		type radial	gamma 1000	
2.34E-08	100	1.00E-05		type anova	gamma 0.8	degree 4
1.03E-06	1.00E+06	0.001		type radial	gamma 0.8	
0.00010006	100	0.01		type radial	gamma 100	
0.00012207	10000	0.01		type radial	gamma 0.05	
0.0003157	10000	1.00E-06		type anova	gamma 0.9	degree 2
0.00110023	1000	0.0001		type polynom	degree 3	
0.00124129	1	1.00E-06		type polynom	degree 3	
0.00201014	10	0.0001		type radial	gamma 0.9	
0.00563079	0.01	0.001		type anova	gamma 0.8	degree 3
0.0335238	10000	1.00E-06		type anova	gamma 0.1	degree 3
0.0475415	10	0.0001		type radial	gamma 0.05	
0.0584346	1	1.00E-05		type radial	gamma 0.3	
0.0730377	0.1	0.01		type polynom	degree 2	
0.0805324	0.1	0.1		type radial	gamma 10	
0.0854286	10		0.2	type radial	gamma 0.3	
0.0857995	1		1	type radial	gamma 0.1	
0.0858337	0.1		1	type radial	gamma 5	

Figure 16 - Résultats du SVM pour le choix d'une fonction « Kernel »

Ici, la fonction surlignée est celle ayant le taux d'erreur (MSE) le plus bas, soit de  $6.66 \times 10^{-9}$ .

Nous prenons alors les données du jeu de validation en utilisant cette fonction (ici, le type défini est radial, avec un gamma de 100, epsilon de 10000 et nu de  $1 \times 10^{-6}$ ).

Nous obtenons cette fois-ci, des résultats s'approchant des résultats attendus:

Target	# exemples y:
1	1.43638
1	1.23852
1	1.47603
1	1.20871
1	1.40547
1	1.29924
1	1.45679
1	1.28952
1	1.16778
1	1.2245
1	1.44361
1	1.38229
1	1.45683
1	1.24336
1	1.36421
1	1.35827
2	2.13903
2	2.27066
2	2.18892
2	2.19415
2	2.05686
2	1.77689
2	2.22334
2	2.2026

**Figure 17 - Résultats du Jeu de Validation**

Il ne nous reste ensuite qu'à répéter l'opération avec un jeu de test, qui soit composé de valeurs non utilisées pour l'entraînement et la validation, pour lesquelles la valeur cible n'est pas connue. Nous obtenons des résultats similaires à ceux-ci:

# exemples ys  
 1.11094 → Classe 1  
 0.988801 → Classe 1  
 1.04764 → Classe 1  
 1.10384 → Classe 1  
 1.97677 → Classe 2  
 1.0524 → Classe 1

Lorsque nous utilisons 31 dimensions sur les données (31 des champs les plus influents), si nous abordons la résolution dans ces paramètres:

**Table 4 - Paramètres d'exécution exemple 1**

MSE	C	epsilon	nu	kernel	
4.06E-08	10000	1.00E-06	1E-06	type radial	gamma 100
Jeu: " Validate AllNorm "					

Nous obtenons les résultats suivants:



**Table 5 - Résultats de l'exemple 1**

Target	Results	delta
1	-306.243	307.243
1	-182.517	183.517
1	-96.0364	97.0364
1	-312.546	313.546
1	-171.809	172.809
1	-254.207	255.207
1	-79.5682	80.5682
1	-83.879	84.879
1	-39.0051	40.0051
1	-123.593	124.593
1	44.5345	-43.5345
1	-99.4542	100.4542
2	-37.5184	39.5184
2	-164.62	166.62
2	577.53	-575.53
2	25.9151	-23.9151
2	38.8458	-36.8458
2	134.143	-132.143
2	235.345	-233.345
3	120.754	-117.754
3	-5.23915	8.23915
3	-79.7061	82.7061
3	39.2816	-36.2816
3	24.1792	-21.1792

Par contre, avec les paramètres suivant, la solution s'améliore de beaucoup:

**Table 6 - Paramètres d'exécution exemple 2**

MSE	C	epsilon	nu	kernel	
4.35E-08	1000	0.0001	n/a	type radial	gamma 10
Jeu: "Validate Complete – Normalized"					

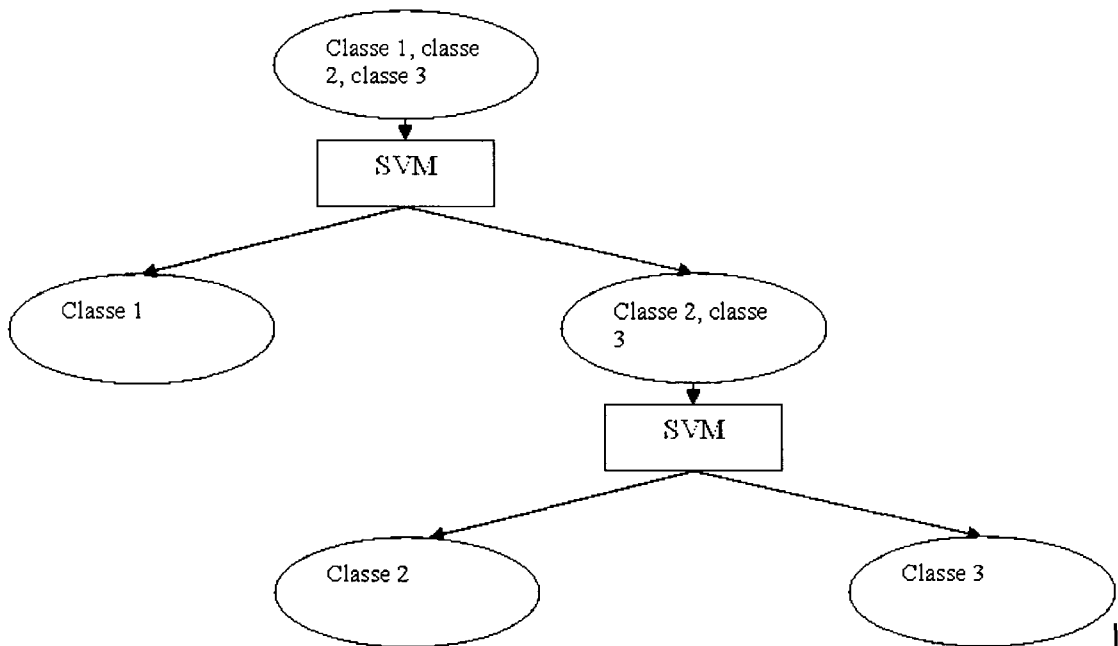
Puisque l'on peut ensuite remarquer une séparation beaucoup plus évidente et claire:

**Table 7- Résultats de l'exemple 2**

Target	Results	Delta
1	1.00004	0%
1	1.00014	0%
1	0.999718	0%

Target	Results	Delta
1	1.00038	0%
1	0.999903	0%
1	0.999891	0%
1	0.999757	0%
1	0.999984	0%
1	0.99998	0%
1	0.999979	0%
1	1.00003	0%
1	0.99959	0%
1	1.00001	0%
1	0.999685	0%
1	0.999934	0%
1	1.00018	0%
2	2.00041	0%
2	2.00014	0%
2	1.99988	0%
2	2.00045	0%
2	2.00021	0%
2	1.99996	0%
2	2.00007	0%
2	2.00007	0%
2	1.9999	0%
2	1.99996	0%
3	1.50005	150%
3	1.49974	150%
3	1.50001	150%
3	1.49985	150%
3	1.5	150%
3	1.4997	150%
3	1.49998	150%
3	1.49996	150%
3	1.50005	150%
3	1.50066	150%

Puisque le SVM effectue des scissions binaires, une approche « diviser pour régner » peut s'avérer très efficace pour scinder des groupes de plus de 2 classes, comme nous pouvons le constater dans le schéma suivant, présenté à la figure 18.



**Figure 18 - Arbre hiérarchique**

Les mêmes résultats peuvent être constatés avec seulement 23 dimensions, si nous utilisons les paramètres suivants:

**Table 8 - Paramètres d'exécution exemple 3**

MSE	C	epsilon	nu	kernel	
4.35E-08	1000	0.0001	n/a	type radial	gamma 2
Jeu: "Validate Complete – Normalized"					

Et aussi avec 24 dimensions, sur le jeu suivant "Validate 24dim – Normalized", mais avec les paramètres standards. Les deux nous offrent des résultats similaires:

**Table 9 - Résultats de l'exemple 3**

Target	Résultats	Delta
1	1.00004	0%
1	1.00014	0%
1	0.999718	0%
1	1.00038	0%
1	0.999903	0%
1	0.999891	0%
2	2.00053	0%
2	2.00003	0%
2	2.00015	0%

Target	Results	Delta
2	2.00002	0%
2	2.00015	0%
2	1.99998	0%
2	1.99994	0%
2	2.00054	0%
3	1.49998	150%
3	1.49996	150%
3	1.50005	150%
3	1.50066	150%
3	1.49944	150%
3	1.49994	150%
3	1.49978	150%

Si l'on augmente drastiquement le nombre de classes, alors là l'algorithme n'y parvient plus, comme dans les résultats suivants:

**Table 10 - Résultats en augmentant le nombre de classes**

Target	# exemples ys 1	# exemples ys 2
4	1.24E+07	-6.34E+06
8	4.93E+07	7.28E+06
8	1.29E+08	6.76E+06
2	5.91E+07	-8.81E+06
8	-8.42E+07	7.71E+06
2	1.98E+08	-2.76E+06
4	2.92E+08	-3.80E+06
4	-1.03E+08	-6.25E+06

## **5.2 Comparaisons**

Nous pouvons regarder les résultats obtenus avec d'autres algorithmes utilisés pour résoudre le problème proposé, ou encore, d'autres résultats obtenus avec la méthode du SVM.

Plusieurs essais ont été effectués à l'aide de la technique du « *K-mean* » (les K plus proches voisins). En utilisant le même modèle de moteur que celui utilisé pour le dernier exemple avec le SVM, voici ce que nous avons obtenu:

Les tests ont été effectués sur un intervalle de 30 classes, afin de déterminer combien de classes l'algorithme détecte dans les données. Les résultats ne sont pas satisfaisants lorsqu'il s'agit d'un nombre de classes inférieur à 11 classes. Aucune des classes n'est séparée efficacement, avec un taux supérieur à 95% de pureté. Cependant, lorsque nous avons 11 classes, il parvient à détecter une des classes avec un niveau de pureté de 100% (les événements de type NUR). Le même scénario se répète ainsi pour un nombre de classes inférieur à 15. Par la suite, il parvient à détecter une nouvelle classes avec un taux de 100% (les événements de type BPR). Encore une fois, le scénario se répète si l'on augmente le nombre de classes, mais lorsque nous atteignons 19 classes, il détecte 2 classes similaires avec un taux de 100% (les événements NUR, se séparent en 2 classes, l'algorithme détecte toujours les BPR aussi). Ce même schéma se répète ensuite, puis avec 22 classes, il détecte 3 classes similaires pour les BPR maintenant. Le tout se complique plus on ajoute de nouvelles classes, si bien qu'avec 30 classes il détecte les groupes suivants (avec un taux de pureté supérieur ou égal à 95%):

7 classes de BPR, 3 classes de BUR, 2 classes de NUR, mais aucune ne contenant que des BUO ou des BPO.

Dans aucun des cas, le temps n'est respecté dans l'attribution des classes.

**Table 11 - Moyenne et Écart-type des CSN**

* Cycles Since Overhaul
* Moyenne = 731,402
* Ecart-type = 791,595

Avec un écart-type supérieur à la moyenne, les données sont très dispersées. Ceci ne constitue pas un bon exemple de classement pouvant nous aider à solutionner le problème initial.

Des tests ont été ensuite effectués avec tous les types de modèles, afin de vérifier si la quantité de données historiques pouvait influencer la séparation des

classes. Toujours en utilisant le « *K-mean* », et en utilisant les données sur les visites en usines passées, ainsi que les données environnementales, pour un nombre de classes égal à 10, il ne détecte que les BPR.

Sur 5 classes, en substituant les données environnementales par les données macro-économiques, il parvient à détecter les BUR, mais le taux de pureté des BPR baisse de 3.15%. Cependant, si nous élevons le nombre de classes à 10, avec les mêmes données, cette fois-ci nous obtenons 3 classes de BUR, 3 classes de BPR et 2 classes de NUR.

Même en ne prenant que les événements prévus, sur 5 classes, le « *K-mean* », détecte 2 classes de BPO et 3 classes de BPR. Sans toutefois effectuer une bonne séparation sur le temps. Il en est de même avec 10 classes. En utilisant seulement les événements non prévus, sur 5 classes, le « *K-mean* » détecte 2 classes de BUR, 2 classes de NUR et une classe de BUO. Il en est sensiblement de même avec 10 classes.

D'autres tests ont été effectués à l'aide de l'algorithme de « *clusterization* » du logiciel SPSS. En considérant l'historique au complet, l'algorithme a de la difficulté à séparer les « *clusters* » jusqu'au nombre de 8. A partir de ce nombre, il associe ensemble certains événements au lieu de percevoir toutes les classes comme distinctes. Il joint ensemble les BPO et les BPR, puis dans une autre classe il joint les BUO et les BUR ainsi que les NUO et les NUR. Si on augmente ensuite le nombre de classes, les résultats sont plutôt similaires. L'algorithme ne parvient pas à effectuer une séparation qui soit efficace pour ce que nous tentons de résoudre.

Afin de vérifier de nouveau l'algorithme du SPSS, nous avons utilisé un ensemble réduit d'événements. Sur un ensemble de 2 à 10 classes, il ne parvient pas à effectuer une séparation qui soit efficace. Des tests ont également été faits pour tenter de séparer sur le temps entre les événements (en nombre d'heures de vol qui soit Zéro, « *Low* », « *Std* » et « *High* »). Sur un ensemble de 2 à 10 classes, encore une fois, rien n'est concluant en termes de séparation efficace. Un second test relatif

au temps à été effectué, en utilisant plus d'enregistrements, mais encore une fois, sans succès. Les résultats sont similaires à l'exemple précédent.

Cinq autres tests ont été effectués, en considérant les événements et le temps. Aucun d'eux n'a révélé de résultats satisfaisants. On peut voir le tableau des résultats en annexe E.11, p157.

En utilisant le SPSS pour les arbres de décision, alors nous observons une séparation possible pour ce qui touche le nombre de jours entre les événements (le temps). En considérant seulement 2 classes, nous percevons déjà une séparation apparente:

Lorsque la distance est faible entre la moyenne et l'écart-type, cela signifie que les données sont regroupées. Si, de plus, la moyenne est égale à l'écart-type, cela signifie que les données sont très dispersées. Si jamais l'écart-type est plus grand que la moyenne, alors là il y a beaucoup de données ayant une valeur faible pour une ou quelques autres données qui ont une valeur élevée.

Nous avons observé ceci sur la première classe:

**Table 12 - Comparatif des moyennes et écarts-types sur 5 caractéristiques (classe 1)**

401 Enregistrements	Moyenne	Écart-type	Diff
* AMU	0.556	0.154	0.402
* Corp Profits US	0.533	0.234	0.299
* Crude Oil Price USD	0.528	0.239	0.289
* Cycles Since Overhaul	0.002	0.184	-0.182
* Days between events	0.184	0.136	0.048

La caractéristique temps, est la seule pour laquelle la valeur de la moyenne est proche de la valeur de l'écart-type. Donc les données se ressemblent beaucoup selon la séparation effectuée. La même observation s'effectue sur la deuxième classe:

**Table 13 -Comparatif des moyennes et écarts-types sur la caractéristique la plus influente (classe 2)**

401 Enregistrements	Moyenne	Écart-type	Diff.
* Days between events	0.151	0.068	0.083

Il s'agit encore de la différence la plus faible constatée.

### **5.3 Analyses de performance**

Avec l'algorithme du « *K-mean* », pour les résultats discutés à la section 2, voici les informations relatives à la performance.

Avec un total de 20 dimensions (Historique de visites en usine et facteurs environnementaux), pour séparer l'ensemble en 10 classes, l'algorithme fait 9 itérations. Pour séparer l'ensemble en 5 classes (Historique de visites en usine et facteurs environnementaux), il effectue 6 itérations.

Si nous utilisons seulement les 14 dimensions relatives à l'historique de visites en usines, afin de séparer l'ensemble en 5 classes, alors l'algorithme effectue 12 itérations. Si l'on augmente le nombre de classes à 10, pour les mêmes données, alors l'algorithme doit effectuer 14 itérations.

En ajoutant les données macro-économiques à l'historique de visites en usine, alors le nombre d'itération diminue à 11. Pour les mêmes données, sur 5 classes plutôt que 10, il doit effectuer 10 itérations, encore en 1 seconde.

Advenant le cas où nous diminuons le nombre de dimensions pour n'utiliser que les données relatives aux événements imprévus, alors le nombre d'itérations grimpe à 18.

Il est donc assez difficile d'atteindre les limites de cet algorithme, le délai pour résoudre un problème est assez court, mais d'un autre côté, les résultats obtenus ne concordent pas nécessairement avec ceux attendus.



Pour ce qui est du SVM, plusieurs tests ont été effectués, au même titre que ceux sur le « *K-mean* ». Les résultats relatifs à la performance varient aussi, selon le nombre de dimensions des données, le nombre d'itérations que devra traiter l'algorithme et le nombre d'enregistrements.

Lorsque nous avons 380 enregistrements, ayant chacun 28 dimensions, l'algorithme effectue 10 itérations. C'est beaucoup plus qu'il n'en faut au « *K-mean* » pour résoudre un problème similaire, cependant, le SVM arrive avec des résultats très près de ceux attendus.

Revenons à 24 dimensions, si nous utilisons 648 enregistrements, l'algorithme s'effectue sur 2 minutes 40 secondes. De plus, l'algorithme donne de bons résultats très près de ce que l'on attend.

Avec 580 enregistrements de 30 dimensions chacun, sur 1 itération, alors il s'effectue sur 3 minutes 8 secondes. En ramenant le nombre de dimensions vraiment bas, de l'ordre de 4, alors il s'effectue sur 1 minute 55 secondes. Cependant, dans les deux cas, les résultats ne sont guère satisfaisants. De plus, si nous augmentons le nombre d'itérations à 6, le délai s'allonge et pour 10 itérations, l'algorithme ne termine pas correctement, il agit ainsi jusqu'à au moins 100 itérations. Nous n'avons pas poussé plus loin les essais.

En utilisant de nouveau SPSS, afin d'utiliser l'algorithme de « *clustering* » hiérarchique, voici les performances obtenues:

Il lui a fallu 5 minutes pour traiter 606 enregistrements non normalisés, ayant chacun 74 dimensions. Évidemment, les résultats ne sont pas très satisfaisants. Il met une seule minute si le nombre de dimensions est de 30. Dans ce cas, les résultats sont encore plus difficilement interprétables. [DataSet2]  
"G:\PWC\Juin 2008\11 JUIN 2008 SPSS\06-06-08 - historique complet +DBE 2.sav"

On peut représenter les paramètres t (Temps), nb.ite (Nombre d'itérations « runs ») et dim (dimension) sur les derniers essais graphiquement aux figures 19 et 20.

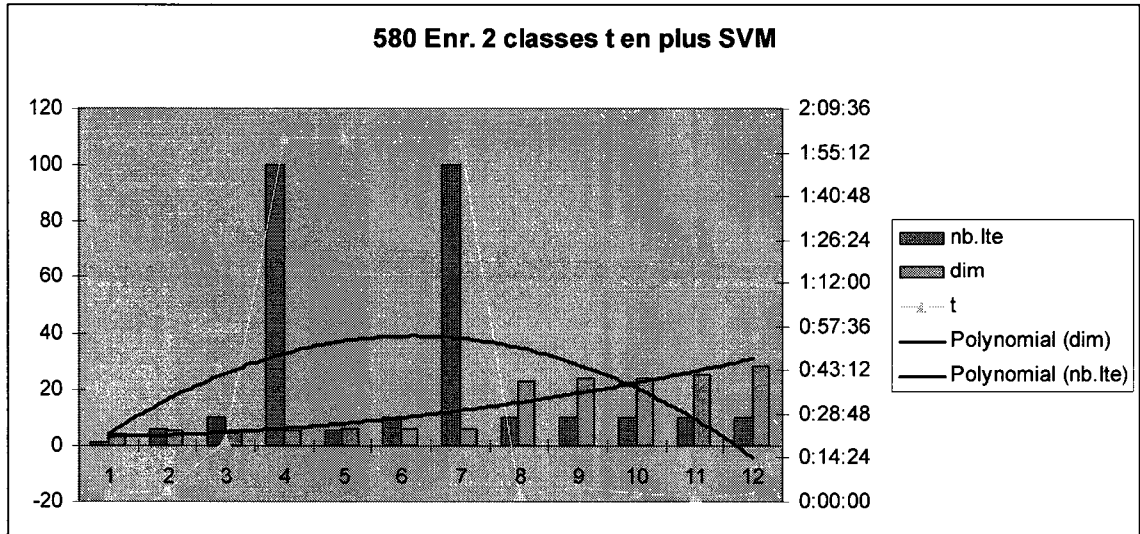


Figure 19 - Paramètres triés en nombre croissant de dimensions (nb de caractéristiques utilisées)

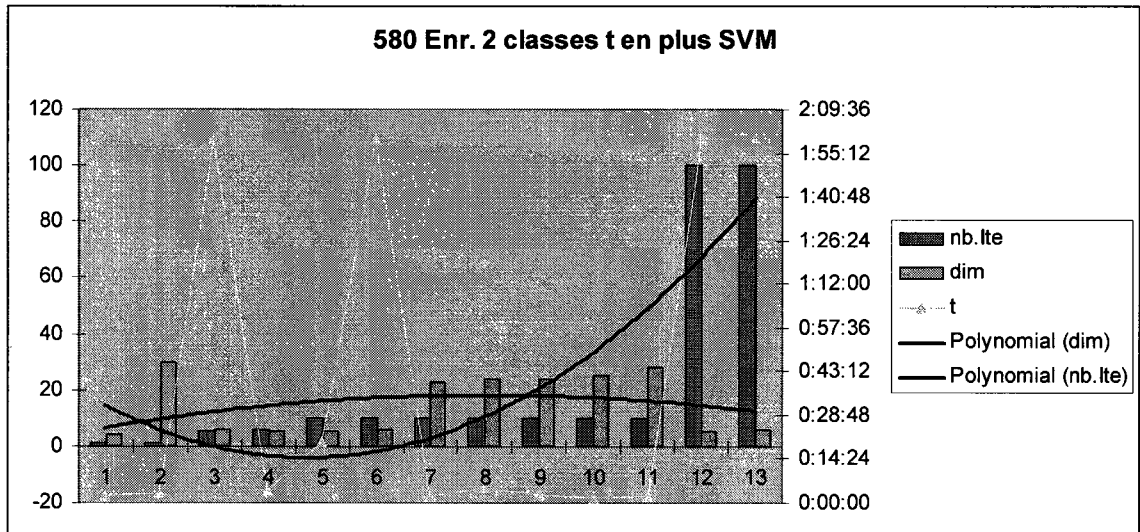


Figure 20 - Paramètres triés en nombre croissant d'itérations

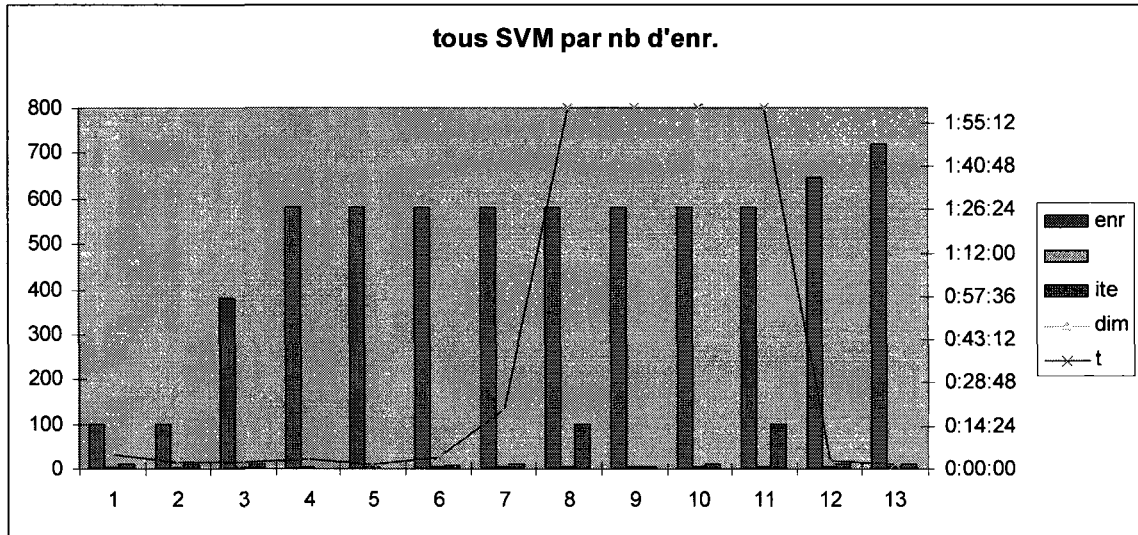


Figure 21 - courbe de temps sur les essais SVM triés par nombre croissant d'enregistrements

En termes de performances, il est clair que l'élément clé demeure principalement le temps d'exécution d'un algorithme, sur sensiblement les mêmes jeux de données. On remarque qu'avec la méthode du SVM, deux éléments influencent directement le temps d'exécution. Ils sont présentés graphiquement à la figure 21. Il s'agit d'abord de la dimension des données; plus le nombre de dimensions est faible, plus il sera difficile de séparer un ensemble en plusieurs groupes et l'algorithme aura besoin de beaucoup de temps pour s'effectuer. Ensuite, le nombre d'itération; plus le nombre d'itération à effectuer est élevé, plus il faudra du temps pour que l'algorithme s'exécute. Si nous poussons ces deux derniers exemples à l'extrême, alors l'algorithme ne se termine pas normalement. Si l'on s'en tient au bon équilibre entre le nombre de dimensions suffisantes et un bon nombre d'itération, l'algorithme donne des résultats satisfaisants.

## Chapitre 6 – Conclusion

### 6.1 Conclusion et travaux futurs

Dans ce travail, nous avons, d'un coté, présenté notre problématique liée à un domaine industriel très important qu'est l'aéronautique et, d'un autre coté, intégré la technique de classification basée sur l'approche du « Support Vector Machines » pour résoudre la partie essentielle de notre projet. Plusieurs implémentations du SVM existent déjà et sont disponibles sur le web. Une liste de réalisations de SVM se trouve à :

<http://www.kernel-machines.org/software.html>;

<http://svmlight.joachims.org/>;

<http://asi.insa-rouen.fr/enseignants/~arakotom/toolbox/index.html>

Nous avons développé un logiciel sous l'environnement oracle avec l'intégration du SVM, permettant d'effectuer les tâches de classification et de prédiction dans les bases de données de grande dimension.

De tout ce qui précède et au complément de la tâche de classification, le SVM est conçu aussi pour accomplir les tâches de « clustering » et de régression. La tâche de régression est notre sujet d'étude dans laquelle nous intégrons le SVM pour résoudre le problème de prédiction des visites des moteurs à l'atelier de

maintenance. C'est-à-dire la prédiction d'une variable par rapport à plusieurs autres variables (régression simple et multi variée). Ceci constitue une nouvelle technique que nous avons appliquée sur la construction de notre modèle, qui consiste en la combinaison du SVM et les arbres de décision pour établir une classification et en même temps extraire les règles de décision qui nous permettent la prédiction des visites en fonction de ces variables.

Afin de s'assurer des performances du modèle, ainsi que de son efficacité à résoudre le problème proposé, nous avons effectué plusieurs tests impliquant chacun plusieurs jeux de données. Chaque jeu ayant des spécifications bien différentes des autres, nous tentions de séparer l'ensemble d'événements selon plusieurs cibles également, afin de vérifier les limites de l'algorithme. Bien entendu, avec un nombre de dimensions égal à 23, ou 28 [voir résultats chap.4, p.53], le SVM donne d'excellents résultats pour notre problème de maintenance de moteurs d'aéronefs, mieux que ne le fait le « K-mean » et l'algorithme des Kppv, sur les mêmes jeux de données. Sur l'aspect type de maintenance à accomplir, maintenant, le SVM domine les arbres de décisions, qui nous donnaient des modèles compliqués, toujours aux limites de notre compréhension vu le nombre de dimensions impliquées. Les classes résultantes n'affichaient pas nécessairement les relations de ressemblances attendues entre les enregistrements d'une même classe. Là où les arbres de décisions ont joué un rôle déterminant, c'est dans leur capacité de résoudre la dimension liée au temps (on peut relire au chap. 4 p. 66).

Ce modèle s'avère être le plus performant des algorithmes évalués pour trouver une solution au problème proposé. Bien sûr, beaucoup de travail reste encore à faire pour bien « fonder » cette méthode comme nouvelle technique hybride de classification/prédiction en forage de données. Ainsi, un arbre de décision qui peut se construire de plusieurs façons, aura maintenant des machines à vecteur de support comme nœuds de scission, cela augmente de beaucoup son efficacité sur les problèmes de grande taille.

Ce projet a duré deux ans et il faudrait encore quelques années pour effectuer une comparaison qui soit juste et valable, sur un autre problème similaire. Ceci reste une piste qui aura certes besoin de développements avant d'atteindre la maturité. Imaginez le pouvoir de classification/prédiction actuel quadruplé, comment cette percée pourrait faire avancer d'autres domaines comme l'étude du génome humain, les classifications d'espèces, de planètes et systèmes solaires, d'œuvres musicales...

# Annexe A - Les différentes sources de données disponibles et la description du contenu de chacune d'elles.

## A.1 Service Bulletins

Format: Documents PDF

Contenu: Le concept modulaire, les intervalles d'entretiens suggérés (TBO, HSI), le temps de service des lames de turbines et de certains accessoires.

### SERVICE BULLETIN

P&WC S.B. No. 18003

TURBOPROP ENGINE  
OPERATING TIME BETWEEN OVERHAULS AND HOT SECTION INSPECTION FREQUENCY

#### 3. Accomplishment Instructions (Cont'd)

TABLE 1,

Model	Gas Generator Module	Power Section Module
PT6A-68	P/N 3047000	P/N 3047200
PT6A-68A	P/N 3120333	P/N 3120334

(3) Each module may be maintained at its own hard time schedule.

#### B. Time Between Overhaul (TBO) Recommendations:

**NOTE: 1.** Industry TBO is the Pratt & Whitney Canada Inc. recommended TBO per this service bulletin which is applicable to all operators.

**NOTE: 2.** Fleet TBO is the TBO level which individual operators have attained for engines of the same model in their possession only.

- (1) Rotor component life limitations outlined in the latest revision of P&WC S.B. No. 18002 override TBO considerations.
- (2) For new operators of PT6A-68 and PT6A-68A engines with no prior PT6 engine experience, it is recommended that one engine be sampled at 4500 hours.
- (3) For operators not on an approved modular program and for military operators, the currently recommended industry TBO for all PT6A-68 and PT6A-68A engines is 4500 hours.
- (4) For operators on an approved modular program, the industry TBO for the two modules are:

TABLE 2,

Model	Gas Generator Module	Power Section Module
PT6A-68	4500	4500
PT6A-68A	4500	4500

(5) Basic engine accessories as defined in the applicable Illustrated Parts Catalogs may be operated to the engine/module TBO plus 500 hours. Where accessories are removed for shop repair and subsequently reinstalled, operating time since new or overhaul must be recorded on the repair tag.

#### C. Hot Section Inspection (HSI) Frequency Recommendations:

- (1) All PT6A-68 engines shall be operated to a scheduled hot section inspection frequency not to exceed 2250 hours.

3. Accomplishment Instructions (Cont'd)
  - (2) Alternatively, the scheduled HSI interval may be based on Engine Condition Trend Monitoring in accordance with Engine Condition Trend Monitoring Analytical Guide Manual P/N 3043607. However, if the trend monitoring method is used, the hot section shall be inspected with a borescope at the nearest convenient time (i.e. fuel nozzle inspection) prior to 2250 hours and at subsequent intervals not to exceed 500 hours until a hot section inspection is accomplished.
4. Appendix  
Not applicable.

## A.2 Maintenance Manuals

Format: Documents PDF

Contenu: Les cycles de vie des pièces, les intervalles d'entretiens suggérés (TBO, HSI)

### SCHEDULED MAINTENANCE CHECKS

#### 1. General

This section gives more details of the inspection and maintenance tasks which are necessary at the times given in Chapter 05-10-00 (Task 05-10-00-990-802). It also refers to the full procedure included in this Engine Manual.

Table 1 gives the maintenance checks for engines on a hard-time maintenance program.

Table 2 gives the maintenance checks for engines on an on-condition maintenance program. The engine is removed from the aircraft when one module/submodule is defective. Refer to Table 2 to get instructions for the defective module/submodule and for analysis of the serviceable module/submodule.

TABLE 1, Scheduled Maintenance Checks for Engine on Hard-Time

AREA	INSPECTION/ TASK	REFERENCE
<b>Overhaul</b>		
Engine	Disassemble, clean and do a full inspection of the detail parts. Do the applicable repairs and reassemble the engine.	72-00-11 thru 72-00-53 to separate modules 72-11-00 thru 72-53-00 to do the overhaul
Reduction Gearbox	Disassemble, clean and do a full inspection of the detail parts. Do the applicable repairs and reassemble the engine.	72-00-11 to remove gearbox 72-11-00 to do the overhaul
<b>Heavy Maintenance</b>		
Hot Section Inspection	Disassemble the engine rear end up to the HP stubshaft. Do a full inspection on all parts and repair damaged parts. Reassemble the rear end of the engine.	72-00-00 Inspection/Check-04 (Task 72-00-00-280-811)
Specific Repair	Remove affected module, disassemble to get access to the detail parts affected and do the applicable inspections and repairs. Reassemble and install the module.	72-00-11 thru 72-00-53 for removal/installation of affected module 72-11-00 thru 72-53-00 to disassemble affected parts, inspections, repairs and reassembly
Limited-life Parts	Disassemble the engine sufficiently to get access and replace the time-expired components.	72-00-00 Inspection/Check-06 (Task 72-00-00-280-830)



### **A.3 Shop Visit Historical Data (TRAC, TRS)**

Format: Base de données reliée à un système SAS

Contenu: La liste des événements passés, leur type, leur description et le nombre d'heure de vol.

### **A.4 Warranty Data**

Format: Base de données reliée à un système UNIX (Oracle)

Contenu: La liste des réclamations de garanties, leur type, leur description et le nombre d'heure de vol.

### **A.5 Part Sales Data**

Format: Base de données reliée à un système UNIX (Oracle)

Contenu: La liste des transactions de pièces, leur numéro et leur description.

### **A.6 Modèles de prévision de chaque département**

Format: Fichiers MS Excel et MS Word.

Contenu: La méthode employée pour estimer le nombre d'événements sur un modèle ou une famille de modèles précis

#### **A.6.1 Exemple de modèle de prédiction actuel fait sur MS Excel (L'anglais est utilisé dans l'industrie):**

*« This current model has been created with MS Excel spreadsheets. Inside a workbook, there are 5 tabs involved in the creation of the model. Here is the description of each tab and their cells:*

*Sheet "Crashed – 25": The list of crashed engines – Engines involved in crashes are no longer flying, so will not require any future repairs or overhauls. Involving these fields:*

*Engine Serial Number, Engine Model, Operator Name, Engine Disposition, Removal Category, Removal Type, Flight Effect, Substantiation, State/Province, Symptom, Condition, Engine Comment, Engine Time Since New, Engine TSN Recorded Date, Time Since Overhaul, Extended Module Operation Hours, Engine Average Monthly Utilization, Extended time to 31 Dec 06.*

*Then Calculation of: Extended Module Operation Hours + (6\*Engine Average Monthly Utilization) And Calculation of: Extended time to 31 Dec 06 > 3500 then YES else NO*

*New Not Installed – 37: List of the new engines, not installed yet – These engines are not mounted on an aircraft yet, so will not be accumulating any flying hours, so they will not be included in any future repairs or overhauls. Once mounted on an aircraft, they then become part of the fleet of other engines. Involving these fields:*

*Engine Serial Number, Engine Model, Operator Name, Engine Disposition, Removal Category, Removal Type, Flight Effect, Substantiation, State/Province, Symptom, Condition, Engine Comment, Engine Time Since New, Engine TSN Recorded Date, Time Since Overhaul, Extended Module Operation Hours, Engine Average Monthly Utilization, Extended time to 31 Dec 06.*

*Calculation of Extended time to 31 Dec 06 > 3500 then YES else NO*

*Sheet 2: The forecasting model itself involving these fields, separated in 2 tables:*

*Year 2002 (thru 2011) Actual / Forecast Current # of engines, Feb 2004 Original Projection of # of Engs Flying, 16 Sep 05 Estimate of # of engines flying, Actual / Forecast Engine Deliveries. Calculation of the years before. Like in 2004:*

*= Actual / Forecast Current # of engines 2004 + Actual / Forecast Current # of engines 2005 (Due to the increase in the number of engs per year, expect the fleet hours to increase accordingly. Note, the AMU used for the flying engines was used in this calculation)*

*Percentage increase in engines per year: Calculation = (I6-I5)/I5*

= 16 Sep 05 Estimate of # of engines flying precedent year - 16 Sep 05  
 Estimate of # of engines flying current year / current year Annual Actual /  
 Forecast Fleet Hours. Calculation for current year

=SOMME(N163:N168)\*2.07

=Sum of all times a factor of 2.07

((<?> hours). Note: If the AMU from the Current Eng Population is used,  
 the AMU is 24.7 hours.)

Actual / Forecast Avg. AMU for Current Eng Pop'n

Avg AMU for Flying Pop'n, Target Reliability Rate, Last two year's average  
 BPR Rate, Last two year's average BUR Rate, Last two year's average NUR Rate,  
 Last two year's average NPR Rate, TSV Rate

Calculation of =SOMME(P7:R7)

= sum (Last two year's average BPR Rate, Last two year's average BUR  
 Rate , Last two year's average NUR Rate)

Table 14 - « calculations for the years presented in the workbook »

	BPR	Overhauls	BUR	NUR	NPR	Total
2006	=\$L7/(1000/P7)	42	=\$L7/(1000/Q7)	=\$L7/(1000/R7)	6	=SOMME(V7:Z7)
2007	=\$L8/(1000/P8)	11	=\$L8/(1000/Q8)	=\$L8/(1000/R8)	6	=SOMME(V8:Z8)
2008	=\$L9/(1000/P9)	6	=\$L9/(1000/Q9)	=\$L9/(1000/R9)	6	=SOMME(V9:Z9)
2009	=\$L10/(1000/P10)	7	=\$L10/(1000/Q10)	=\$L10/(1000/R10)	6	=SOMME(V10:Z10)
2010	=\$L11/(1000/P11)	28	=\$L11/(1000/Q11)	=\$L11/(1000/R11)	6	=SOMME(V11:Z11)
2011	=\$L12/(1000/P12)	47	=\$L12/(1000/Q12)	=\$L12/(1000/R12)	6	=SOMME(V12:Z12)

This table fits the years presented in the field Year Ex: 2002 (thru 2011). Still in  
 this table, there are notes about the fields:

BPR => BPRs are separate from O/Hs i.e. HSIs. There have only been 81  
 BPRs that were not BPOs since 1995, so there are only about 7 per year. XX200

*does not have scheduled HSIs; however, occasionally one is done and they are recorded as BPRs. Campaigns would also be categorized as BPRs.*

*BPR is calculated here with Annual Actual / Forecast Fleet Hours \* (1000/Last two year's average BPR Rate)*

*Overhauls => See the separate worksheet on XX200 O/Hs herein*

*BUR is calculated here with Annual Actual / Forecast Fleet Hours \* (1000/Last two year's average BUR Rate)*

*NUR is calculated here with Annual Actual / Forecast Fleet Hours \* (1000/Last two year's average NUR Rate)*

*NPR => Typically these involve rental removals, or convenience removals by the Operator i.e. One eng is removed in a twin eng a/c and the Operator decides to send the other one at the same time for some type of maintenance/repair*

*TOTAL => Sum of BPR, Overhauls, BUR, NUR, NPR.*

*At the end of that table comes a total row for all "rates" fields "Two year average used to calculate removals for OMM", involving the sum of each column.*

*Sheet "2006 Overhauls": A list of the Overhauls for 2006*

*Involving these fields:*

*Engine Serial Number, Engine Model, Operator Name, Engine Disposition, Removal Category, Removal Type, Flight Effect, Substantiation, State/Province, Symptom, Condition, Engine Comment, Engine Time Since New, Engine TSN Recorded Date, REVISED Engine Time Since New, REVISED Engine TSN Recorded Date, Time Since Overhaul, 19 Jul 06 Extended Module Operation*

*Hours, Engine Average Monthly Utilization, REVISED Engine Average Monthly Utilization, Extended time to 31 Dec 06. Calculation of \$R3+(6\*\$S3)*

*= 19 Jul 06 Extended Module Operation Hours+ (6 \* Engine Average Monthly Utilization) Ext TSO eff. 31 Dec 06*

*Calculation of U3-3500*

*= Extended time to 31 Dec 06 - 3500 2006 overhaul (y.n)*

*Calculation of SI(U3>3500,"Y","N")*

*=IF Extended time to 31 Dec 06 > 3500 THEN Y ELSE N REVISED*

*Extended TSN as of Dec 31, 2006. Calculation of O3 + ((\$Y\$1-P3)/(365/12)\*T3)*

*= REVISED Engine Time Since New + ((31/12/2006 – REVISED Engine TSN Recorded Date) / (352/12) \* REVISED Engine Average Monthly Utilization)*

*REVISED 2006 Overhaul (y-n). Calculation of SI(Y3>3500,"Y","N")*

*=IF REVISED Extended TSN as of Dec 31, 2006 > 3500 THEN Y ELSE N*

*Considering also:*

**Table 15 - « OH Completed by CSC Berlin in 2006 and Belgian Police listing »**

BB0008	XX206B	Air Lloyd Luftfahrt Technik GmbH
BB0006	XX206B	ESP / Copterline OY
206040	XX206A	ESP/Noordzee Helikopters Vlaanderen
BA0020	XX206A	PHuSt Niedersachsen
BA0024	XX206A	PHuSt Niedersachsen
<b>Belgian Police</b> Aircraft 900-0034 Last time at Maintenance Facility March 2006 Aircraft Hours 4399 hrs LHE 4225 hrs RHE 2388 hrs Probably won't be in for overhaul this year.		

Aircraft 900-0045  
 Last time at Maintenance Facility June 2006  
 Aircraft Hours 4002 hrs  
 LHE 999 hrs  
 RHE 891 hrs  
 Probably won't be in for overhaul this year.

Aircraft 900-0038  
 Last time at Maintenance Facility July 2006  
 Aircraft Hours 2800 hrs  
 LHE 2674 hrs (this is probably the one I did the extension on)  
 RHE 1643 hrs  
 Probably won't be in for overhaul this year.

Sheet "Model200 Overhaul Master":

*Involving these fields:*

*Engine Serial Number, Engine Model, Operator Name, Engine Time Since New, Engine TSN Recorded Date, Time Since Overhaul, REVISED Engine Time Since New. Calculation of RECHERCHEV (A99, ADACXX206\_060601!\$C\$2:\$F\$51, 4, FAUX)*

*=VerticalSearch in sheet " ADACXX206\_060601", searching with the Engine Serial Number, to get " TW" with the exact value only.*

*REVISED Engine TSN Recorded Date, 19 Jul 06 Extended Module Operation Hours, Engine Average Monthly Utilization*

*Extended time to 31 Dec 06: Calculation of \$R99+(6\*\$S99)*

*= 19 Jul 06 Extended Module Operation Hours + (6 \* Engine Average Monthly Utilization)*

*Ext TSO eff. 31 Dec 06: Calculation of T99-\$O99-M99*

*= Extended time to 31 Dec 06 - Time Since Overhaul - Engine Time Since New*

*2006 overhaul (y.n): Calculation of SI(U99>3500,"Y","N")*

*=IF Ext TSO eff. 31 Dec 06 > 3500 THEN Y ELSE N*

*Extended time to 31 Dec 07: Calculation of  $\$S99*12+T99$*

*= Engine Average Monthly Utilization \* 12 + Extended time to 31 Dec 06  
Ext TSO eff. 31 Dec 07: Calculation of  $\$S99*12+U99$*

*= Engine Average Monthly Utilization \* 12 + Ext TSO eff. 31 Dec 06  
2007 overhaul (y.n): Calculation of  $SI(X99>3500,"Y","N")$*

*=IF Ext TSO eff. 31 Dec 07 > 3500 THEN Y ELSE N*

*Extended time to 31 Dec 08: Calculation of  $W99+12*\$S99$*

*= Extended time to 31 Dec 07 + 12 \* Engine Average Monthly Utilization  
Ext TSO eff. 31 Dec 08: Calculation of  $12*\$S99+X99$*

*=12 \* Engine Average Monthly Utilization + Ext TSO eff. 31 Dec 07  
2008 overhaul (y.n): Calculation of  $SI(AA99>3500,"Y","N")$*

*=IF Ext TSO eff. 31 Dec 08 > 3500 THEN Y ELSE N*

*Extended time to 31 Dec 09: Calculation of  $Z99+12*\$S99$*

*= Extended time to 31 Dec 08 + 12 \* Engine Average Monthly Utilization  
Ext TSO eff. 31 Dec 09: Calculation of  $AA99+12*\$S99$*

*= Ext TSO eff. 31 Dec 08 + 12 \* Engine Average Monthly Utilization  
2009 overhaul (y.n): Calculation of  $SI(AD99>3500,"Y","N")$*

*Ext TSO eff. 31 Dec 09 > 3500 THEN Y ELSE N*

*Extended time to 31 Dec 10: Calculation of  $AC99+12*\$S99$*

*= Extended time to 31 Dec 09 + 12 \* Engine Average Monthly Utilization  
Ext TSO eff. 31 Dec 10: Calculation of  $AD99+12*\$S99$*

*=Ext TSO eff. 31 Dec 09 + 12 \* Engine Average Monthly Utilization*

2010 overhaul (y.n): Calculation of SI(AG99>3500,"Y","N")

=IF Ext TSO eff. 31 Dec 10 > 3500 THEN Y ELSE N

Extended time to 31 Dec 11: Calculation of AF99+12\*\$\$S99

= Extended time to 31 Dec 10 + 12 \* Engine Average Monthly Utilization

Ext TSO eff. 31 Dec 11: Calculation of AG99+12\*\$\$S99

= Ext TSO eff. 31 Dec 10 + 12 \* Engine Average Monthly Utilization

2011 overhaul (y.n): Calculation of SI(AJ99>3500,"Y","N")

=IF Extended time to 31 Dec 11 > 3500 THEN Y ELSE N

Sheet "ADACXX206\_060601" Involving these fields, in date of: 1st June 2006:

Table 16 - ADACXX206\_060601 zone 1

TYPACSNR	PARTNUMBER	SERIALNR	TW	PARTNAME	TT_STD
L1350007	3117200-01	XX0008	1	TRIEBWERK XX206B	3217.42

Table 17 - ADACXX206\_060601 zone 2

TT_ICY	TT_GCY	TT_PCY	SB28280	Remarks	
10162.00	7231.00	3935.00	incorporated		

Considering this model as an example of the current forecasting methods, there is a real need for improvement. Note: The expert has advised that the forecast for year 2006 was for 192 engine removals; there were 194, accuracy to within 1.1%. However, previous years forecasting efforts were off by as much as 50%. Mike has stated that his methodology has evolved, which has allowed for better event forecasting. For example, the average monthly usage (AMU) information provided from TRS is reviewed for every engine, at the time, there were more than 1,400 engines to review. Discussion with Service Engineering personnel was done to see if the AMU was accurate or not. Service Engineering personnel have an intimate understanding of their engine model operators. Having the correct AMU is one of the most important elements to forecasting, since knowing



*the number of hours an operator flies per year and knowing his maintenance interval, i.e. an overhaul is required every 3,500 hours for a XX200, is key to this task. »*

### **A.6.2 Exemple de notes relatives aux rencontres avec les experts:**

*« Meeting with Mike Parent. 03 May 2007. 9:00-9:30: Model 308/305.*

*Every quarter they meet with Service engineering and Plant 5, to talk about shop visit, over 2012. Their data is to be considered from a parts point of view and HIS / O/H / Campaigns are considered.*

*Service engineering sends to JF Desrosiers, who sends to the 300 group and is then sent to Anna Cerone (for regression). Warranty campaigns are to consider too, then we can build a sale's plan (input/output RPO). SB Compliance codes (most important) Range 1-10. Ranges 1-6 generate most activities (most priority) for the workscope in the shop.*

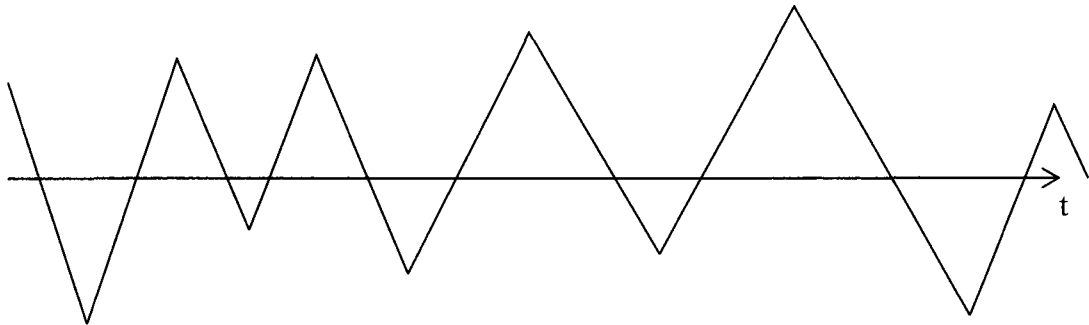
*For stock requirements see in SAP (main MRP driver for parts). Plant 17 got the spare engines, Based on the sales history / Consumption history (Usually go out 3 years, 26 000 parts for 308/305).*

*An automatic forecast is done by the system for Bulk items, some parts have a consistent consumption (it's Automatic). Delay to react (1-2 mo.) goes out of balance.*

*Also use: O/H manuals, Parts manuals, Maintenance manuals*

*Key figure: Stock value (SAP). Review of the inventory for critical parts, such as fast movers (expensive parts) (X-parts). They are 20% of the parts, but generate 80% of the revenues.*

*Consumption curve shall look like this:*



*It's never really a flat line or curve... like this one*



»

## **A.7 Notes des administrateurs de système**

Format: Fichiers MS Word

Contenu: Les notes, commentaires, recommandations des administrateurs sur l'utilisation et le sens des données

Exemple de notes des administrateurs systèmes: « *Part Sales Meeting July-17-2007, from 1:30 to 2:45* »

Guy Malette, Louise, JF Dubois et Jean Heydra.

Voici les questions que j'avais préparées:

« *A document with simulation/actual thing about part sales?* »

- « *A part thru the complete process?* »
- « *What are the flat parts? Trend is flat.* »
- « *Parts that are selling fast / slow?* »
- « *Parts gradually sold less... sales slowing down? Why?* »
- « *Parts for which sales are increasing slowly?* »
- « *Do you look at service bulletins to understand what parts will be needed to increase engine design?* »
- « *Are parts available when they incorporate SBs?* »
- « *Is there a process for parts availability?* »
- « *TBO, TSN, TCN... → See Plant 5 or SBs (ex: model 900)?* »

Ils ont des données Transactionnelles.

Seulement des pièces neuves. Pas de reconditionnées, ni réparées.

Approx. 25 000 items par année.

Je veux tenter d'obtenir à qui on vend ces pièces, en quelle quantité, à quel prix, pour quel moteur... NON. On n'a pas le « *model name* ». On a seulement le « *model family group* ».

2 types d'information dans files. 22 000 P/N (Part Number)

- 1- Variations d'inventaires
  - a. Chaque mois
  - b. Chaque quart (quater)
  - c. Selon la préférence de livraison du client
  - d. Selon l'usage du client (son assiduité) et sa région géographique
- 2- Consommation

En analysant les données, voici le genre de renseignement qu'il serait utile de connaître:

- les pièces se vendent, mais pas seulement par P&WC. Des fois un atelier va réparer 500 moteurs, mais il va acheter seulement 50 pièces pour ses 500 moteurs. Pourquoi? Parce qu'il répare des pièces, il a peut-être d'autres fournisseurs (« *suppliers* »).

- Quelles pièces P&WC est le seul à vendre?

Donc, voici les champs de données que je vais recevoir. Ça provient de UNIX. C'est Jean Heydra qui va me sortir l'information, par années, de 2002 à Juin 2007.

P/N , Qty , Date , SoldTo , Cust Group , Keyword , IPC# , ELM

Cust.Group = {DOF, OEM, PWC/SCN}

ELM = {1=LRP, 2=HIS, 3=R, 4=O/H} où LRP = LineReplacementPart, qui va « toffer » jusqu'au prochain O/H ou HSI, R = Repaired. C'est le plus permissif. Les pièces pour le O/H seront classées comme O/H.

IPC = Config du moteur pour ensuite regardes dans les SBs. Par exemple, pour le 901, y'a un seul SB.

P/N = il fait référence au « *Part Number* », mais il se peut que dans l'historique une pièce cesse soudainement d'être vendue. Pourquoi? Une autre pièce la remplace, avec un autre numéro. Dans ce cas, Jean va trouver le « *reference-tail #* », qui sera le même pour toutes les pièces.

Donc, si on récapitule, il y a un problème possible ici: les pièces changent de no. de série.

PDC → A des pièces réparables = n'apparaissent pas dans leur chiffres.

L'idée (selon Jean Hydra):

En regardant cette l'équation (6) sur un modèle mature, ce serait un bon exemple:

$$\frac{\text{AMU T}^*}{\text{TBO}} \quad (6)$$

Par rapport à : « Fleet flown hours » & « fleet size ».

\* AMU cumulatif de l'appareil.

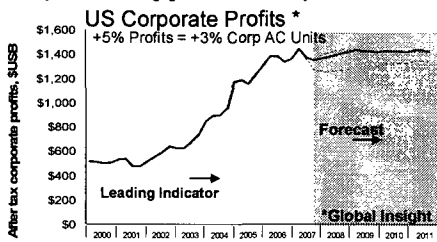
### A.8.1 Les facteurs macro-économiques

Format: Présentation MS PowerPoint

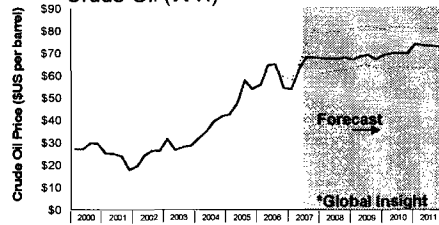
Contenu: US Corp. Profits, Oil price, S&P 500, World Real GDP, Nickel  
LIME cash price, Cobalt price, Titanium price, US PPI Metals & Metal  
product

#### September 2007

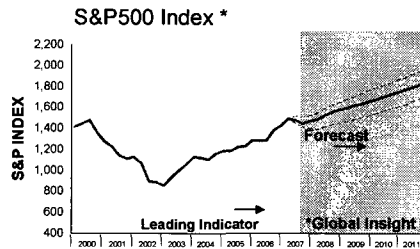
Profits leading Corp. AC deliveries by 2 to 3 years. Strong growth in delivery foreseen.



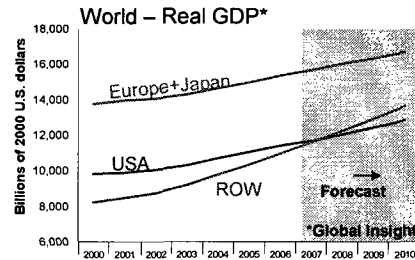
Oil stabilize at ~\$73. Good for Helicopter (exploration) and Regional T/Prop (vs. T/Fan).  
**Crude Oil (WTI) \***



S&P leads Corp. AC mkt. downturn by 9 to 12 months. Expect continuous growth.



World Economic balance is shifting. Emerging market orders are on the rise.



## A.8.2 Les facteurs environnementaux

Format: Classeur MS Excel

Contenu: Les caractéristiques environnementales (en taux) des principales villes et aéroports du monde.

Environment assessment for :  
05:Nov-07

Country	City or Airport Name	Zone	Sand	Salt	Volcano	Pollution	Elevation
Afghanistan	Kabul	2	0	1	0	0	0
Alaska	Anchorage Intl	4	0	1	0	0	20
Alaska	Aniak	3	1	0	0	0	615
Alaska	Bethel	3	0.25	0.25	0	0	18
Albania	Tirani	4	0	0.25	0	0	640
Algeria	Adrar	4	0.25	1	0	0	205
Algeria	Algiers Boumediene	2	0	0	0	0	64
Algeria	Bejaia	2	0.25	1	0	0	20
Algeria	Constantine	3	0.5	0	0	0	915
Algeria	El Golea	3	0.5	1	0	0	88
Algeria	El Oued	2	0	0	0	0	203
Algeria	Ghardaia	4	0	0.5	0	0	184
Algeria	In Amenas	2	0	1	0	0	16
Algeria	In Salah	2	0	0	0	0	1228
Algeria	Oran Es Senia	2	1	0	0	0.75	0
Algeria	Rhourd Nouss	2	0	0.1	0	0	2301
Algeria	Tebessa	2	0	0	0	0	285
Algeria	Touggourt	2	0	1	0	0	232
Angola	Benguela	3	0	0	0	0	0
Angola	Cabinda	2	0	0	0	0	24
Angola	Luanda	2	0	1	0	0	88
Angola	Soyo	3	0	0	0	0	82

## **Annexe B – Réponses aux questions posées dans la série d'étapes du plan de traitement des entrées.**

À l'aide du plan proposé, nous avons du répondre à plusieurs questions concernant les intrants du système et ainsi nous avons pu démystifier les éléments ayant une importance dans la résolution du problème. Voyons comment nous avons répondu aux différentes questions de chacune des étapes présentées.

### ***Étape B.1 – Ce qui est fait (Ce qui existe déjà)***

Quel(s) est (sont) L'(les)objectif(s) du modèle? Il existe 3 objectifs:

- 1- La prévision de ventes de pièces (pour le « *just in time delivery* »).
  
- 2- La prévision de capacité des centres de service (s'assurer que le nombre de moteurs que l'on attend ne surpassera pas la capacité physique et opérationnelle des centres).
  
- 3- Comme support pour les locations (Quand un moteur entre en centre de service, un moteur loué peut se retrouver sur l'avion. Il faut avoir assez de moteurs de location disponibles. Ce nombre dépend directement du nombre de moteurs que l'on attend dans les centres de service).

A quel niveau d'analyse considère-t-on les moteurs? Individuels, en groupes...? On doit considérer les moteurs comme individuels (pour effectuer la prévision d'événements sur un nombre  $x$  d'années), cependant plusieurs critères affectant l'état du moteur ne dépendent pas du moteur lui-même, mais plutôt de contexte dans lequel ce moteur est opéré. On doit donc considérer également l'opérateur (ainsi, ce moteur est intimement lié à tous les autres moteurs de cet opérateur). Revoyons la structure hiérarchique mise en place ci-dessous où on utilise une fonction d'appartenance. Par exemple, 1 opérateur détient de 1 à  $N$  aéronefs. Sur chaque aéronef est installé de 1 à 3 moteurs, et ainsi de suite. On peut voir la suite à la figure 22.

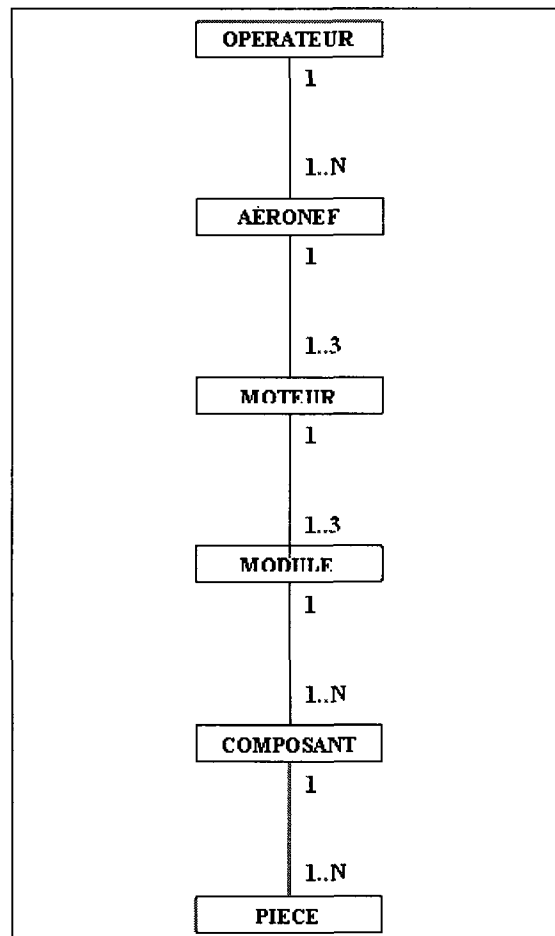


Figure 22 - Hiérarchie des éléments liés au moteur



De plus, chaque opérateur opère dans quelques régions et ces régions ont un impact sur la performance des moteurs. On parle de 4 régions types: **les régions polluées** (comme par exemple la Chine et la région de Beijing (la capitale mondiale de la pollution) ou encore les provinces de Shanxi au centre, Jiangxi et Zhejiang au Sud Est du pays (près de Shanghai), une des régions les plus peuplées et polluées de Chine), **les régions en haute altitude** (comme par exemple Mexico, au Mexique. Cette ville est située très haut en altitude 2240M (7349'). Il est donc plus difficile d'y opérer un moteur, puisqu'il se fatigue plus rapidement), **les régions désertiques** (comme par exemple le Nigéria, situé en Afrique du Nord sur la côte ouest. On y opère beaucoup de moteurs puisqu'il s'y fait de l'exploitation pétrolière. Il s'agit d'une région polluée et gravement affectée par le réchauffement climatique. On y voit une détérioration constante de l'environnement. Dès que les moteurs sont opérés près de sources sablonneuses, la maintenance doit être effectuée beaucoup plus tôt. Les moteurs deviennent plus fragiles et à haut risque d'événements imprévus), et finalement, **les régions salées**, généralement en zones chaudes entourées des mers (comme les caraïbes par exemple. Ces régions sont souvent de très fortes destinations vacances, donc les moteurs que l'on y opère sont souvent soumis à la chaleur mélangée au sel de la mer, ce qui augmente la possibilité de corrosion). Il nous faut donc, avec ces informations, dresser un profil d'opérateur. Ce profil doit contenir la mission de l'opérateur, son utilisation mensuelle (MRT « *Monthly Running Time* »), les paramètres du moteur (sa configuration: ses modules, ses SBs (« *Service Bulletins* »), ses MM (Manuels de Maintenance) et ses « *Airworthiness Documents* »).

L'expérience de l'analyste compte pour beaucoup. Il connaît sa flotte. Voici les commentaires quant aux types d'opérateurs:

On a 2 principaux types de maintenances. ST « *Soft-Time* » ou HT « *Hard-Time* ». « **Hard-Time** », c'est facile. Les HIS et les O/H sont à des heures fixes. Pour les « **Soft-Time** », c'est plus compliqué:

Les moteurs sont « *On-Condition* », Pour le HSI, les moteurs sont reliés à un système de transmission automatique des données (WebECTM) et des analystes suivent de près l'évolution des données sur la performance des moteurs. Nous devons nous baser sur l'expérience. Pour le O/Hs, le « *Soft-Time* » est de beaucoup plus économique. Les coûts majeurs sont les LCFs (« *Life-Limited Parts* »). On doit avoir une liste des pièces qui deviennent « *Time expired* ».

Comment est-elle générée la prévision?

1. On regarde la population totale. Ici on a approximativement 262 moteurs en fonction.
  - a. Pour chaque moteur:
    - i. Quelle est son historique? Les événements vécus jusqu'à maintenant, depuis la mise en service.
    - ii. Sa configuration? A-t-on incorporé certains SBs?
    - iii. Quels sont les LCFs? Voir la liste dans le MM.
    - iv. Quel est le type d'opérateur?
  - b. Exemple: pour un « *forecast* » sur 10 ans. Un moteur XX306B. Donc, à partir d'aujourd'hui (19OCT2007) jusqu'à dans 10 ans (19OCT2017), on veut tous les événements possibles dans ce temps. Une boucle qui s'arrête lorsque la date de la prévision est atteinte, ou que la prochaine prévision dépasse la date limite.
    - i. On commence avec le moteur 1 (le premier numéro de série). On regarde ses caractéristiques. Quel est la prochaine étape? Un HSI après x heures, basé sur usage mensuel (AMU: « *average monthly usage* ») = 1 Janvier 2008, on doit remplacer telles pièces. On met à jour la configuration pour la prochaine prévision.
    - ii. Moteur 2, mêmes étapes que le moteur 1.
    - iii. ... jusqu'au moteur N.
2. Quel serait le cas parfait? Deux moteurs sur un nouvel avion. Les deux sont enlevés en même temps, disons à 4000 heures. On a un « *Turn Around Time* » TAT de 30 jours pour faire la maintenance sur le moteur. Ensuite on doit considérer de 10 à 15 jours de transit et pour faire l'installation sur l'avion. Le deuxième événement devra considérer ce temps, puisqu'il n'y a aucun usage pendant ce temps.

a. Prenons un exemple. Un moteur débarque le 1<sup>er</sup> janvier, il sort le 15 février suivant et est installé le 1<sup>er</sup> mars. Il n'a accumulé aucune heures pendant ce temps. Cependant, dès le 1<sup>er</sup> mars, il recommence à accumuler des heures. On ré-simule le prochain événement, qui sera cette fois-ci un O/H dans 5000 heures, selon l'AMU. Nous devons toujours garder à l'esprit que la prévision reste toujours une probabilité, ce n'est jamais sûr à 100%. Tout cela est en partie basé sur l'expérience des centres de service et celle des opérateurs.

3. On reprend la prévision encore, en recommençant un 3<sup>ème</sup> cycle. On arrive à mi-2010, où l'on doit remplacer les LCFs.

Comparons les XX306B avec des moteurs plus récents (XX600), pour des événements prévus:

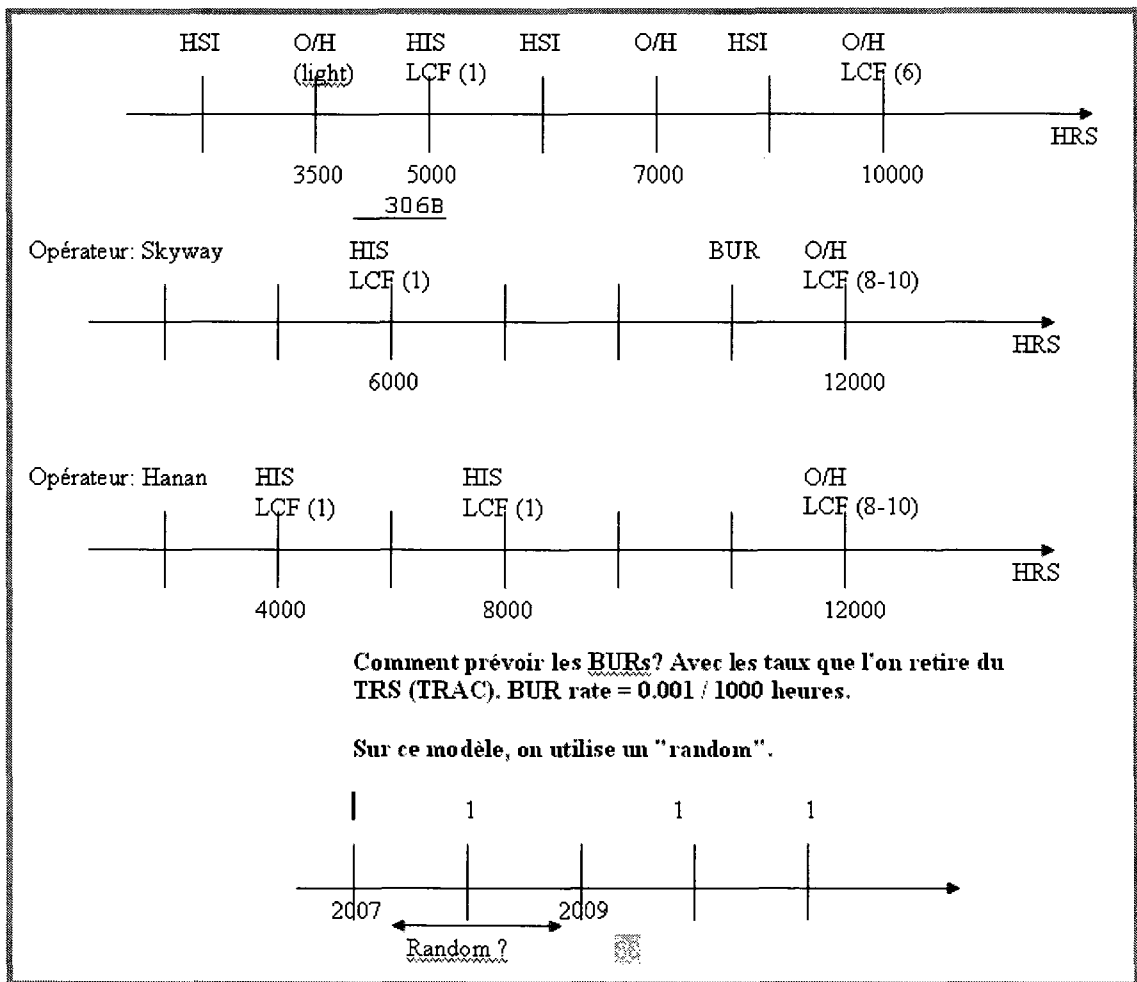


Figure 23 - Prévisions départementale, comparant le 306B au modèle le plus récent, le 600

Quelles sont donc les avenues d'améliorations possibles pour éviter que la prédiction ne soit que le fruit du hasard? On pourrait, par exemple prendre la somme des heures d'un opérateur et trouver les chances d'avoir un événement non prévu. On peut trouver le TAT pour l'ensemble de la flotte en utilisant un tableau ou un graphique, comme dans les figures 24 et 25.

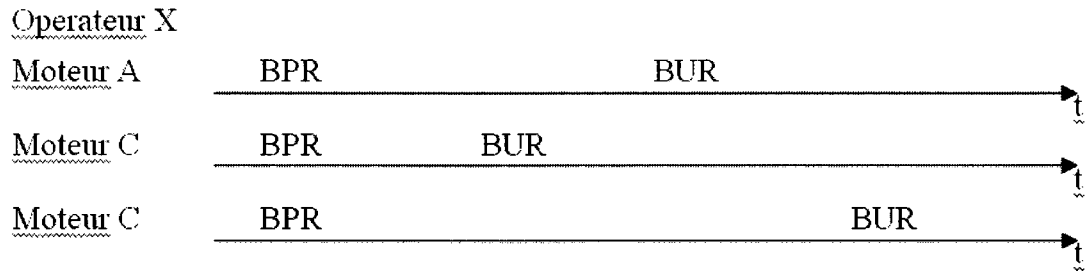


Figure 24 - Tableau de prévisions

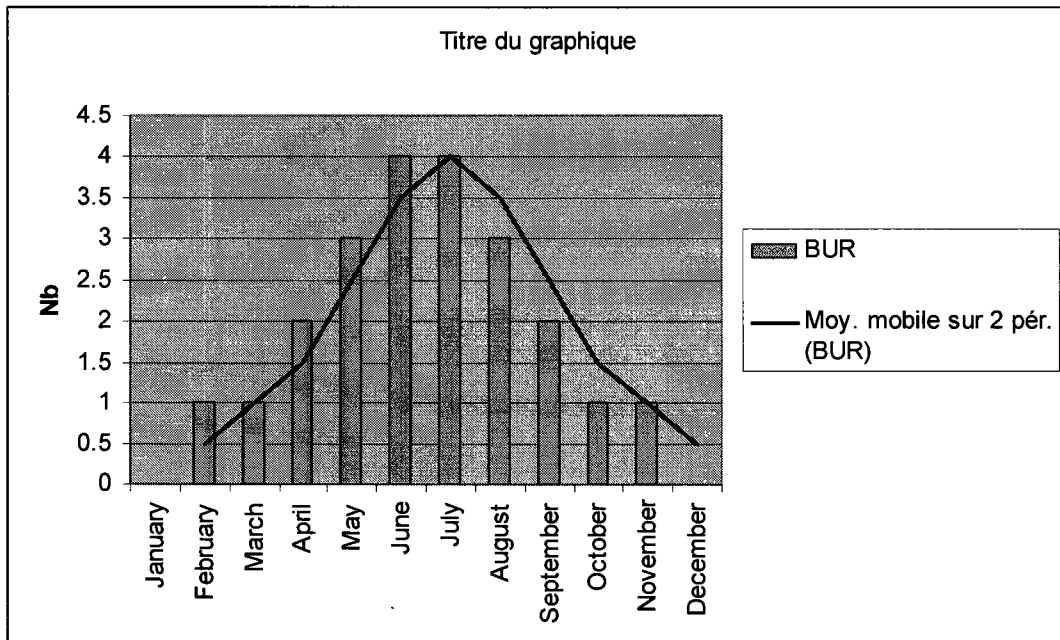


Figure 25 - Graphique de prévisions

Si on récapitule: on doit y aller moteur par moteur, on fait toutes les prévisions et ensuite on passe au moteur suivant. Lorsque tous les moteurs ont été faits, on ajoute les BURs sur toute la flotte entière.

Lorsque la flotte est faite, on regarde les tables des LCFs. Il y a trois types d'impact sur les LCFs. Type 1: 50/50 (pollué/non pollué). Type 2: 80/20 (80% propre et 20% de sable et sel). Type 3: « *Very Harsh* » (Dans le désert toujours).

On doit regarder les numéros de série pièces, l'original vendue avec le moteur n'a pas le même numéro de série que les pièces d'extension (Celles qui allongent la vie du moteur).

On peut trouver ces renseignements dans le « *Airworthiness Documents* », sur l'Intranet de l'entreprise. Voici la table des LCFs Utilisée pour la prévision:

Part Description	Part Num	Limit Type	AWL	Type 2 (Benign 80-20)		Type 1 (Harsh 50-50)		Very Harsh Equatorial		Type	PN Post Visit	SB
				Limit	Disposition	Limit	Disposition	Limit	Disposition			
Hub-Fan	30B4080-01	Cycle	14000	14000	3000	14000	3000	14000	3000	LCF	30B4080-01	-
Shaft LP Compressor	30B3897-01	Cycle	30000	30000	3000	30000	3000	30000	3000	LCF	30B3897-01	-
HPC1 Rotor	30B4011-01	Cycle	6000	6000	3000	6000	3000	6000	3000	LCF	30B4011-01	-
HPC1 Rotor	30B3951-01	Cycle	15000	15000	3000	15000	3000	15000	3000	LCF	30B4811-01	25082
HPC1 Rotor	30B4811-01	Cycle	15000	15000	3000	15000	3000	15000	3000	LCF	30B4811-01	25135
Rotor Drive	30B4149-01	Cycle	3000	3000	3000	3000	3000	3000	3000	LCF	30B4746-01	-
Rotor Drive	30B4539-01	Cycle	3000	3000	3000	3000	3000	3000	3000	LCF	30B4746-01	25046
Rotor Drive	30B4725-01	Cycle	3000	3000	3000	3000	3000	3000	3000	LCF	30B4746-01	25064
Rotor Drive	30B4728-01	Cycle	3000	3000	3000	3000	3000	3000	3000	LCF	30B4746-01	2506
Rotor Drive	30B4746-01	Cycle	3000	3000	3000	3000	3000	3000	3000	LCF	30B4746-01	25069
HPC2 rotor	30B4652-01	Cycle	15000	15000	3000	15000	3000	15000	3000	LCF	30B6082-01	25082
HPC2 Rotor	30B6082-01	Cycle	15000	15000	3000	15000	3000	15000	3000	LCF	30B6082-01	25177
HPC3 Rotor	30B3933-01	Cycle	15000	15000	3000	15000	3000	15000	3000	LCF	30B3933-01	25082

Figure 26 - LCFs utilisés dans la prédiction

Il existe aussi une table des opérateurs. Voir le transit time. Ça lui prend combien de temps à envoyer leurs moteurs, afin de monter une cédule d'entrée pour les usines, illustrant bien les délais. Par exemple, les opérateurs d'Europe prennent en moyenne 15 jours, alors que la Chine met 45 jours.

Le temps (délais total) peut se calculer ainsi: « *Transit from* » + TAT + « *Transit to* » + « *installation time* ». Voici la table des opérateurs utilisée:

Customer	Monthly Utilisation		Overhaul & Major Refurbishment		Type	Location	H.S.I	Maintenance Type	Contract	Contract #	Status	Applicatio
	Hours	Cycles	Limit	Disp								
328WW LLC	200	200	14000	2000	Type 2	England	3000	-	ESP	0923-01		Corporate
CLUB328	200	200	6000	2000	Type 2	England	3000	Hard Time	FMP	05-31		Corporate
Delta Simulation	100	100	14000	2000	Type 2	USA	5000	Continuous	FMP	04-1978	Parked	Avline
Aero Dienst	50	50	6000	2000	Type 2	Germany	3000	Hard Time	ESP	0922-01 1076-01		Corporate
Air Vallee	103	103	14000	2000	Type 2	Italy	5000		FMP	03-21		Corporate
Alvia	83	83	14000	2000	Type 2	USA	5000		JSSI	-		Corporate
AvCraft	1	1	14000	2000	Type 2	USA	5000		-	-	Parked	
Cirrus Attlog	100	100	14000	2000	Type 2	Germany	5000		-	-		Corporate
CS Aviation	50	50	14000	2000	Type 2	France	5000		-	-		Corporate
FFU2	50	50	14000	2000	Type 2		5000		-	-		Corporate
Gandalf	1	1	14000	2000	Type 2	Italy	5000		-	-	Parked	Corporate
Gessman	35	21	14000	2000	Type 2	Austria	5000		-	-		Corporate
Hainan	180	170	14000	3000	Type 1	China	5000	Continuous	FMP	01-1434		Avline
Independence Air	200	200	14000	2000	Type 2	USA	5000	Continuous	-	-	Parked	Avline
Itali Airline	145	172	14000	2000	Type 2	Italy	5000		-	-		Avline
Johnson	100	140	14000	2000	Type 2	USA	5000		JSSI	-		Corporate
Leasco	100	100	14000	2000	Type 2	Canada	5000	Continuous	-	-		Corporate
Nordic Partner	100	100	14000	2000	Type 2	Iceland	5000		-	-		Corporate
Oldenduff Carriers	50	50	14000	2000	Type 2	Germany	5000		-	-		Corporate

Figure 27 - Liste des opérateurs utilisée dans la prédiction

Y a-t-il d'autres connaissances liées? Afin de savoir, on peut regarder le « *work scope* ». (le TAT associé). Ils définissent un moteur par modules (voir la hiérarchie). Le « *WSName* » nous dit quel module entre à l'atelier. Par exemple: HPT (LCF 1) + LPT = 45 jours. Le transit time, le TAT et le « transit to » sont tous des délais, sans AMU, des temps morts. Un exemple de « *work scope* » est présenté à la figure 28.

Close

WS#	Name	Definition					TAT
1	BUR						15
2	Hub Fan Only (On Wing)	Hub					30
3	Hub Fan & Compressor	Hub	HPC				45
3	Hub Fan & Compressor	Hub	HPC	IMP			45
3	Major Refurbishment/Overhaul (Fan Hub, Compressor, Impeller, LPT)	Hub	HPC	IMP	LPT		90
11	Major Refurbishment/Overhaul (Fan Hub, Compressor, Impeller, HPT, LPT)	Hub	HPC	IMP	HPT	LPT	90
1	Hub Fan, Compressor & High Turbine	Hub	HPC	HPT			60
1	Hub Fan, Compressor, High Turbine, low turbine	Hub	HPC	HPT	LPT		60
4	Hub Fan & Impeller	Hub	IMP				45
5	Hub Fan, Impeller, HPT	Hub	IMP	HPT			50
5	Major Refurbishment/Overhaul (Fan Hub, Impeller, HPT, LPT)	Hub	IMP	HPT	LPT		90
6	Fan Hub & HPT	Hub	HPT				45
7	Fan Hub, HPT & LPT	Hub	HPT	LPT			45
8	Fan Hub & Low Pressure Turbine	Hub	LPT				45
12	Compressor	HPC					45
13	Compressor & Impeller	HPC	IMP				45
14	Major Refurbishment/Overhaul (Compressor, Impeller, HPT)	HPC	IMP	HPT			90
14	Major Refurbishment/Overhaul (Compressor, Impeller, HPT, LPT)	HPC	IMP	HPT	LPT		90
15	Compressor & Turbine	HPC	HPT				60
15	Major Refurbishment/Overhaul (Compressor, HPT, LPT)	HPC	HPT	LPT			90
16	Major Refurbishment/Overhaul (Compressor, LPT)	HPC	LPT				90
19	Impeller	IMP					45
20	Impeller & HPT	IMP	HPT				45
21	Impeller, HPT & LPT	IMP	HPT	LPT			60
22	Impeller & LPT	IMP	LPT				45
23	HPT	HPT					45
24	HPT & LPT	HPT	LPT				45

Figure 28 - Un exemple de « work scope ».

Ils existent des « campaign reviews », dans lesquels on incorpore certains SBs. Voici un exemple de « campaign reviews »:

Close

Campaign Name	Campaign Number	SB	Cost	TOTAL To Do	Completed	Remain	Percent Complete
Tie Shaft	A2000024	SB25082					
External BOV	A2002005	SB25090	\$90,000				
Phase 4	-	SB25234					
Anti Rotation	-	SB25156					
Fuel Flow Meter	A1000486	SB25252					
EEC Reprogramming	A2004010	SB25237					
L SEAL Clamping	-	SB25254					
P3 Tube Modification	-	SB25279					

Figure 29 - Un exemple de « campaign reviews »

Dans son exemple, la simulation nous dit combien de temps il reste pour les pièces, mettre des «1» aux pièces qu'on pourrait remplacer. Voici les résultats obtenus lorsque l'algorithme a roulé:

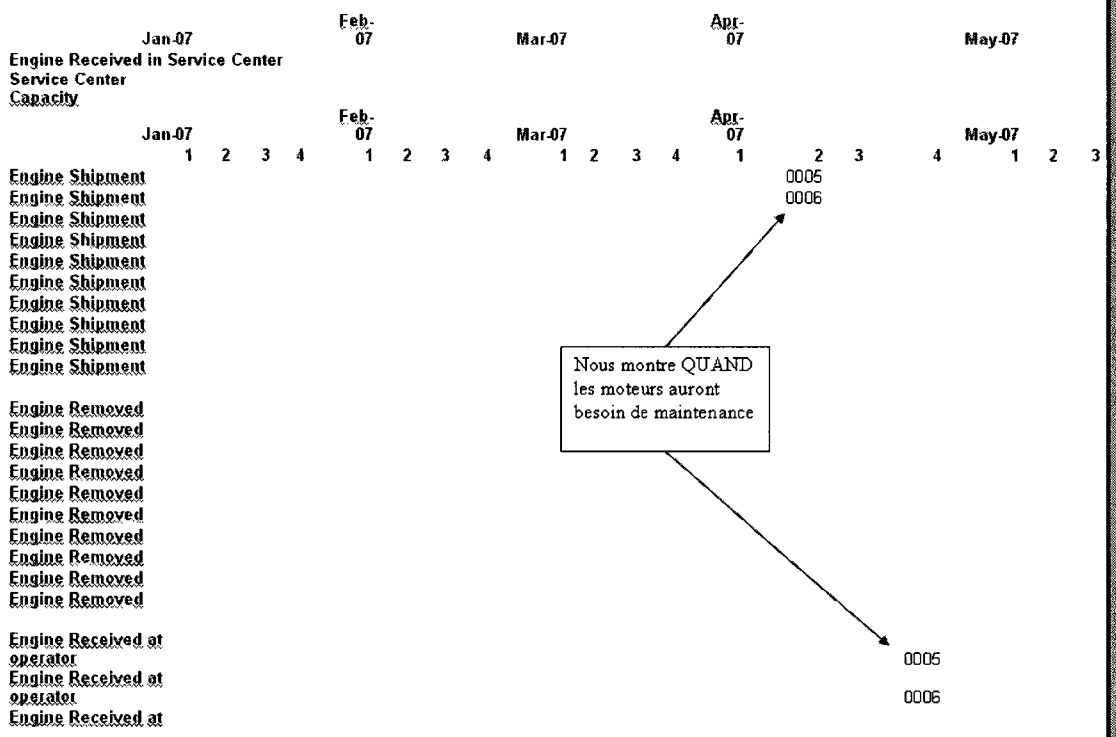
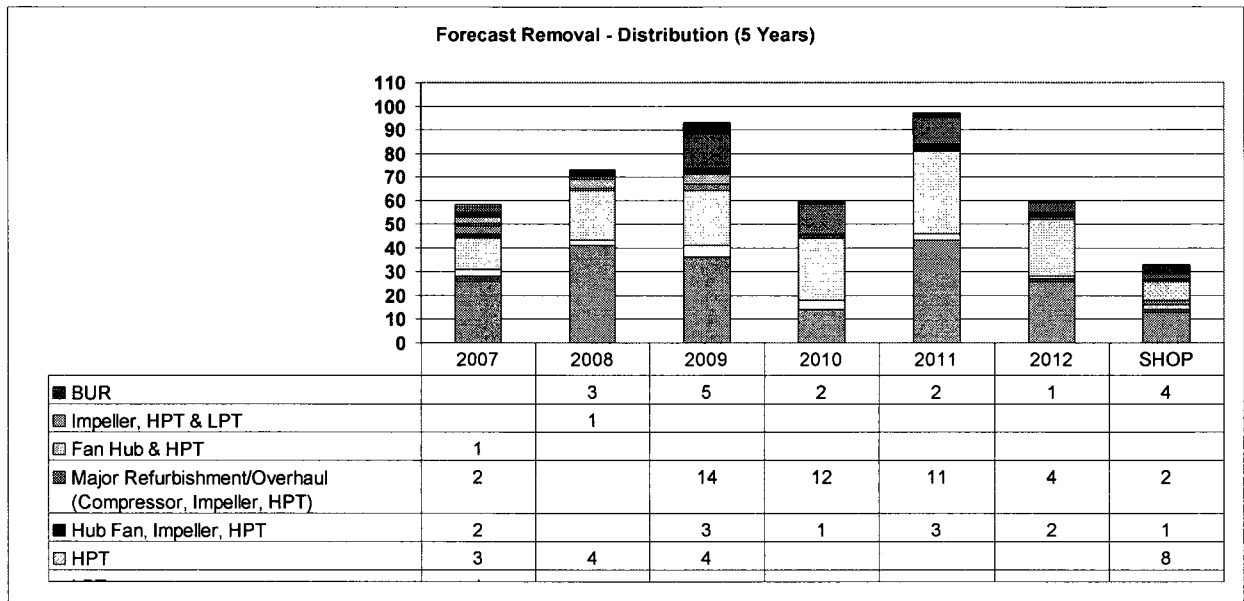


**Workscope Name**  
**Impeller & HPT**  
**Hub Fan Only (On Wing)**  
**Major Refurbishment/Overhaul (Compressor, Impeller, HPT, LPT)**  
**Major Refurbishment/Overhaul (Fan Hub, Compressor, Impeller, HPT, LPT)**  
**LPT**  
**HPT**  
**Major Refurbishment/Overhaul (Fan Hub, Impeller, HPT, LPT)**  
**Hub Fan & Compressor**  
**Hub Fan, Impeller, HPT**  
**Major Refurbishment/Overhaul (Compressor, Impeller, HPT)**  
**Fan Hub & HPT**  
**Impeller, HPT & LPT**  
**BUR**  
**Impeller**

**Figure 30- Résultats de l'algorithme sur MS Excel**



**Table 18 - Distribution des maintenances « Moteur Enlevé » sur 5 ans, par MS Excel**



**Étape B.2 – La base (Ce que l'on recommande)**

SB et MM

Regrouper les événements prévus aux moments similaires.

Tracer un cheminement commun pour tous les moteurs de ce modèle.  
 Relever les variables intemporelles (constantes) ou contraintes non horaires.

« *On condition* »

Ces éléments ont été considérés comme faisant partie de l'information intrinsèque contenue dans l'historique de données.

### Étape B.3 – L'historique (Que s'est-il passé avant?)

Combien d'événements pour tous les moteurs 306B? 771

En moyenne par moteur? 3.28

Le max? 7

Le Min? 1

L'événement le plus fréquent? BPRNC (262).

L'événement le moins fréquent? BPONS, BPRNU, BURCC, BURPC, NURAU, NURDC, NURIC, NURIS, NURNU (1).

Engine Disposition	Classification
BPONC Count	5
BPONS Count	1
BPRN+ Count	5
BPRNC Count	262
BPRNS Count	163
BPRNU Count	1
BUONC Count	2
BURAC Count	14
BURAS Count	11
BURCC Count	1
BURIC Count	7
BURIS Count	5
BURNC Count	179
BURNS Count	56
BURNU Count	3
BURPC Count	1
NURAC Count	5
NURAS Count	3
NURAU Count	1
NURDC Count	1
NURIC Count	1
NURIS Count	1
NURNC Count	30
NURNS Count	12
NURNU Count	1
Grand Count	771

Figure 31 - Décompte des types d'événements

**Table 19 - Sommaire des événements regroupés**

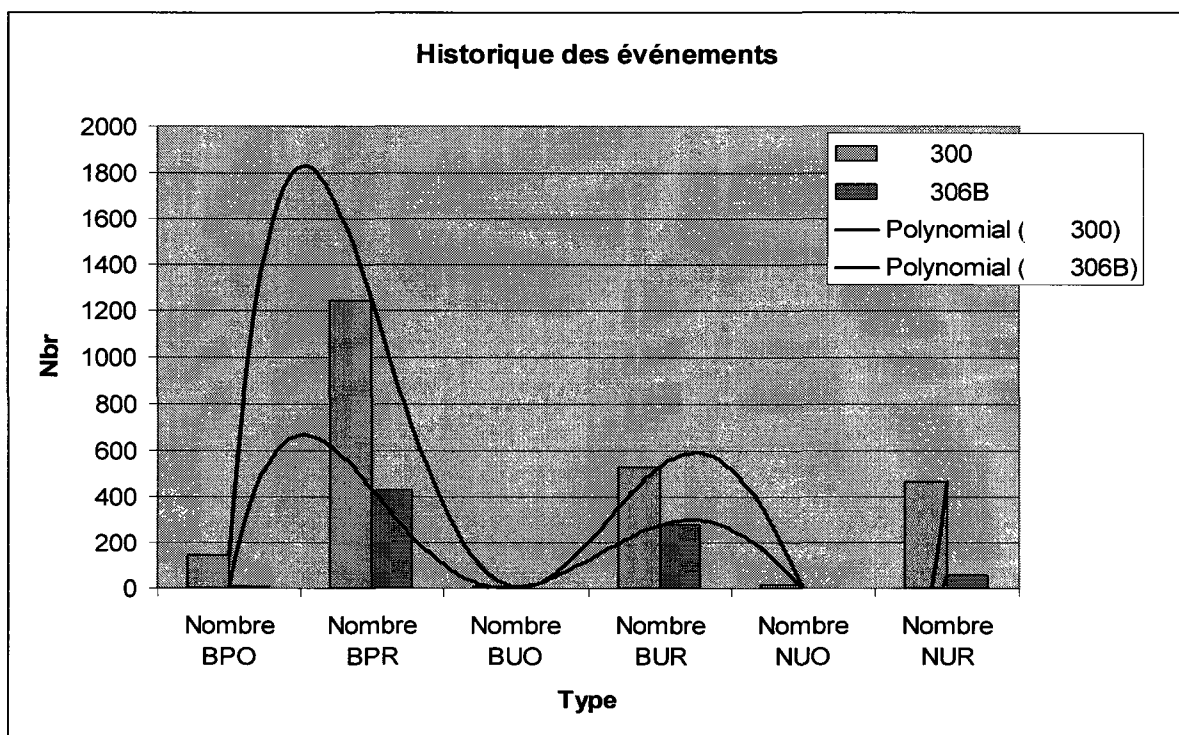
BPO	6	0.78%	Basic Planned Overhauled No Fl. eff
BPR	431	55.90%	Basic Planned Repaired
BUO	2	0.26%	Basic Unplanned Overhauled
BUR	277	35.93%	Basic Unplanned Repaired
NUR	55	7.13%	Non Basic Unplanned Repaired

Comparé à toute la flotte de modèles 300:

**Table 20 - Sommaire des événements pour toute la flotte des modèles 300**

Substantiation	Event Classification 3		
<b>Nombre BPO</b>	6.1%	146	Basic Planned Overhauled No Fl. eff
<b>Nombre BPR</b>	51.7%	1241	Basic Planned Repaired
<b>Nombre BUO</b>	0.3%	7	Basic Unplanned Overhauled
<b>Nombre BUR</b>	22%	529	Basic Unplanned Repaired
<b>Nombre NUO</b>	0.5%	11	Non Basic Unplanned Overhauled
<b>Nombre NUR</b>	19.5%	467	Non Basic Unplanned Repaired
<b><u>Nbval</u></b>		<b><u>2401</u></b>	

Graphiquement:



**Figure 32 - Historique des événements sur les modèles 300**

Peut-on regrouper aux heures similaires? Oui. On peut voir la concentration si on regarde tous les modèles 300 en général pour un BPO:

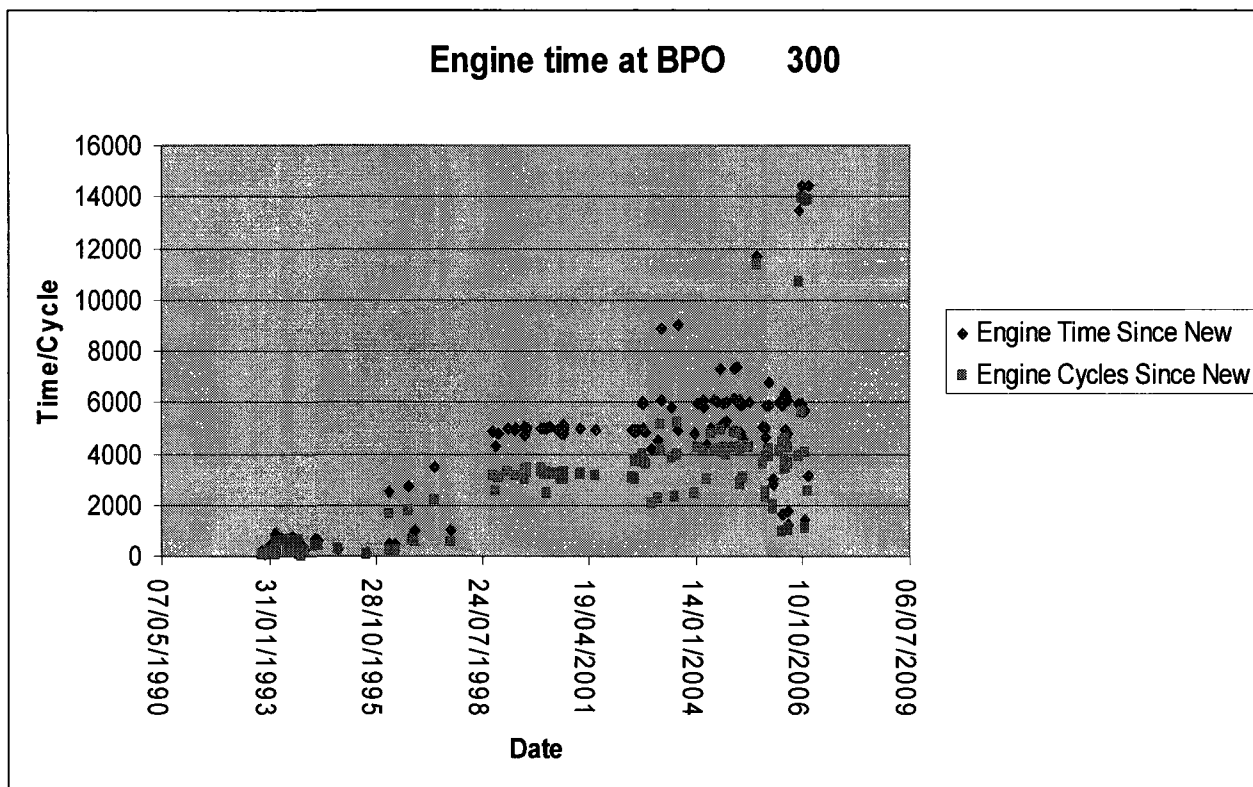


Figure 33 - Les BPOs sur les modèles 300

Lesquels sont des XX306B?

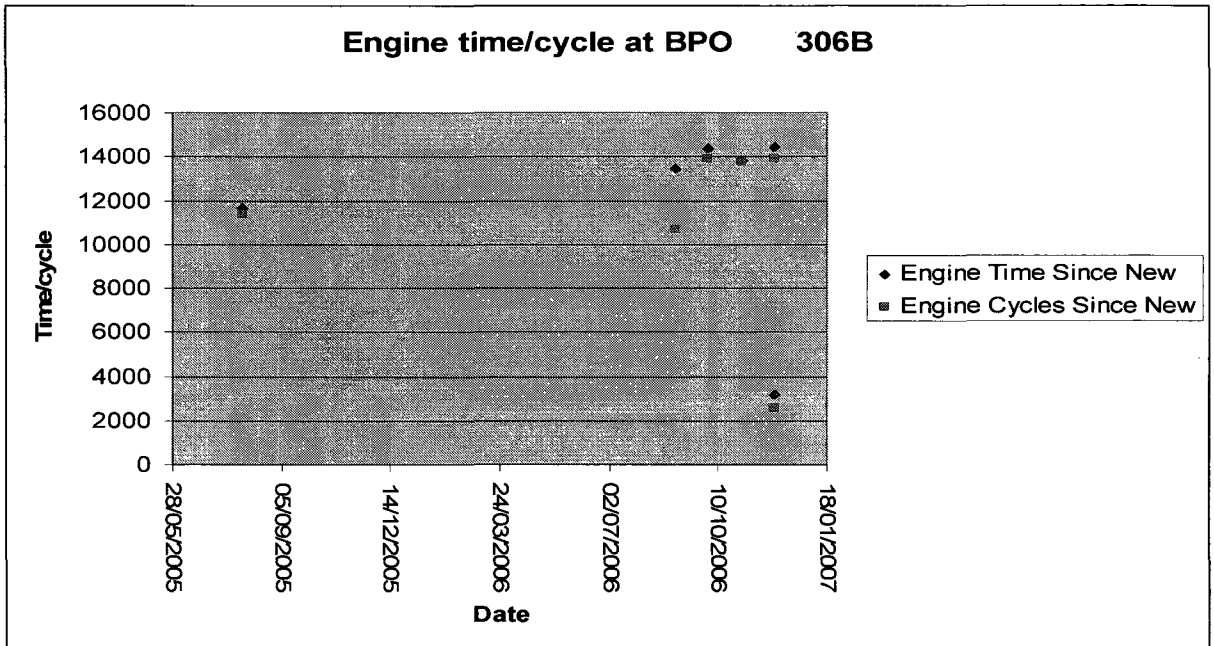


Figure 34 - Les BPOs sur les modèles 306B

A noter que sur les prochains graphiques, « Série1 » signifie « *Engine Time Since New* » et « Série2 » signifie « *Engine Cycles Since New* ». Voyons ce que nous avons pour les BPRs:

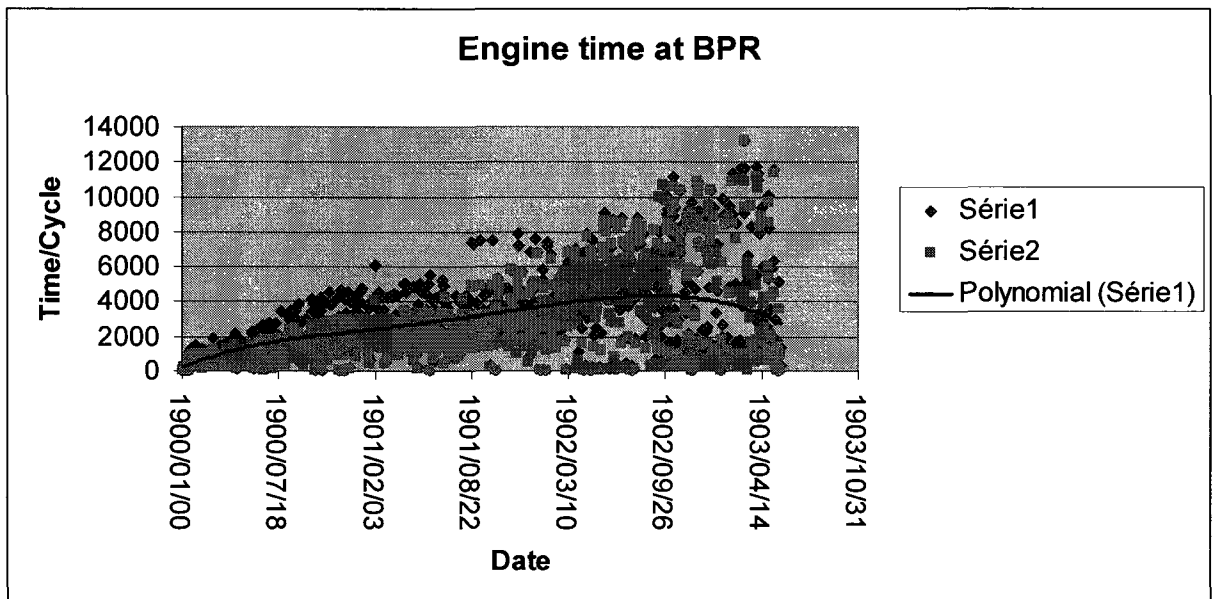


Figure 35 - Les BPRs sur les modèles 300

Lesquels sont des XX306B?

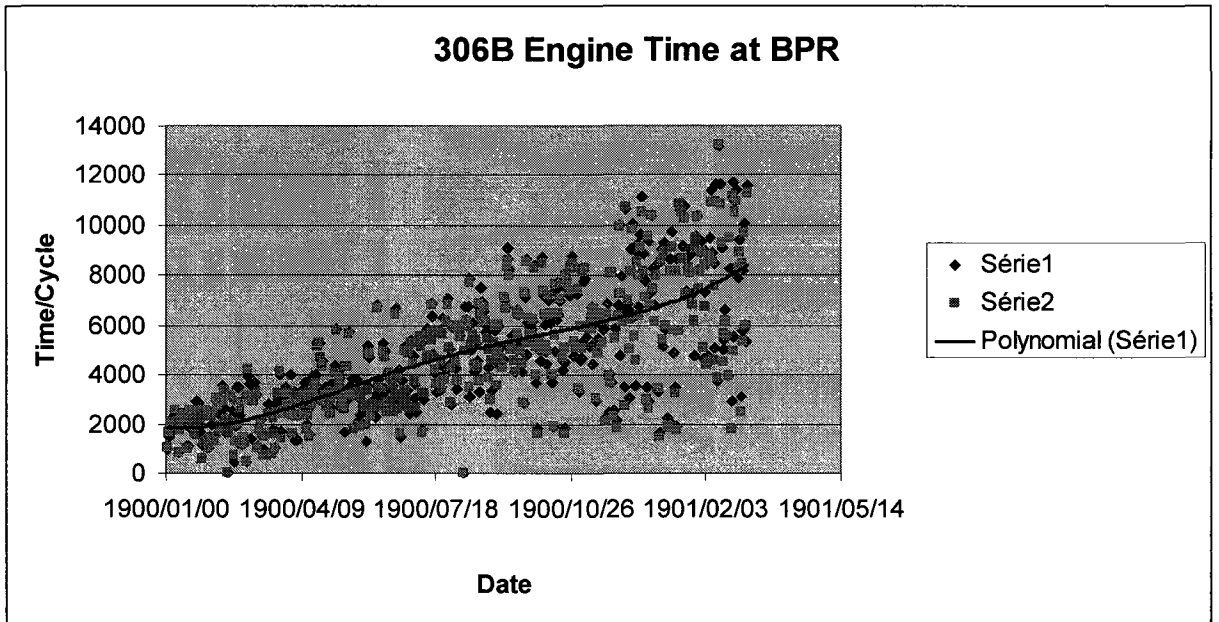


Figure 36 - Les BPRs sur les modèles 306B

Et les BUOs? Il n'y en a aucun pour les XX306B, seulement sur les XX300:

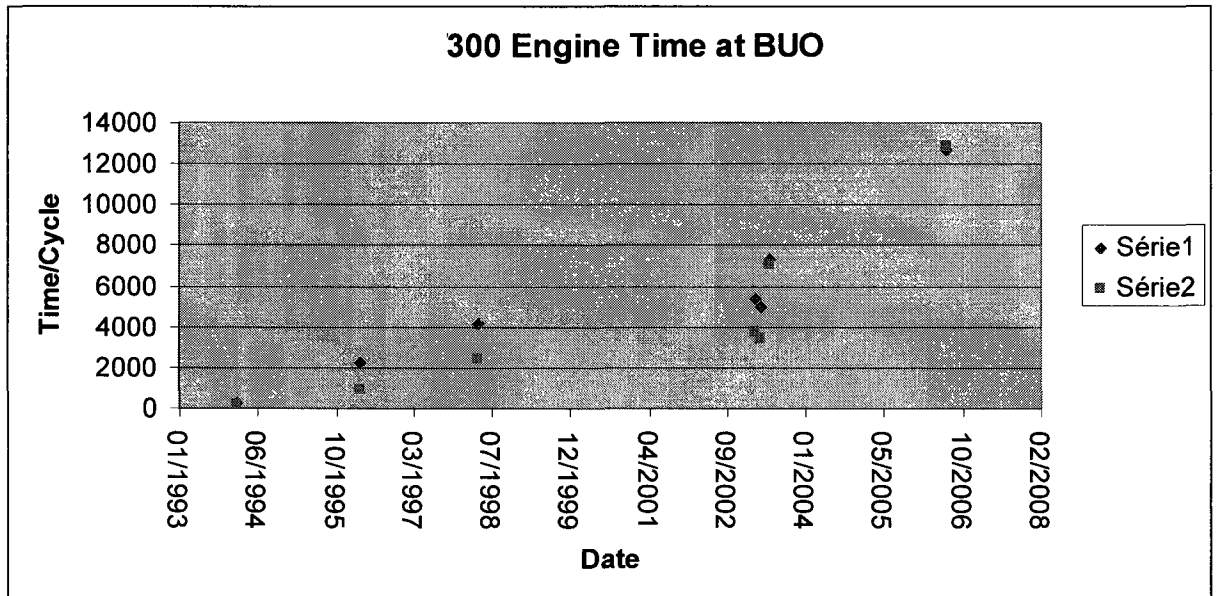


Figure 37 - Les BUOs sur les modèles 300

Les BURs:

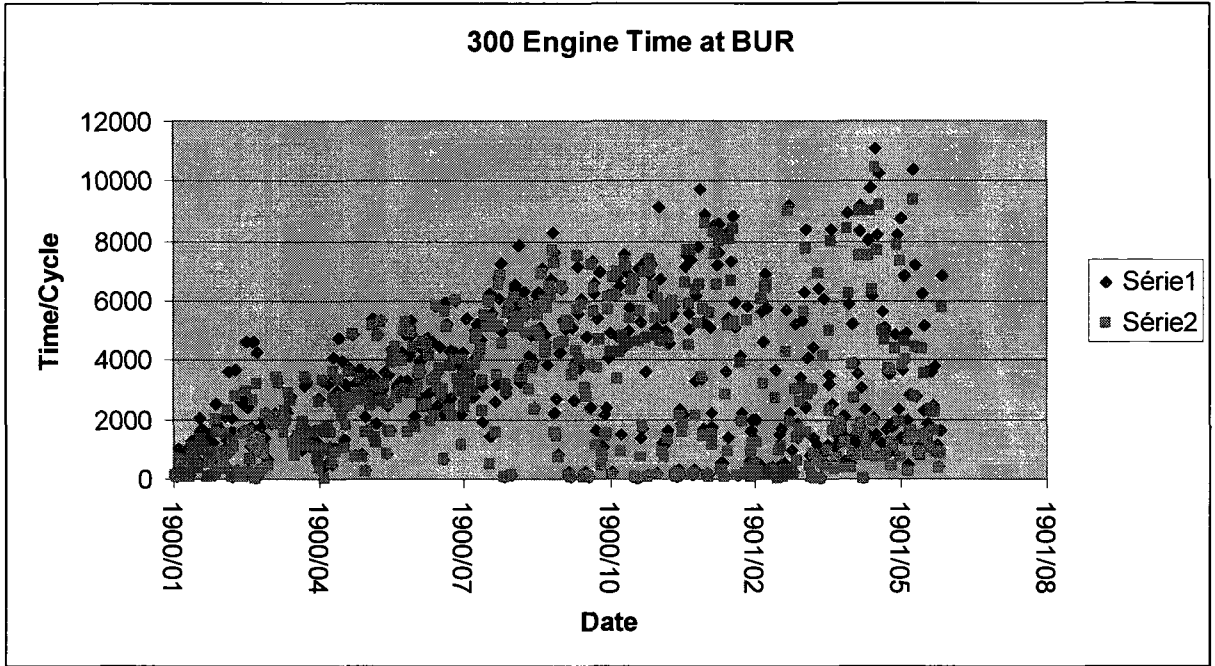


Figure 38 - Les BURs sur les modèles 300

Ceux sur les XX306B:

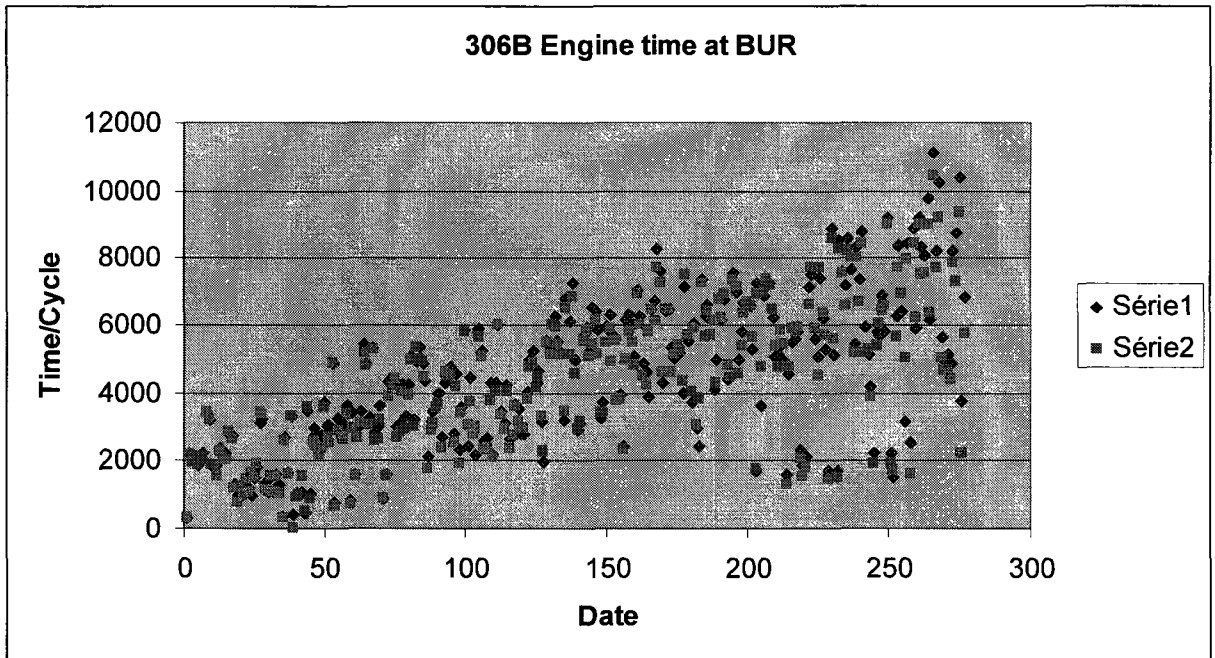


Figure 39 - Les BURs sur les modèles 306B

Les NUOs? Il n'y en a aucun pour les XX306B, seulement les 300:

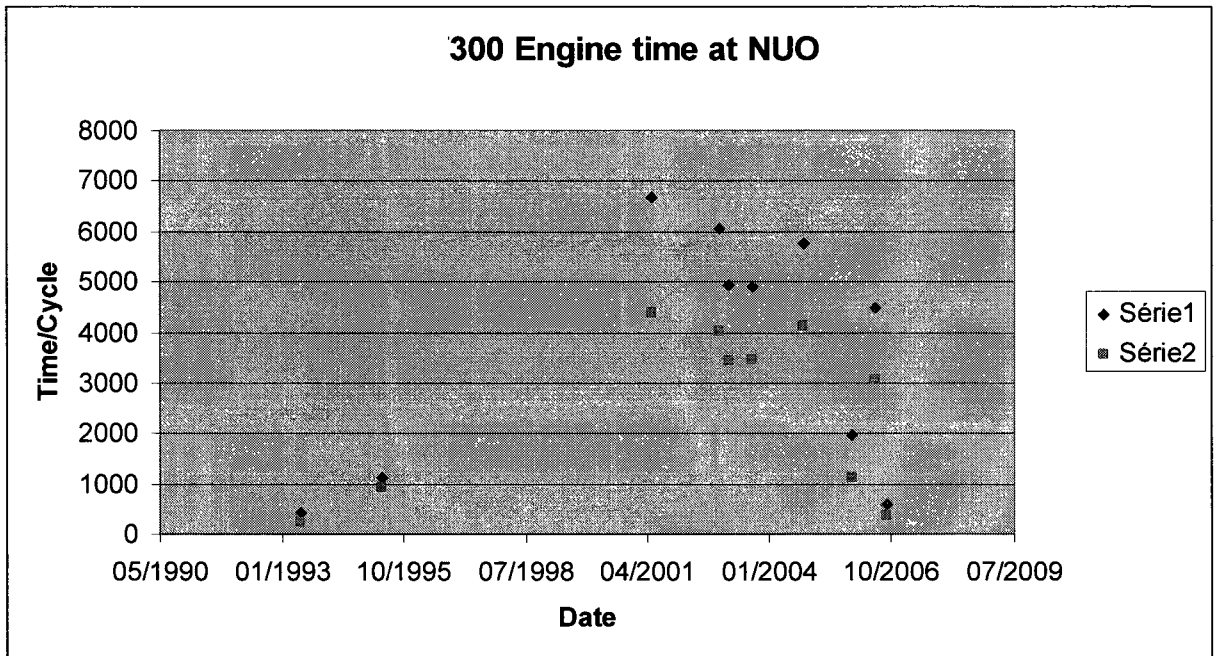


Figure 40 - Les NUOs sur les modèles 300

Finalement, les NURs:

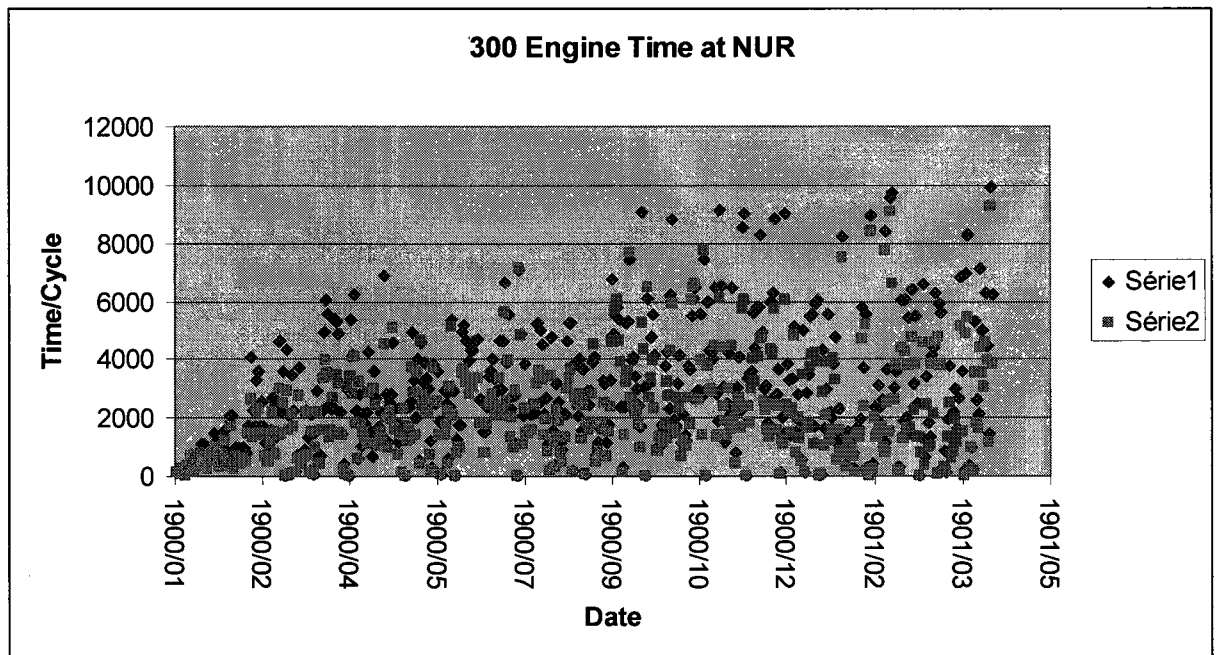


Figure 41 - Les NURs sur les modèles 300

Pour le XX306B:



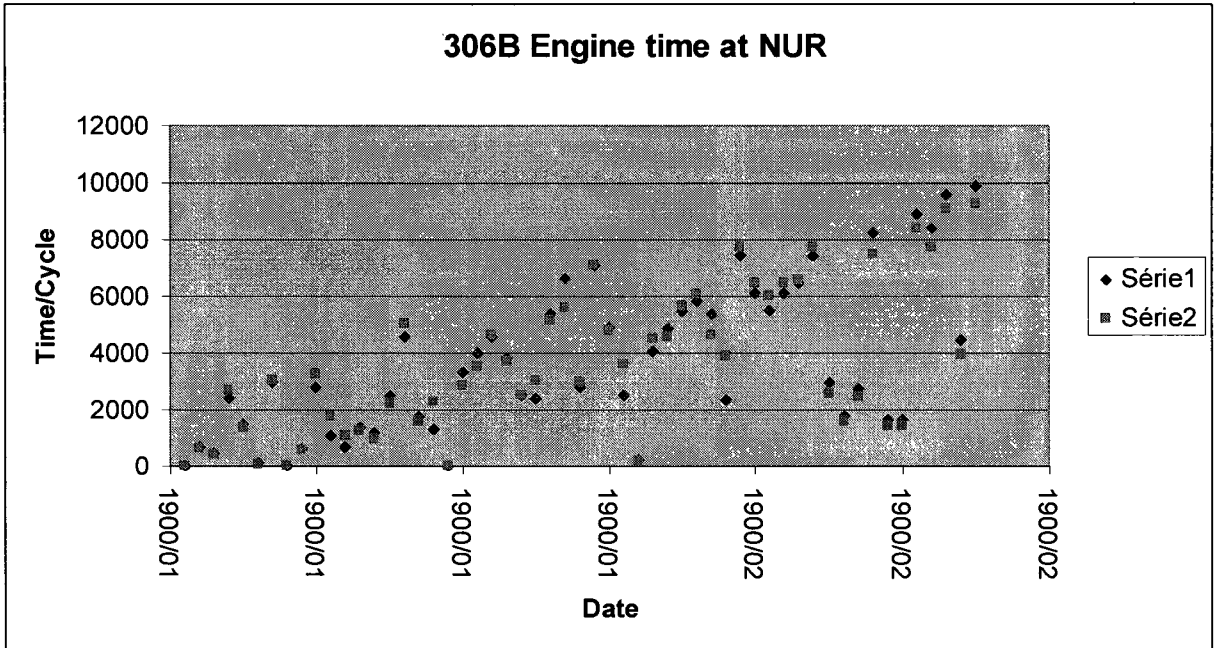


Figure 42 - Les NURs sur les modèles 306B

Si on regarde chacun des moteurs séparément, on peut regarder l'ensemble des événements le concernant:

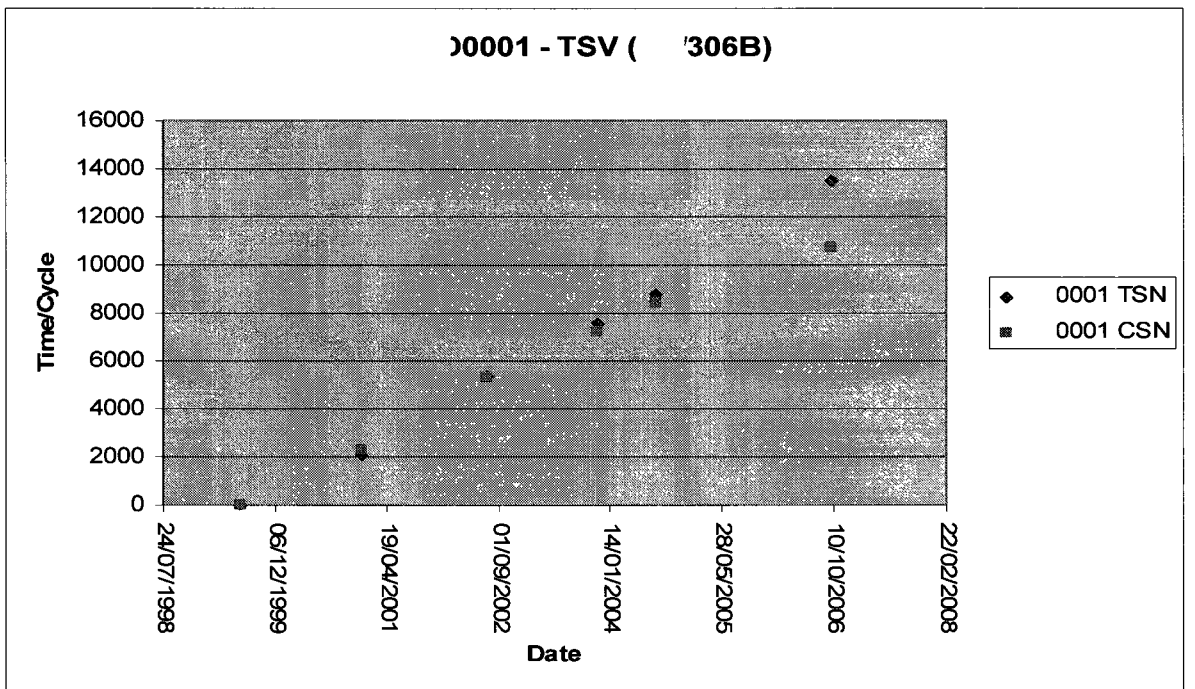


Figure 43 - Événements sur le moteur 00001 des 306B

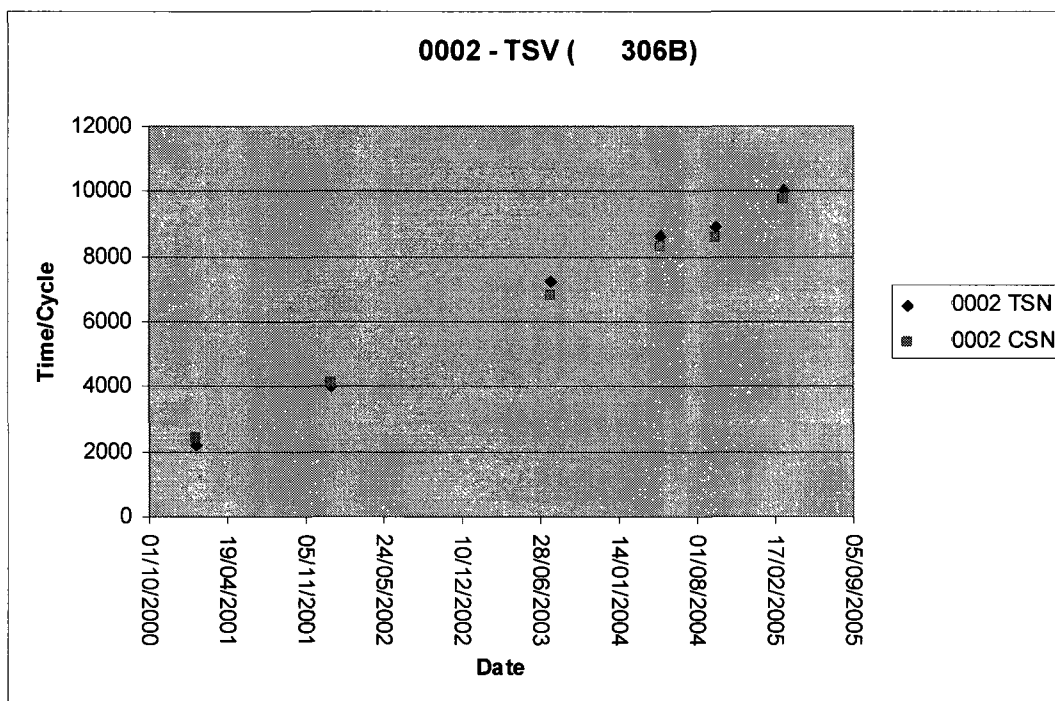


Figure 44 - Événements sur le moteur 00002 des 306B

Dans les figures 43 et 44, les deux moteurs sont montés sur le même avion, voilà pourquoi ils ont sensiblement les mêmes événements. Ces graphiques nous apprennent certaines choses: on y remarque les délais entre les événements, et leur fréquence. On peut aussi constater si l'opérateur effectue des cycles inférieurs, semblables ou supérieurs à 1 heure de vol. Dans les deux prochaines figures (45 et 46), on peut clairement constater un changement dans la durée des cycles.

Dans la figure 46, on constate que les cycles sont égaux avec les « *Effective Flying Hours – EFH* ». Ce moteur (000011) aura besoin de maintenance beaucoup moins rapidement que le 000008 de la figure 41. Plus il y a de cycles par heures, plus il y a de rallumages et cela demande le pouvoir maximum du moteur.

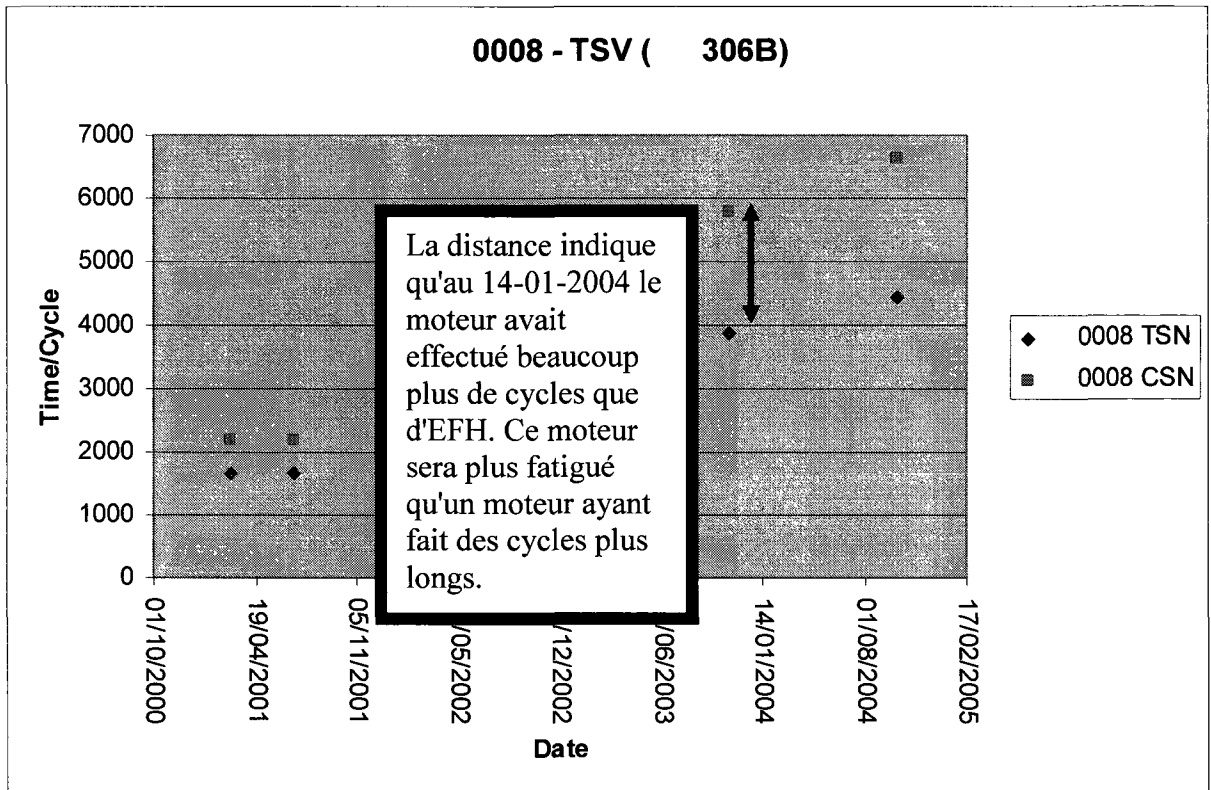


Figure 45 - le TSN comparé aux CSN pour un modèle 306B

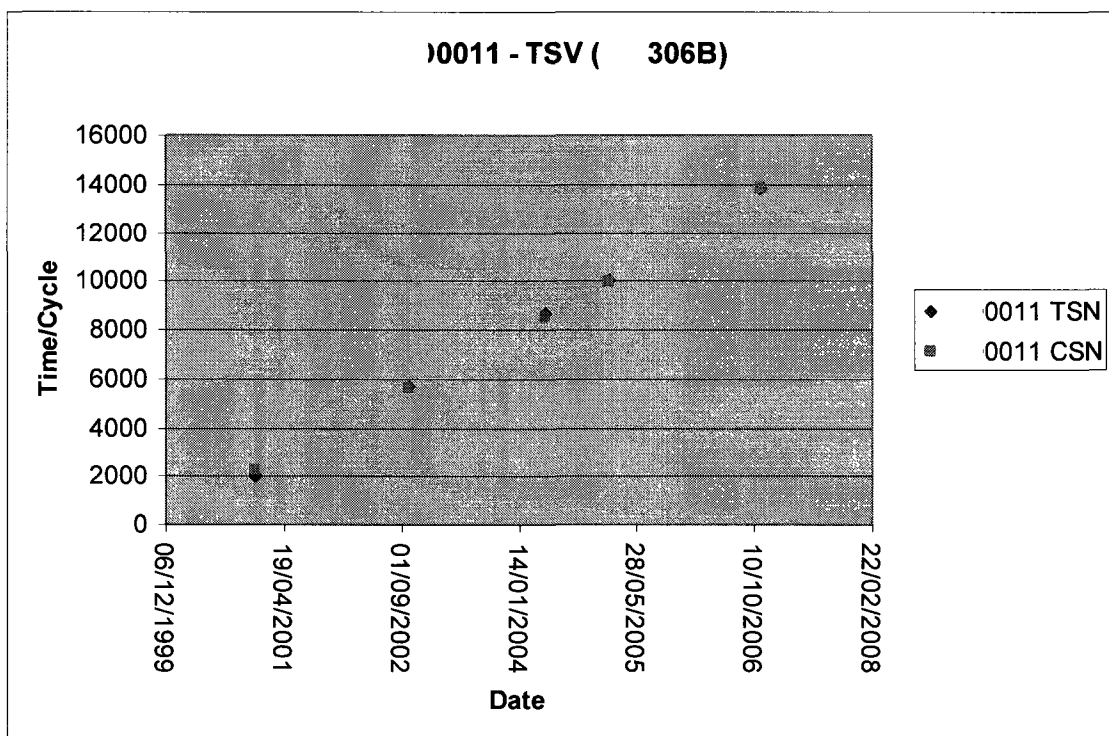


Figure 46 - le TSN comparé aux CSN pour un autre modèle 306B

Regardons un autre exemple. Dans la figure 47, la distance est encore très grande entre les temps (EFH) et les cycles. Pourtant, cette fois-ci, le moteur se fatiguera encore moins que le 000011 de la figure 46 (considéré comme étant lui-même moins fatigué que le 000008, notre premier exemple dans la figure 45). Le moteur 000012 aura besoin de maintenance beaucoup plus tardivement que les deux premiers. Il sera d'ailleurs beaucoup moins susceptible de contracter un événement imprévu.

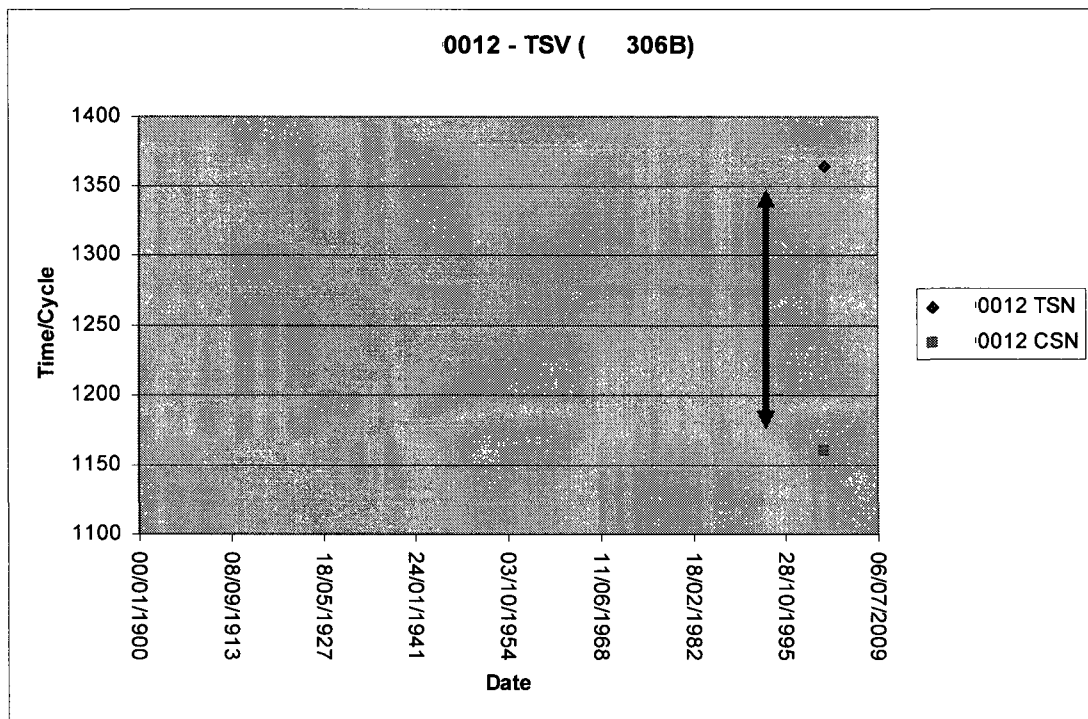
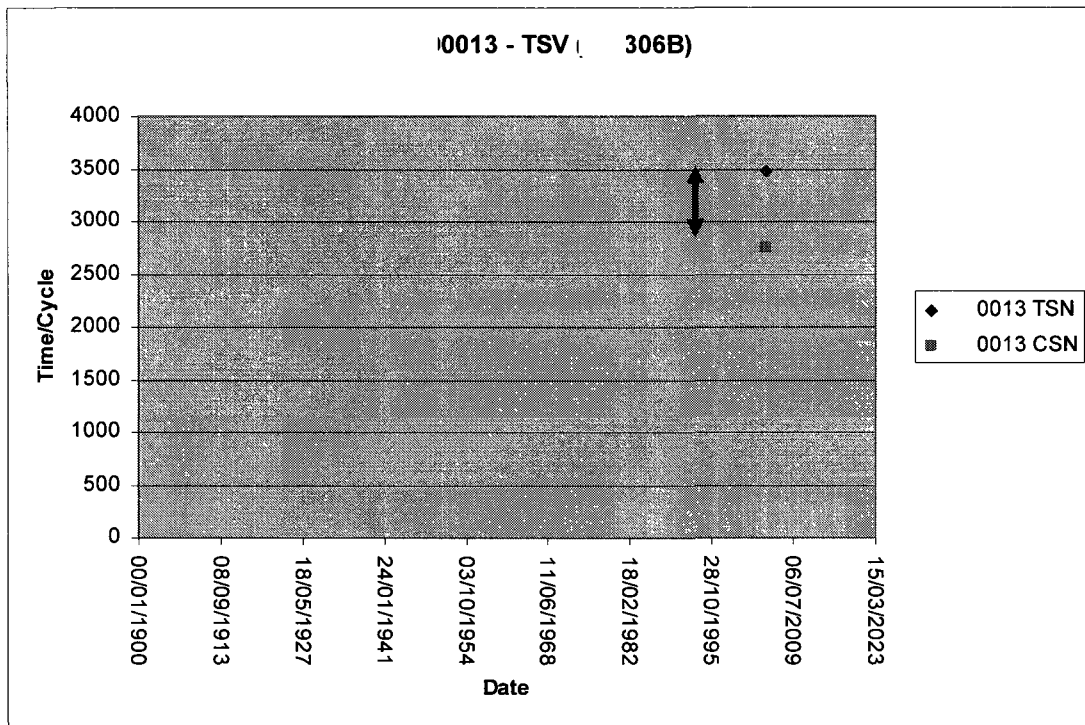
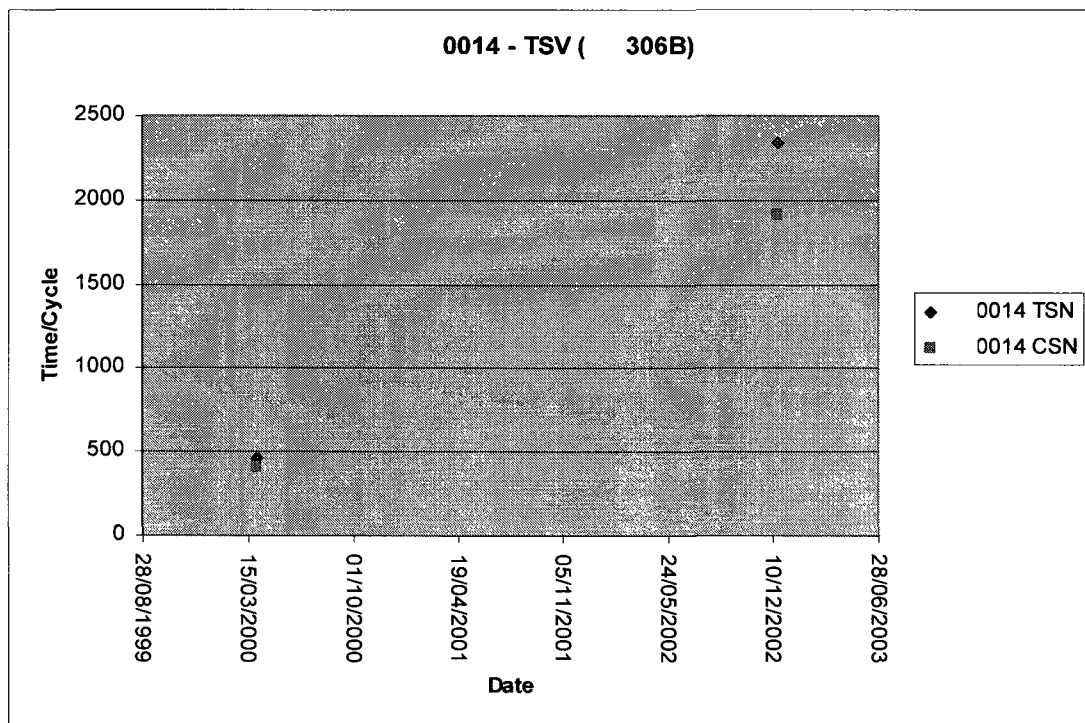


Figure 47 - Distance marquée entre le TSN et les CSN d'un 306B

Regardons encore un autre exemple avec un moteur « en forme ». Dans la figure 48, le moteur 000013 a des cycles plus courts que son nombre d'heure de vol. C'est un moteur qui n'aura pas besoin de maintenance très souvent. Cependant, puisque les temps et les cycles sont quand même rapprochés (si on compare avec le moteur 000012 de la figure 47), il se fatiguera quand même et aura probablement plus de visites en usines que le 000012, mais beaucoup moins que les deux précédents: 000008 et 000011.



**Figure 48 - Distance entre le TSN et les CSN d'un 306B**



**Figure 49 - Distance entre le TSN et les CSN d'un autre 306B**

Dans la figure 49, le moteur a des cycles plus courts que son nombre d'heures de vol, ce qui signifie qu'il effectue de longs vols. On peut également remarquer que le moteur n'a pas eu, jusqu'à présent, besoin de maintenance. Malheureusement, le graphique ne remonte pas jusqu'à nos jours et tout porte à croire qu'il n'y ait pas encore eu d'événement. Les figures 50 et 51 nous montrent deux extrapolations possibles où les paramètres suivant sont considérés:

L'événement aurait du avoir lieu, mais il ne s'est pas manifesté. Le moteur peut « extensionner » ses visites puisqu'il fait des voyages plus longs. Ainsi, la majorité du temps, il flotte en l'air et le moteur utilise très peu sa puissance maximale.

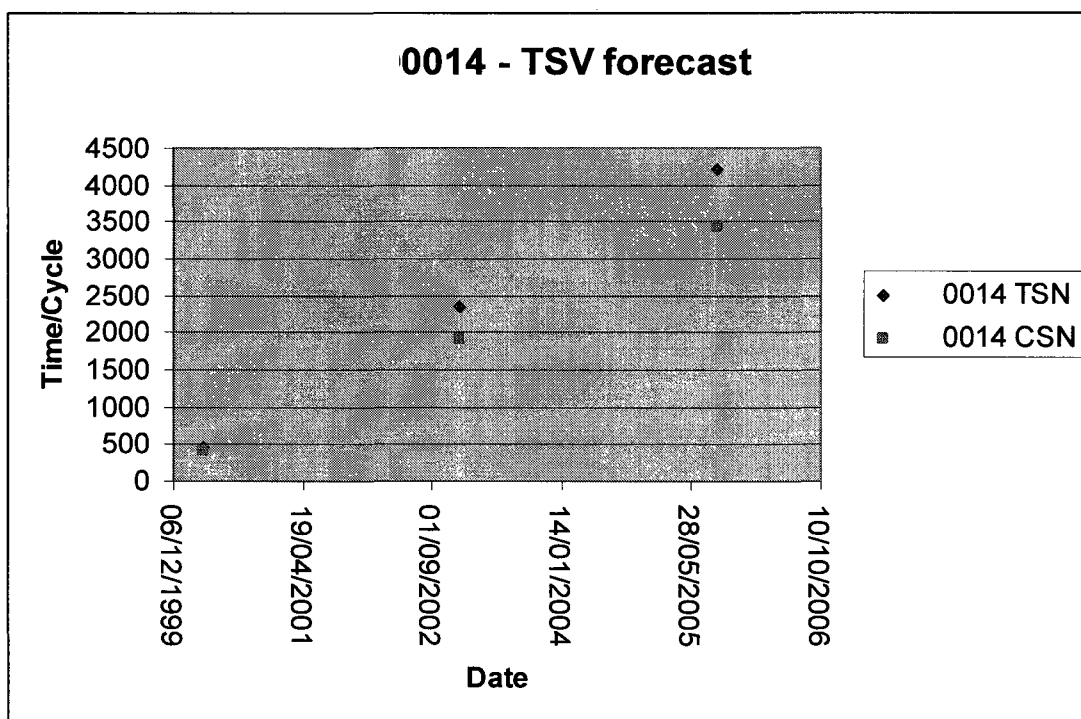


Figure 50 - Prévission d'événement possible sur le moteur 000014

Enfin, il serait surprenant d'avoir un événement aussi prochainement. On devrait plutôt assister à un prochain événement comme le présente le graphique de la figure 51.

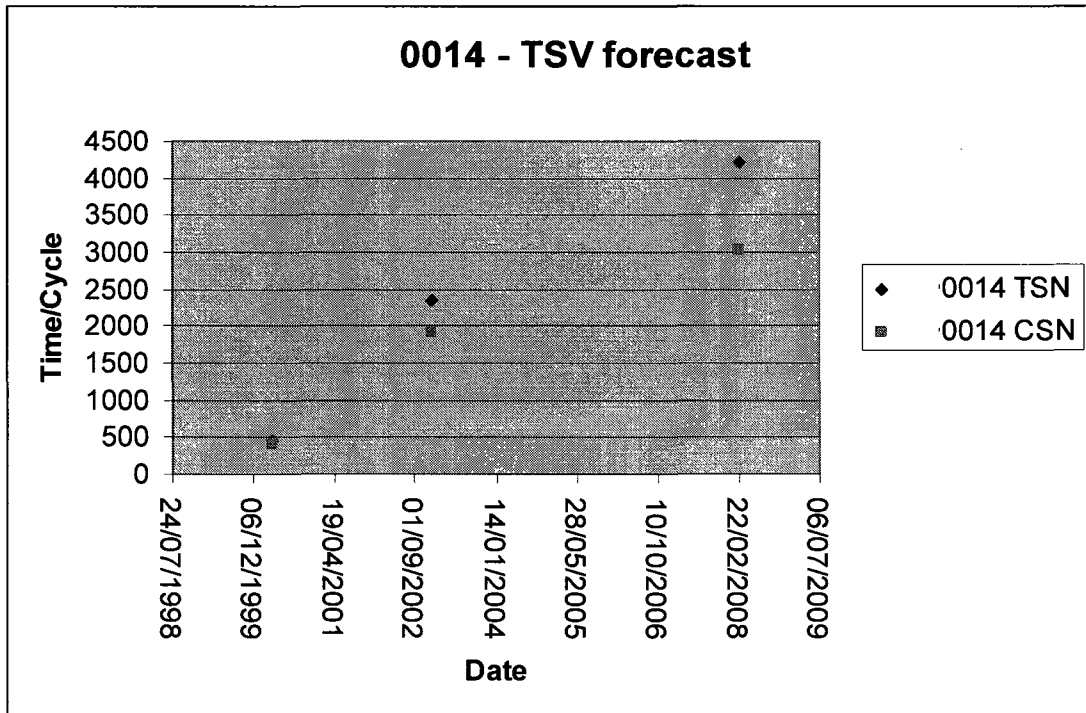
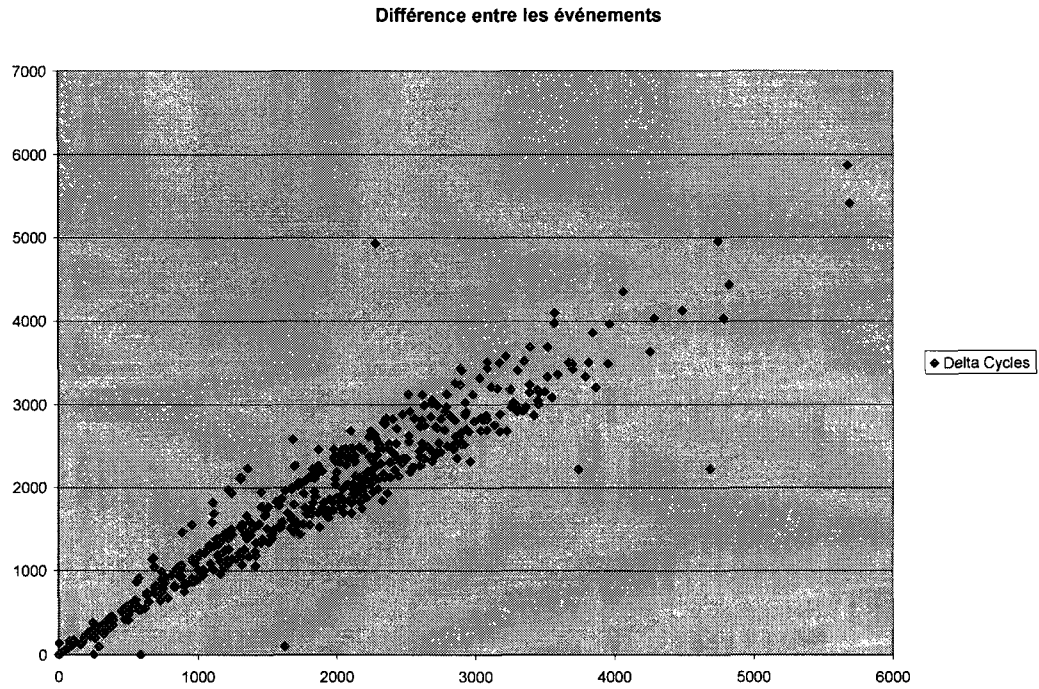


Figure 51 - Pr vision d' v nement possible sur le moteur 000014 #2

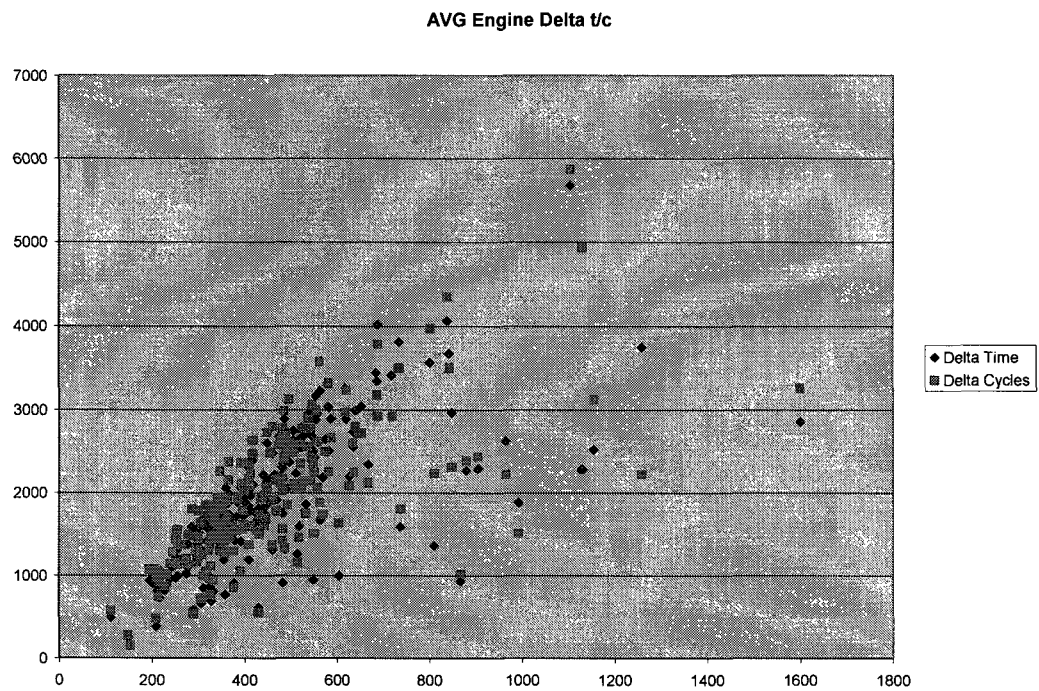
En effet, sur ce graphique,   la figure 51, si la pr vision est exacte, la maintenance devrait avoir lieu beaucoup plus tard. Voil  donc pourquoi nous n'aurions pas d j  enregistr  d' v nements encore dans nos syst mes. Si on compare le moteur 000014   d'autres moteurs comme le 000001 et le 000002, ces derniers ont des temps et des cycles  gaux, ils ont aussi visiblement plus d' v nements   leurs actifs. Peut-on d duire une tendance dans tout cela?

On peut bien regarder tous les moteurs un   un, mais cette m thode serait longue et fastidieuse. Nous nous sommes plut t pench s sur le lien qu'il pouvait y avoir entre l' cart de trois facteurs pour chaque  v nement. Je parle ici des temps, des cycles et des jours. En calculant l' cart entre chaque  v nement, nous avons regard  si les  carts de jours pouvaient  tre corr l s avec les  carts de temps et cycles. La distance entre les  v nements peut s'afficher ainsi, si on regarde,   la figure 52, les temps par rapport aux cycles. La corr lation est effectivement tr s forte, juste   voir la forme de la distribution ( $R= 0.97$ ).



**Figure 52 - Les temps par rapport aux cycles**

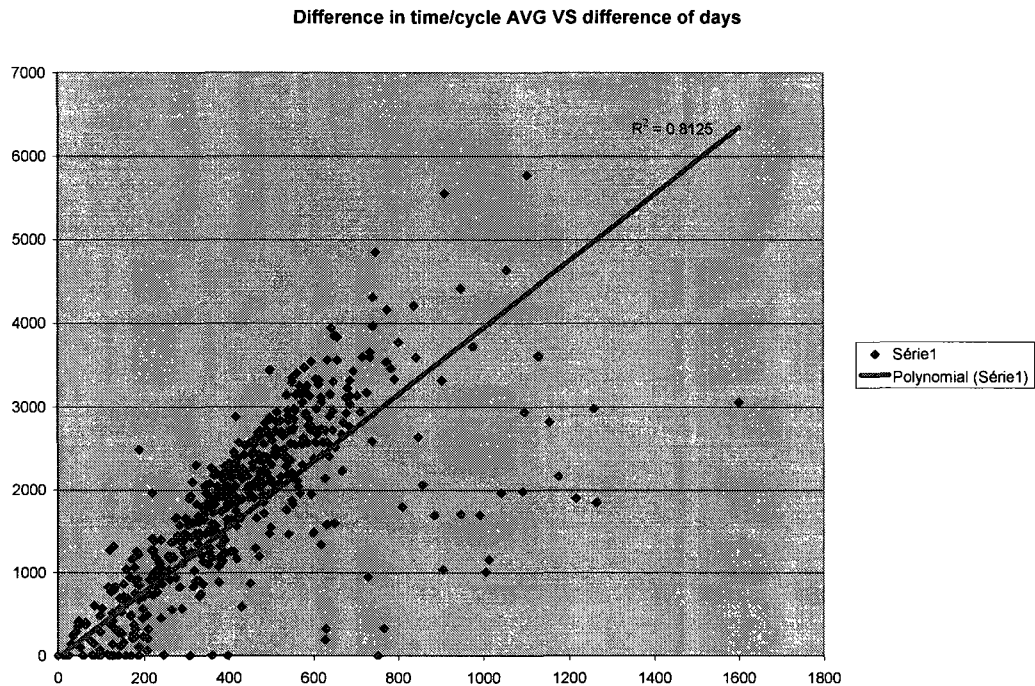
A la figure 53, La distribution des temps et cycles est tout aussi corrélée:



**Figure 53 - Distribution des temps et cycles**

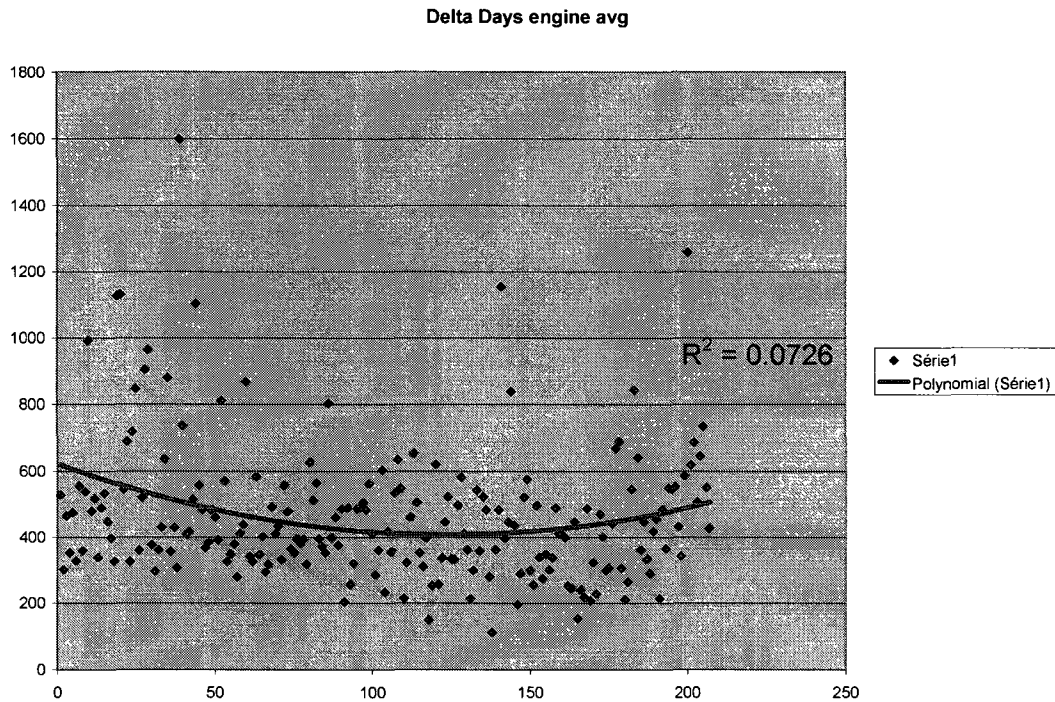


Si on regarde l'ensemble des différences entre chaque événement pour les facteurs de jours, de temps et de cycles, on obtient ceci:



**Figure 54 - différences entre chaque événement en nombre de jours et de temps/cycles**

La corrélation passe de 0.97 à 0.81 dès que l'on y ajout le facteur temps. Pourquoi cela? Parce que selon le temps et selon aussi quelques autres facteurs, dont l'environnement (qui n'est pas encore considérée dans la corrélation, mais qui en fait indirectement partie, radicalement liée avec les jours) et la macro-économie du moment (qui joue surtout sur l'usage et l'usure des moteurs), il existe plusieurs profils d'utilisation (souvent lié à la mission, dont, directement aux cycles).Voilà aussi pourquoi il n'y a pas vraiment de corrélation entre les différences des jours et les événements. La figure 55 nous le démontre.



**Figure 55 - Corrélation entre les différences des jours et les événements**

Cependant, si on pouvait repérer des patterns types pour des séquences dans le temps, alors là on aurait non plus une seule corrélation, mais des sous-ensembles corrélés. Peut-on déduire un cycle dans tout cela?

Comme si ce si beau casse-tête pouvait en fait se résumer à ces quelques variables, suffirait-il simplement de savoir ensuite comment les calibrer? En utilisant un algorithme de « *Clustering* », il est ainsi possible de calibrer les variables en fonction des jours et détecter des classes (ou groupes) de moteur appartenant à certains « cycles » dans le temps, selon l'utilisation que l'on fait de l'appareil.

On peut également regarder les taux d'événements imprévus au cours des derniers mois pour le modèle XX306B, appareils militaires inclus, en date du 1<sup>er</sup> octobre 2007:

**Table 21 - Taux publiés de BUR et IFSD sur un 306B**

Publication Date	Series	Rate
1-Oct-05	Published BUR Rate	0.151
1-Oct-05	Published Basic IFSD Rate	0
1-Nov-05	Published BUR Rate	0.149
1-Nov-05	Published Basic IFSD Rate	0
1-Dec-05	Published BUR Rate	0.126
1-Dec-05	Published Basic IFSD Rate	0
1-Jan-06	Published BUR Rate	0.064
1-Jan-06	Published Basic IFSD Rate	0
1-Feb-06	Published BUR Rate	0.052
1-Feb-06	Published Basic IFSD Rate	0
1-Mar-06	Published BUR Rate	0.039
1-Mar-06	Published Basic IFSD Rate	0
1-Apr-06	Published BUR Rate	0.029
1-Apr-06	Published Basic IFSD Rate	0
1-May-06	Published BUR Rate	0.035
1-May-06	Published Basic IFSD Rate	0
1-May-06	Published Non Basic IFSD Rate	0
1-Jun-06	Published BUR Rate	0.044
1-Jun-06	Published Basic IFSD Rate	0
1-Jun-06	Published Non Basic IFSD Rate	0
1-Jul-06	Published BUR Rate	0.029
1-Jul-06	Published Basic IFSD Rate	0
1-Jul-06	Published Non Basic IFSD Rate	0
1-Aug-06	Published BUR Rate	0.036
1-Aug-06	Published Basic IFSD Rate	0
1-Aug-06	Published Non Basic IFSD Rate	0
1-Sep-06	Published BUR Rate	0.037
1-Sep-06	Published Basic IFSD Rate	0
1-Sep-06	Published Non Basic IFSD Rate	0
1-Oct-06	Published BUR Rate	0.044
1-Oct-06	Published Basic IFSD Rate	0.007
1-Oct-06	Published Non Basic IFSD Rate	0
1-Nov-06	Published BUR Rate	0.044
1-Nov-06	Published Basic IFSD Rate	0.007
1-Nov-06	Published Non Basic IFSD Rate	0
1-Dec-06	Published BUR Rate	0.059
1-Dec-06	Published Basic IFSD Rate	0.007
1-Dec-06	Published Non Basic IFSD Rate	0
1-Jan-07	Published BUR Rate	0.089
1-Jan-07	Published Basic IFSD Rate	0.015
1-Jan-07	Published Non Basic IFSD Rate	0
1-Feb-07	Published BUR Rate	0.086
1-Feb-07	Published Basic IFSD Rate	0.014
1-Feb-07	Published Non Basic IFSD Rate	0
1-Mar-07	Published BUR Rate	0.092
1-Mar-07	Published Basic IFSD Rate	0.014

Publication Date	Series	Rate
1-Mar-07	Published Non Basic IFSD Rate	0
1-Apr-07	Published BUR Rate	0.093
1-Apr-07	Published Basic IFSD Rate	0.021
1-Apr-07	Published Non Basic IFSD Rate	0
1-May-07	Published BUR Rate	0.078
1-May-07	Published Basic IFSD Rate	0.021
1-May-07	Published Non Basic IFSD Rate	0.007
1-Jun-07	Published BUR Rate	0.056
1-Jun-07	Published Basic IFSD Rate	0.014
1-Jun-07	Published Non Basic IFSD Rate	0.007
1-Jul-07	Published BUR Rate	0.028
1-Jul-07	Published Basic IFSD Rate	0.007
1-Jul-07	Published Non Basic IFSD Rate	0.007
1-Aug-07	Published BUR Rate	0.014
1-Aug-07	Published Basic IFSD Rate	0.007
1-Aug-07	Published Non Basic IFSD Rate	0.007
1-Sep-07	Published BUR Rate	0.007
1-Sep-07	Published Basic IFSD Rate	0.007
1-Sep-07	Published Non Basic IFSD Rate	0.007
1-Oct-07	Published BUR Rate	0.043
1-Oct-07	Published Basic IFSD Rate	0
1-Oct-07	Published Non Basic IFSD Rate	0.007

### ***Étape B.4 – Le possible (Que peut-il arriver?)***

Cette section figure parmi les informations contenues intrinsèquement dans l'historique. Tous les types d'événements rencontrés au cours des années dressent un portrait suffisamment complet des possibilités.

### ***Étape B.5 – La Visio économique (Un lien Macro?)***

Nous allons regarder comment les facteurs économiques sont corrélés avec le nombre d'événements survenus aux cours des dix dernières années. Prenons le prix du baril de pétrole brut en dollars américains « *Crude Oil Prices – COP* » et Le nombre de mises en service de nouveaux moteurs. Est-ce que ce modèle, à la figure 56, présente une corrélation?

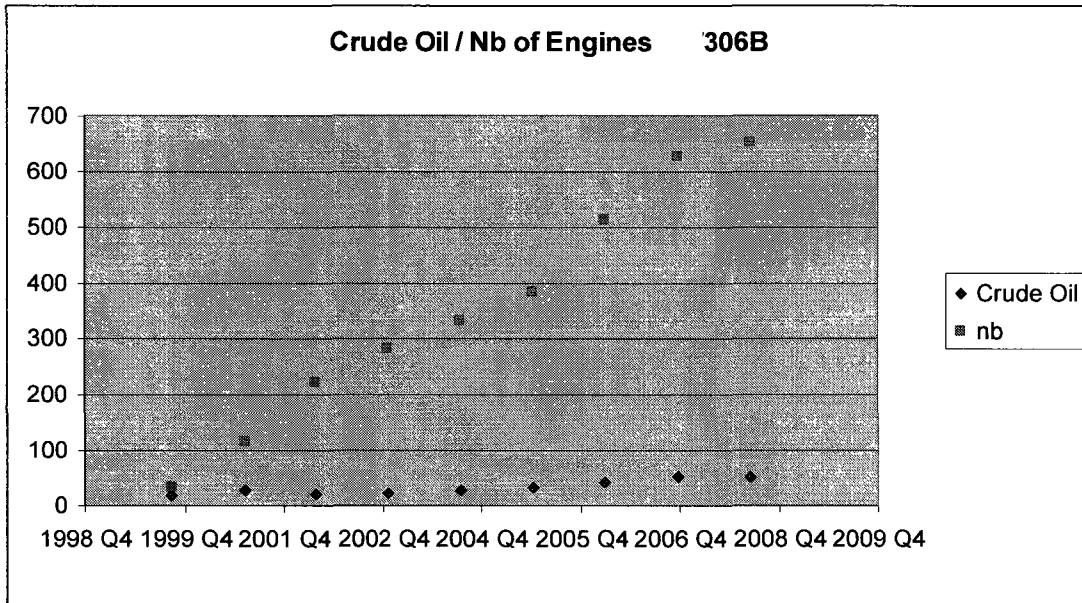


Figure 56 - « Crude Oil » VS Nombre de moteurs mis en circulation

Le nombre de mises en service ne semble pas diminuer lorsque le prix du pétrole monte. Il faudrait probablement une très forte hausse prolongée pour que cela change. Regardons plutôt le nombre d'heures de vol versus le COP, à la figure 57.

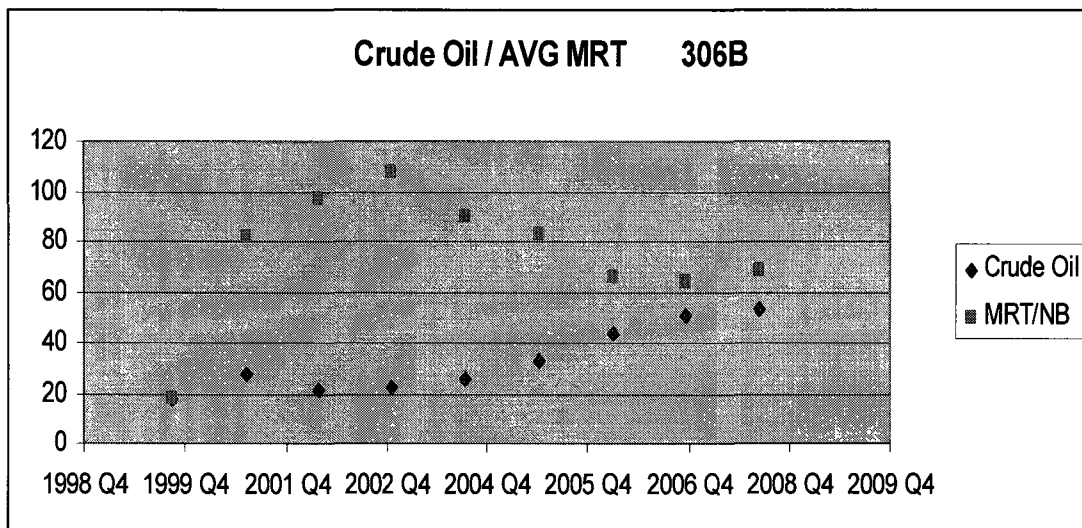


Figure 57 - « Crude Oil » VS Usage mensuel en nombre d'heures de vol

En effet, lorsque le prix du pétrole augmente, ce n'est pas le nombre de mises en services qui est affecté, mais plutôt le nombre d'heures de vol. Si on regarde maintenant les profits corporatifs aux États-Unis (« US Corporate Profits » - USCP) par rapport au nombre de mises en services, cela s'améliore. Lorsque les profits corporatifs augmentent, les ventes d'avions régionaux augmentent également.

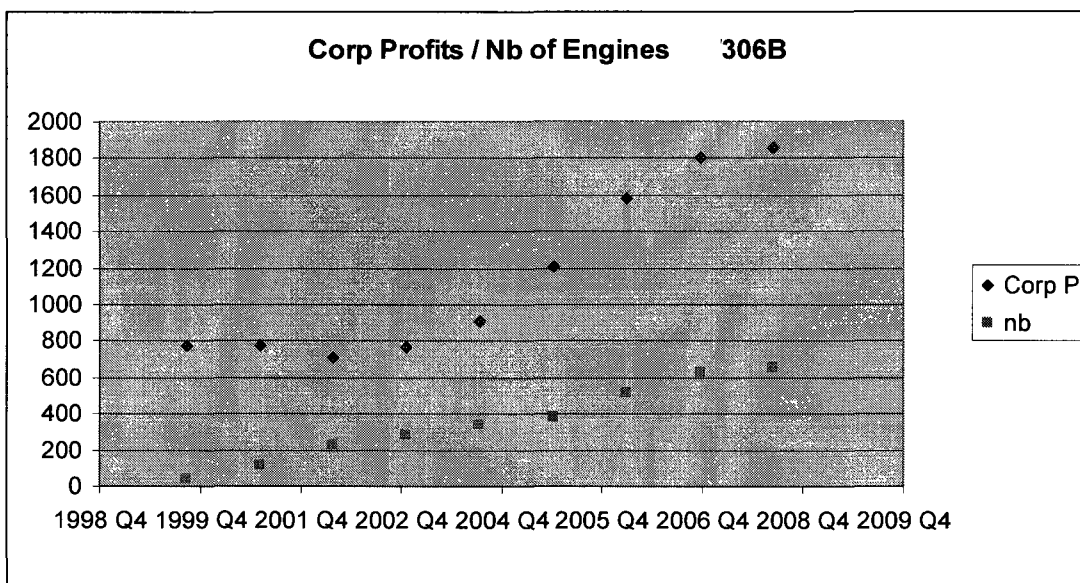


Figure 58 - Le nombre de mises en service VS les USCP

Dans la figure 59, le nombre d'heures de vol ne semble pas vraiment affecté par la hausse des USCP. On constate que de 2001 à 2002, bien que les profits corporatifs eurent diminués, le nombre d'heure de vol à continué de croître.

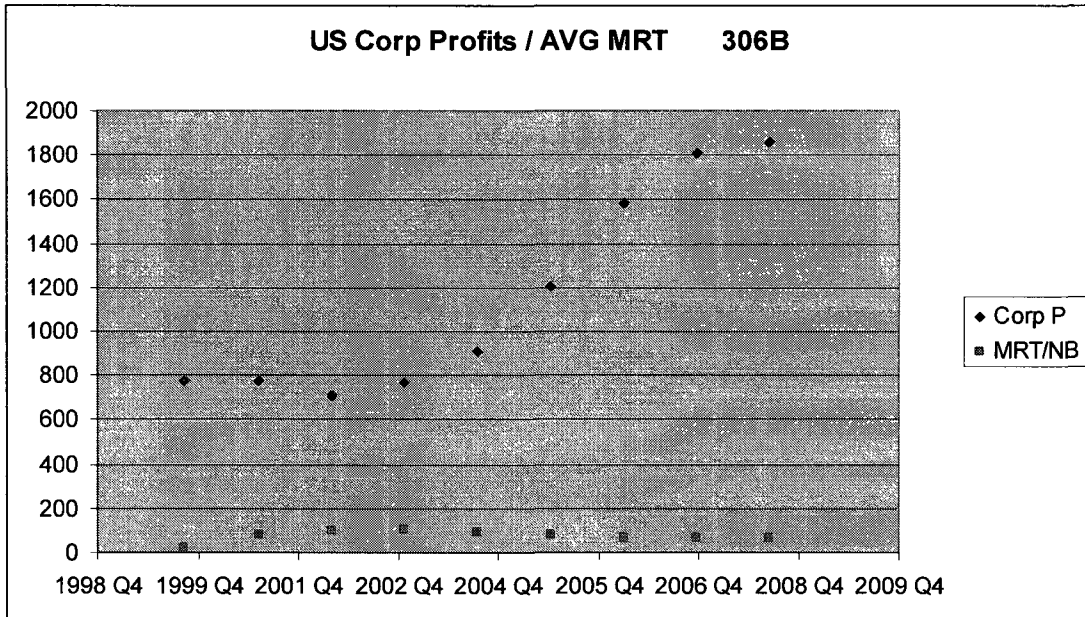


Figure 59 - Usage mensuel en nombre d'heures de vol VS USCP

Nous avons également évalué le produit brut mondial « *World's Gross Domestic Product – WGDP* » par rapport au nombre de moteurs mis en service, à la figure 60, puis par rapport à l'usage mensuelle en nombre d'heures de vol, à la figure 61.

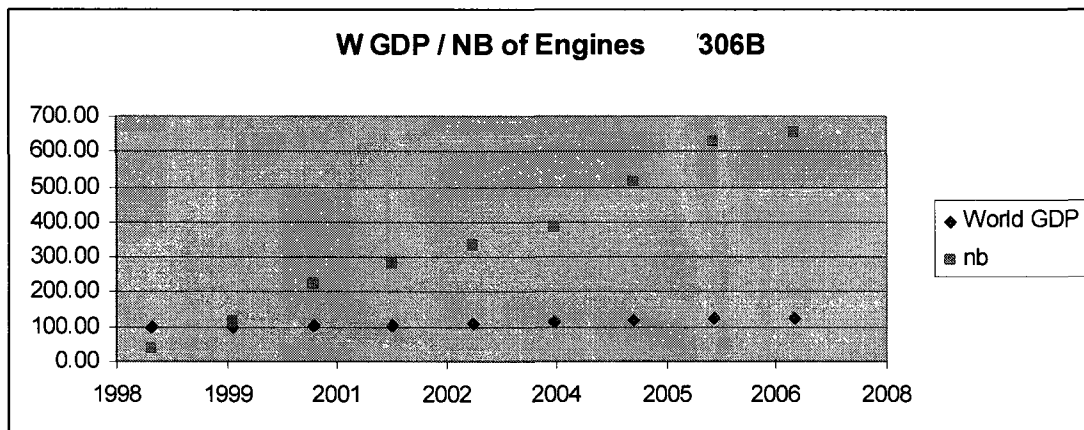


Figure 60 - Le nombre de mises en service VS WGDP

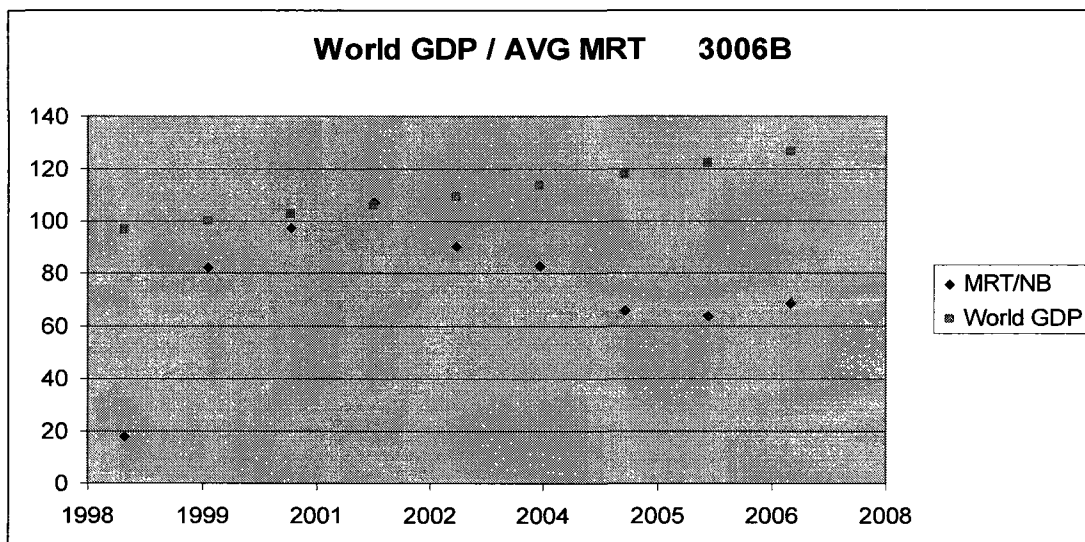
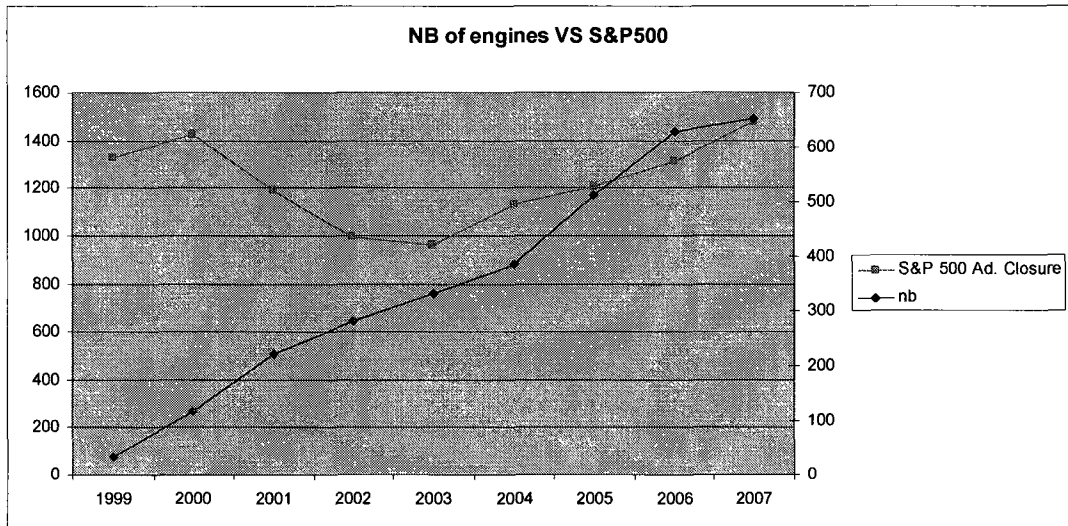


Figure 61 - L'usage mensuel en nombre d'heures de vol VS WGD

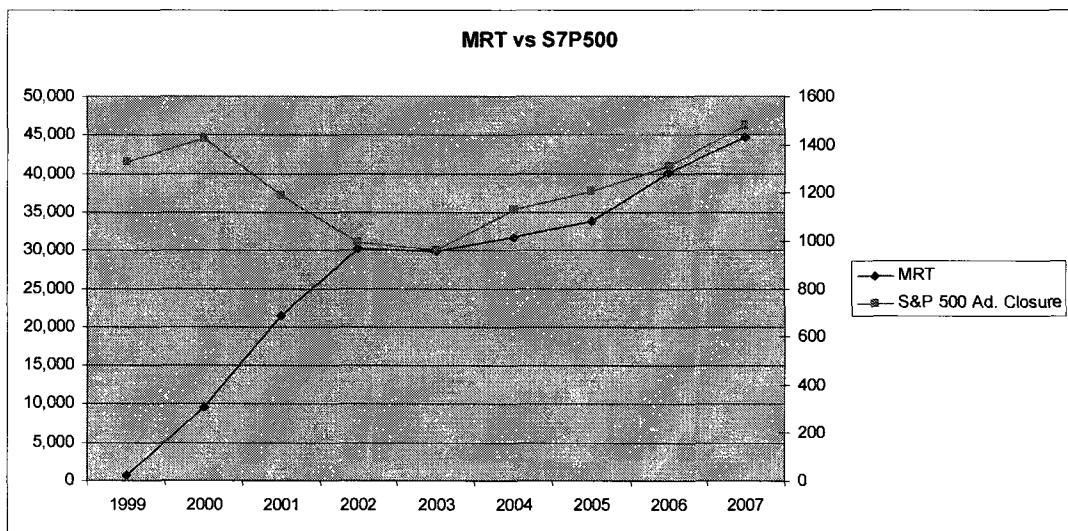
Le nombre de mises en services semble réagir plus au WGD que le nombre d'heures de vol, mais encore, les deux croissent, mais rien ne nous indique une très forte corrélation.

Prenons maintenant l'Indice S&P500 (500 compagnies américaines incluses dans l'indice maintenu par Standard & Poor's). Nous souhaitons voir si le nombre de mises en service est atteint par les fluctuations de l'Indice S&P500, à la figure 62, puis si le nombre d'heures de vol mensuel est atteint aussi par le S&P500, à la figure 63.



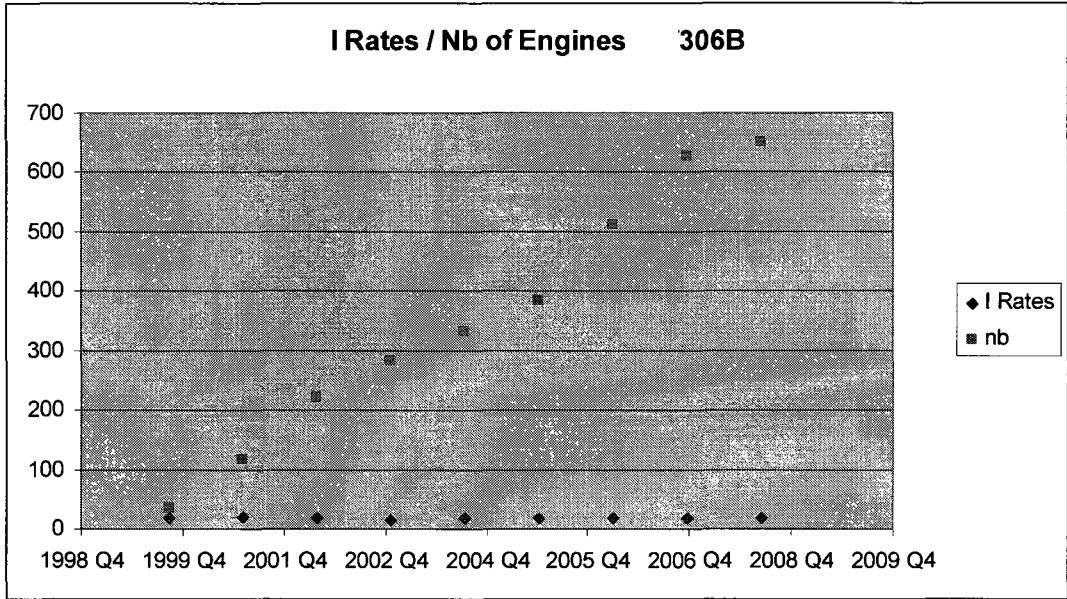


**Figure 62 - Nombre de mises en service VS S&P500**

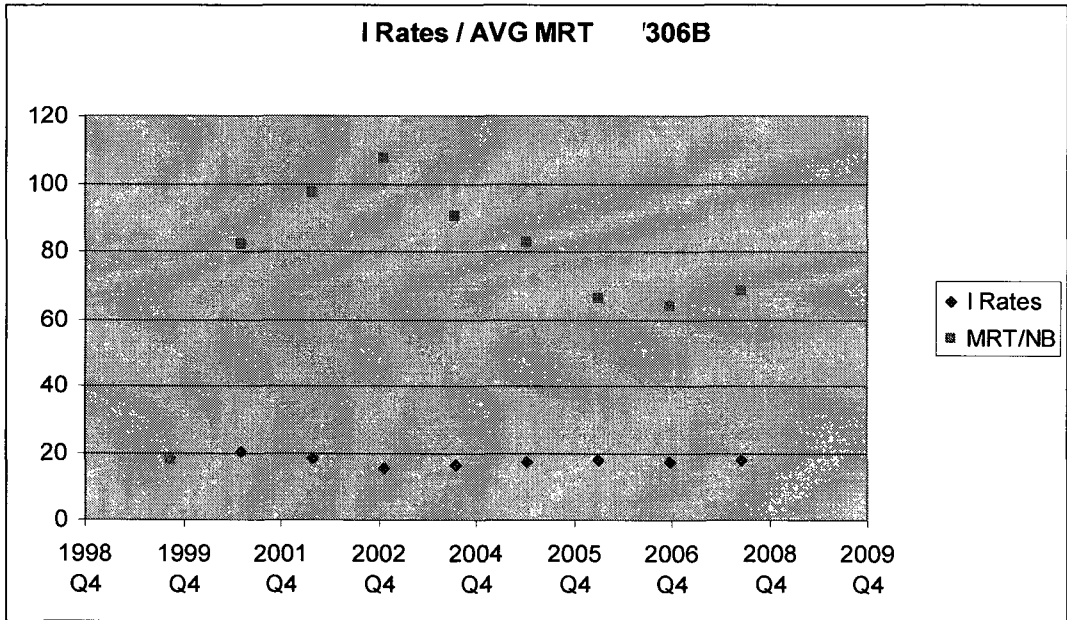


**Figure 63 - L'usage mensuel en nombre d'heures de vol VS S&P500**

Le nombre d'heures de vol et le S&P500 ont une forme similaire, beaucoup plus visible que le nombre de mises en services comparé au S&P500. Finalement, qu'en est-il des taux d'intérêts aux États-Unis (USIR)? La variation des taux d'intérêts affecte-t-elle le nombre de mises en services ou le nombre d'heures de vol? Selon les graphiques des figures 64 et 65, il ne semble pas y avoir de corrélation.



**Figure 64 - le nombre de mises en services VS USIR**



**Figure 65 - L'usage mensuel en nombre d'heures de vol VS USIR**

## **Annexe C - la description détaillée des nombreuses caractéristiques retenues en entrée.**

### **Numéro**

Source: TRAC Database (Entrepôt de données historiques)

Utilité: Attribuer un numéro unique à l'événement rencontré. Sert de clé primaire. Il s'agit d'un numéro séquentiel. Nombre entier  $0 < x < n$ , où n correspond au nombre total d'événements pour ce modèle de moteur.

Importance : Ce champ n'a aucune influence sur l'usure du moteur et la prédiction de futures visites en usine.

### **Engine Serial Number**

Source: TRAC Database

Utilité: Chaque moteur possède un numéro de série unique, composé de 2 lettres et 4 chiffres, prenant la forme suivante « XX0000 ». Les deux premières lettres identifient le type de moteur.

Importance : Ce champ nous permet de savoir à quel moteur l'événement est relié. Un même moteur peut avoir plusieurs événements.

### **Days between events**

Source: Champ calculé

Utilité: Ce champ calcule le délai entre les événements d'un même moteur en nombre de jours. En classant les événements en ordre croissant de dates, on peut identifier ce délai. Il suffit, pour un même numéro de série, de soustraire la date du dernier événement de celle de l'événement courant.

Importance : Ce champ est l'un des 2 champs cibles. Nous souhaitons prédire quand aura lieu le prochain événement pour un moteur précis.

### **Event Date**

Source: TRAC Database

Utilité: Chaque événement sur un moteur se produit à une date donnée, prenant la forme « JJ-MMM-AA ».

Importance : Ce champ nous permet de savoir à quelle date un événement a eu lieu. Ce champ sert à calculer le nombre de jours entre les événements. Je ne crois pas

que ce champ soit nécessaire comme entrée pour prédire les prochaines visites en usine.

### **Political Belonging**

Source: TRAC Database

Utilité: Chaque moteur appartient à une entité politique (Pays/Région).

Importance : Ce champ nous permet de savoir à quelle région politique ce moteur appartient. Ce champ est important pour la prédiction des visites en usines, puisque certaines régions se conforment plus aux exigences et délais de maintenance que d'autres.

Pour les XX306B, nous avons les possibilités présentées à la table 22.

**Table 22 - Répartition des pays d'appartenance pour les 306B**

Austria Count	7
Canada Count	3
China Count	285
French Polynesia Count	4
Germany Count	6
Italy Count	41
Nigeria Count	27
U.S.A. Count	397
United Kingdom Count	1

### **Engine Time Since New**

Source: TRAC Database

Utilité: Chaque moteur a un nombre d'heures de vol depuis sa mise en service. Ce champ contient le nombre total d'heures de vol depuis sa mise en service.

Importance : Ce champ nous permet de savoir combien d'heures de vol le moteur a volé au total. Il est important pour la prédiction des visites en usine, puisqu'un moteur ayant moins d'heures de vol est moins susceptible d'avoir besoin d'une visite en usine.

### **Engine TSN Recorded Date**

Source: TRAC Database

Utilité: Chaque moteur a un nombre d'heures de vol depuis sa mise en service. Ce champ contient la date à laquelle le nombre total d'heures de vol depuis la mise en service du moteur a été enregistrée.

Importance : Ce champ nous permet de savoir quand on a enregistré le nombre d'heures de vol depuis la mise en service du moteur. Je ne crois pas que ce champ soit nécessaire comme entrée pour prédire les prochaines visites en usine.

### **Engine Cycles Since New**

Source: TRAC Database

**Utilité:** Chaque moteur a un nombre de cycles depuis sa mise en service. Ce champ contient le nombre total de cycles depuis sa mise en service. Un cycle correspond à un intervalle <décollage, vol, atterrissage>.

**Importance :** Ce champ nous permet de savoir combien de cycles le moteur a effectué au total. Il est important pour la prédiction des visites en usine, puisqu'un moteur ayant moins de cycles est moins susceptible d'avoir besoin d'une visite en usine. D'ailleurs, plus les cycles sont longs, moins le moteur sera fatigué et moins il aura besoin d'avoir de visites en usine. L'inverse est vrai aussi : plus les cycles sont fréquents et courts, plus les chances augmentent de voir ce moteur entrer en usine.

### **Engine CSN Recorded Date**

**Source:** TRAC Database

**Utilité:** Chaque moteur a un nombre de cycles depuis sa mise en service. Ce champ contient la date à laquelle le nombre total de cycles depuis la mise en service du moteur a été enregistré.

**Importance :** Ce champ nous permet de savoir quand on a enregistré le nombre de cycles depuis la mise en service du moteur. Je ne crois pas que ce champ soit nécessaire comme entrée pour prédire les prochaines visites en usine.

### **Time Between Overhaul**

**Source:** TRAC Database

**Utilité:** Chaque moteur a un délai fixe de maintenance prévu (Overhaul). Ce champ indique la durée de l'intervalle de maintenance prévue sur les « *Overhaul* » (OH). Il est possible d'obtenir une extension de ce délai si l'opérateur est fidèle aux délais prescrits. (Bonne conduite – aussi appelé « *Soft-Time* »). Si, par contre, l'opérateur n'est pas fidèle aux délais prescrits, il ne pourra pas bénéficier de délais supplémentaires (habituellement par tranches de 500h), mais il peut se voir attribuer une réduction du délais (-500h, par exemple, pour mauvaise conduite – aussi appelé « *Hard-Time* »).

**Importance :** Ce champ nous permet de savoir quel est le délai de maintenance fixe pour les OH. Ceci est un des principaux indicateurs pour savoir quand le moteur entrera en usine pour une maintenance prévue. (Type=BPO). Ce champ est TRES important pour la prédiction des visites en usines.

### **Time Since Overhaul**

**Source:** TRAC Database

**Utilité:** Chaque moteur a un délai depuis sa dernière maintenance prévue (BPO). Ce champ contient le nombre d'heures de vol depuis le dernier événement OH prévu.

**Importance :** Ce champ nous permet de savoir quand on a effectué la dernière maintenance (BPO) sur ce moteur, en termes de nombre d'heures depuis. Ce champ est très important pour effectuer une prédiction sur la prochaine visite possible.

### **Cycles Since Overhaul**

**Source:** TRAC Database

Utilité: Chaque moteur a un délai depuis sa dernière maintenance prévue (BPO). Ce champ contient le nombre de cycles effectués depuis le dernier événement OH prévu.

Importance : Ce champ nous permet de savoir combien de cycles ont été effectués depuis la dernière maintenance (BPO) sur ce moteur. Ce champ est très important pour effectuer une prédiction sur la prochaine visite possible. Encore une fois, plus le nombre de cycles sera élevé, plus le risque d'une visite en usine sera élevé et encore plus si la durée des cycles est de courte durée.

### **Aircraft Serial Number**

Source: TRAC Database

Utilité: Chaque moteur est utilisé sur un aéronef (avion, hélicoptère, ...). Un aéronef peut utiliser 1 ou plusieurs moteurs (maximum de 8). Un moteur est interchangeable, alors il peut changer de position sur l'aéronef (position 1, 2, 3, 4, 5, 6, LH, RH, spare) et il peut également se retrouver sur un autre aéronef quelques semaines plus tard. Ce champ contient le numéro de série de l'aéronef sur lequel il était installé lorsque l'événement a eu lieu.

Importance : Ce champ nous permet de savoir sur quel aéronef le moteur était installé lors de l'événement. Ce champ peut être important, dans la mesure où le fait d'installer ce moteur sur un aéronef plutôt qu'un autre peut réduire sa durée de vie, ou augmenter l'usure qu'il subira.

### **Operator Class**

Source: TRAC Database

Utilité: Chaque opérateur d'aéronef, sur lequel figure un ou plusieurs moteurs appartient à une classe d'opérateur prédéterminée. Ce champ contient la classe de l'opérateur.

Importance : Ce champ nous permet de savoir quelle est la classe de l'opérateur qui utilise ce moteur. Certaines classes sont plus susceptibles d'entraîner de la fatigue sur les moteurs. Ce champ est important pour la prédiction, s'il peut apporter d'autres informations que celles en relation avec le nombre de cycles

**Table 23 - Répartition des classes d'opérateurs pour les 306B**

AIRLINE Count	77
COMMUTER Count	643
CORPORATE/UTILITY Count	47
EQUIPMENT MANUFACTURER Count	4

### **Operator Name**

Source: TRAC Database

Utilité: Chaque opérateur d'aéronef, sur lequel figure un ou plusieurs moteurs appartient possède un nom. Ce champ contient le nom de l'opérateur.

Importance : Ce champ nous permet de savoir quel est le nom de l'opérateur qui opère l'aéronef sur lequel le moteur était installé lors de l'événement. Ce champ peut être important, dans la mesure où les opérateurs n'ont pas tous le même type d'utilisation, ou encore le même type de rigueur pour la maintenance.

Table 24 - Répartition des noms d'opérateurs pour le 306B

ADAC LUFTRETTUNG GMBH Count	2
AERO-DIENST GMBH Count	1
AIR VALLEE S.P.A. Count	10
ALTRIA GROUP, INC. Count	2
ATLANTIC COAST AIRLINES Count	255
AVCRAFT AVIATION LLC Count	1
BRISTOW HELICOPTERS GROUP LTD. Count	8
CHANG AN AIRLINES CO., LTD Count	212
CIRRUS AIRLINES Count	1
FAIRCHILD DORNIER GMBH Count	1
GANDALF AIRLINES Count	27
GREAT PLAINS AIRLINES Count	14
HAINAN AIRLINES CO LTD Count	74
ITALI AIRLINES SRL Count	2
JOHNSON CONTROLS INC. Count	4
OZARK MANAGEMENT INC. Count	2
PRATT & WHITNEY CANADA LEASING INC. Count	3
SHELL PETROLEUM DEVELOPMENT CO. NIGERIA Count	20
SKYWAY AIRLINES Count	111
TYROL AIR AMBULANCE Count	3
TYROLEAN AIRWAYS GMBH Count	1
TYROLEAN JET SERVICE GMBH Count	3
ULTIMATE JET CHARTER Count	5
UNION PLANTERS NATIONAL BANK Count	3
UNKNOWN Count	4
WANAIR Count	2

### Operator Short Name

Source: TRAC Database

Utilité: Ce champ contient une version réduite du nom de l'opérateur.

Importance : Ce champ nous permet de savoir quel est le nom de l'opérateur qui opère l'aéronef sur lequel le moteur était installé lors de l'événement. Ce champ peut être important, dans la mesure où les opérateurs n'ont pas tous le même type d'utilisation, ou encore le même type de rigueur pour la maintenance. Ce champ est également la répétition du champ « *Operator Name* ». Seulement un des deux devra être considéré. Reste à trouver lequel ?

### Owner Name

Source: TRAC Database

Utilité: Ce champ contient le nom du détenteur. Ce n'est pas nécessairement celui qui l'opère.

VALEURS MANQUANTES: 372 valeurs sur 771.

Importance : Ce champ nous permet de savoir à qui appartient l'aéronef sur lequel est installé le moteur. Cependant, il y a beaucoup de valeurs manquantes. Ce champ ne sera pas considéré (contient 48% de valeurs manquantes).

**Table 25 - Répartition des noms courts des détenteurs pour le 306B**

AERO-DIENST GMBH Count	2
AIR VALLEE S.P.A. Count	5
ASTRAL SKYWAY Count	9
ATLANTIC COAST AIRLINES Count	29
ATLANTIC COAST AIRLINES HOLDINGS INC Count	2
ATLANTIC COAST AIRLINES HOLDINGS, INC Count	2
AVCRAFT AEROSPACE Count	1
AVCRAFT AVIATION LLC Count	1
AVIA CONSULT FLUGTRIEBSWERK - GMBH Count	1
BRISTOW HELICOPTERS GROUP LTD. Count	6
CHANG AN AIRLINES CO., LTD Count	2
DORNIER Count	1
DORNIER LUFTFAHRT GMBH Count	3
GANDALF A/L Count	1
GANDALF AIRLINES Count	7
GREAT PLAINS AIRLINES Count	1
HAINAN Count	1
HAINAN AIRLINES Count	88
HAINAN AIRLINES CO LTD Count	47
ITALI AIRLINES SRL Count	2
JOHNSON CONTROLS INC. Count	4
OZARK AIRLINES Count	2
PRATT & WHITNEY CANADA LEASING INC. Count	100
PRATT & WHITNEY ENGINE SERVICES, INC. Count	1
PRATT AND WHITNEY CANADA LEASING INC. Count	5
PWCL(CSSP) Count	1
SHELL PETROLEUM DEVELOPMENT CO. NIGERIA Count	5
SKYWAY AIRLINES Count	61
TYROLEAN JET SERVICE GMBH Count	2
ULTIMATE JET CHARTER Count	2
UNION PLANTERS NATIONAL BANK Count	3
WELCOME AIR LUFTFAHRT GMBH & CO KG Count	2
Grand Count	399

**Country**

Source: TRAC Database

Utilité: Ce champ contient le nom du pays où l'événement a eu lieu.

Importance : Ce champ nous permet de savoir où exactement l'événement a eu lieu.

Il est possible que ce champ ait de l'influence sur les visites en usine, mais je crois



que les facteurs environnementaux auront un plus grand impact sur la prédiction des visites en usine que ce champ.

**Table 26 - Répartition des pays d'événements pour le 306B**

Austria Count	7
Canada Count	3
China Count	285
French Polynesia Count	4
Germany Count	6
Italy Count	41
Nigeria Count	27
U.S.A. Count	397
United Kingdom Count	1

**State/Province**

Source: TRAC Database

Utilité: Ce champ contient le nom de l'état ou de la province où l'événement a eu lieu.

Importance : Ce champ nous permet de savoir où exactement l'événement a eu lieu, et encore plus précisément que seulement le nom du pays. Il est possible que ce champ ait de l'influence sur les visites en usine, mais je crois que les facteurs environnementaux auront un plus grand impact sur la prédiction des visites en usine que ce champ.

VALEURS MANQUANTES: 68 « UNKNOWN » sur 771.

**Table 27 - Répartition des états/provinces d'événements pour le 306B**

Bavaria Count	3
Bergamo Count	8
England Count	1
Hainan Dao Count	73
Illinois Count	1
Kentucky Count	1
Lagos Count	3
Massachusetts Count	4
Michigan Count	4
Missouri Count	7
New York Count	1
North Carolina Count	1
Ohio Count	5
Oklahoma Count	8
Pescara Count	2
Quebec Count	2
Saarland Count	1
Shaanxi Count	209

South Carolina Count	19
Tahiti Archipelago Count	4
Tennessee Count	3
Tirol Count	3
UNKNOWN Count	68
Virginia Count	230
Wisconsin Count	110

### Event Classification 3

Source: TRAC Database

Utilité: Ce champ contient le type de l'événement.

Importance : Ce champ nous permet de savoir quel est le type de l'événement qui a eu lieu. Il s'agit du deuxième champ cible. Il est très important.

**Table 28 - Répartition des événements du 306B**

BPO Count	6	« <i>(Less WORSE) Basic Planned Overhaul</i> »
BPR Count	431	« <i>Basic Planned Removal</i> »
BUO Count	2	« <i>Basic Unplanned Overhaul</i> »
BUR Count	277	« <i>Basic Unplanned Removal</i> »
NUR Count	55	« <i>(WORSE) Non-basic Unplanned Removal</i> »

### Engine age in days

Source: TRAC Database

Utilité: Ce champ contient l'âge du moteur en nombre de jours.

Importance : Ce champ nous permet de savoir quel est l'âge du moteur. Je crois qu'il est important pour la prédiction, puisque plus un moteur est âgé, plus il est susceptible de nécessiter une visite en usine. On représente la fonction de retraite comme un graphique présenté à la figure 66.:

**Table 29 - Statistiques sur l'âge des 306B**

MIN	1742.2425
MAX	3141.15
AVERAGE	2521.802539

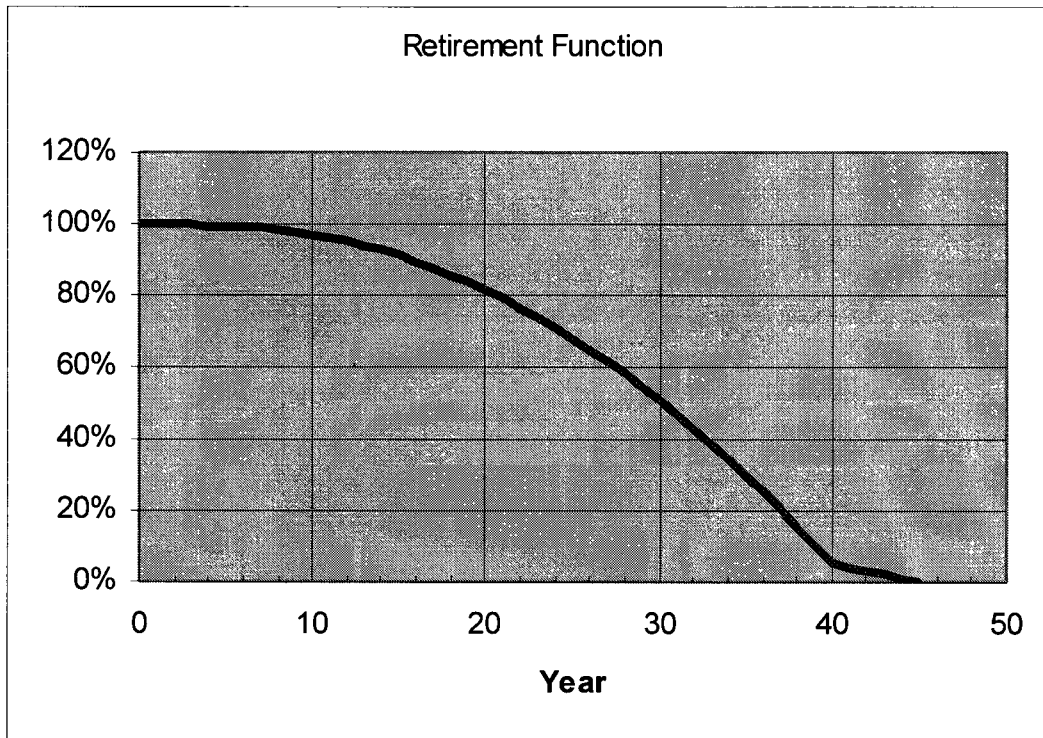


Figure 66 - Fonction de retraite des moteurs

### Engine Age

Source: TRAC Database

Utilité: Ce champ contient l'âge du moteur en nombre d'années.

Importance : Ce champ nous permet de savoir quel est l'âge du moteur. Je crois qu'il est important pour la prédiction, puisque plus un moteur est âgé, plus il est susceptible de nécessiter une visite en usine. Cependant, ce champ duplique l'information contenue dans « **Engine age in days** ».

### Module Age

Source: TRAC Database

Utilité: Ce champ contient la date d'entrée en service (EIS) du moteur.

Importance : Ce champ nous permet de savoir quel est l'âge du moteur. Je crois qu'il est important pour la prédiction, puisque plus un moteur est âgé, plus il est susceptible de nécessiter une visite en usine. Cependant, ce champ duplique l'information contenue dans « **Engine age in days** » et dans « **Engine Age** ».

### Number Of Events

Source: TRAC Database

Utilité: Ce champ contient une valeur unique définissant s'il s'agit de l'événement le plus récent. Malheureusement, il n'a pas été utilisé à bon escient.

Importance : Ce champ est inutilisable.

### Engine Target BUR

**Source:** TRAC Database  
**Utilité:** Ce champ contient une valeur unique définissant le taux de BUR à atteindre pour ce modèle de moteur.  
**Importance :** Ce champ est inutilisable.

### **Engine Target IFSD**

**Source:** TRAC Database  
**Utilité:** Ce champ contient une valeur unique définissant le taux de In Flight Shut-Down (IFSD) à atteindre pour ce modèle de moteur.  
**Importance :** Ce champ est inutilisable.

### **Installation Type**

**Source:** TRAC Database  
**Utilité:** Ce champ contient une valeur unique définissant sur quel type d'aéronef le moteur est installé.  
**Importance :** Ce champ est inutilisable.

### **Sand**

**Source:** Modèle des coûts de maintenance – Recherche environnementale.  
**Utilité:** Ce champ contient le taux moyen de sable présent à l'aéroport/la région où est opéré l'aéronef. Cette valeur est obtenue du modèle de coûts de maintenance (DMC).  
**Importance :** Ce champ important. Il a un impact direct sur l'usure prématurée du moteur.

### **Salt**

**Source:** Modèle des coûts de maintenance – Recherche environnementale.  
**Utilité:** Ce champ contient le taux moyen de sel présent à l'aéroport/la région où est opéré l'aéronef. Cette valeur est obtenue du modèle de coûts de maintenance (DMC).  
**Importance :** Ce champ important. Il a un impact direct sur l'usure prématurée du moteur.

### **Volcano**

**Source:** Modèle des coûts de maintenance – Recherche environnementale.  
**Utilité:** Ce champ contient le taux moyen d'éruptions volcaniques présentes à l'aéroport/la région où est opéré l'aéronef. Cette valeur est obtenue du modèle de coûts de maintenance (DMC).  
**Importance :** Ce champ important. Il a un impact direct sur l'usure prématurée du moteur.

### **Pollution**

**Source:** Modèle des coûts de maintenance – Recherche environnementale.  
**Utilité:** Ce champ contient le taux moyen de pollution présent à l'aéroport/la région où est opéré l'aéronef. Cette valeur est obtenue du modèle de coûts de maintenance (DMC).

Importance : Ce champ important. Il a un impact direct sur l'usure prématurée du moteur.

### **Elevation**

Source: Modèle des coûts de maintenance – Recherche environnementale.

Utilité: Ce champ contient le taux moyen d'élévation présent à l'aéroport/la région où est opéré l'aéronef. Cette valeur est obtenue du modèle de coûts de maintenance (DMC).

Importance : Ce champ important. Il a un impact direct sur l'usure prématurée du moteur.

### **Crude Oil Price**

Source: Global Insight – Web reference on Macro-economics factors.

Utilité: Ce champ contient le prix du pétrole brut au moment où l'événement a eu lieu.

Importance : Ce champ important. Il a un impact direct sur l'utilisation du moteur.

### **USD**

Source: Global Insight – Web reference on Macro-economics factors.

Utilité: Ce champ contient le prix du dollar américain au moment où l'événement a eu lieu.

Importance : Ce champ important. Il a un impact direct sur la mise en marché de nouveaux moteurs corporatifs.

### **S&P500**

Source: Global Insight – Web reference on Macro-economics factors.

Utilité: Ce champ contient la cote boursière du S&P500 au moment de l'événement.

Importance : Ce champ important. Il a un impact direct sur l'utilisation du moteur.

### **Corp Profits US**

Source: Global Insight – Web reference on Macro-economics factors.

Utilité: Ce champ contient la valeur des profits corporatifs aux US à la date de l'événement.

Importance : Ce champ important. Il a un impact direct sur la mise en marché de nouveaux moteurs corporatifs.

### **Interest Rates US**

Source: Global Insight – Web reference on Macro-economics factors.

Utilité: Ce champ contient la valeur des taux d'intérêts aux US à la date de l'événement.

Importance : Ce champ important. Il a un impact direct sur la mise en marché/l'utilisation de nouveaux moteurs corporatifs.

### **W GDP**

Source: Global Insight – Web reference on Macro-economics factors.

**Utilité:** Ce champ contient la croissance mondiale à la date de l'événement.  
**Importance :** Ce champ important. Il a un impact direct sur la mise en marché de nouveaux moteurs corporatifs.

### **USDCAD=X**

**Source:** Global Insight – Web reference on Macro-economics factors.  
**Utilité:** Ce champ contient la valeur du dollar US par rapport au dollar Canadien à la date de l'événement.  
**Importance :** Ce champ important. Il a un impact direct sur la mise en marché de nouveaux moteurs corporatifs.

### **Nb of Engines**

**Source:** TRAC Database.  
**Utilité:** Ce champ contient le nombre de moteurs en service pour l'ensemble de la famille de moteurs à la date de l'événement.  
**Importance :** Ce champ permet de diviser le total des heures volées par le nombre de moteurs en service et avoir une moyenne d'utilisation par moteur.

### **MRT**

**Source:** TRAC Database.  
**Utilité:** Ce champ contient le nombre d'heures de vol mensuel (Monthly Running Time) pour l'ensemble de la famille de moteurs à la date de l'événement.  
**Importance :** Ce champ permet de diviser le total des heures volées par le nombre de moteurs en service et avoir une moyenne d'utilisation par moteur.

### **TRT**

**Source:** TRAC Database.  
**Utilité:** Ce champ contient le nombre d'heures de vol total (Total Running Time) pour le moteur en service.  
**Importance :** Ce champ permet de diviser le total des heures volées par le moteur.

### **AMU**

**Source:** TRAC Database.  
**Utilité:** Ce champ contient le nombre moyen d'heures de vol mensuelles pour tous les moteurs de la famille à la date de l'événement.  
**Importance :** Ce champ permet de diviser le total des heures volées par le nombre de moteurs en service et avoir une moyenne mensuelle d'utilisation par moteur.

## Annexe D - la conversion des valeurs catégorielle en valeurs numériques (« *ranking* »).

Parmi les caractéristiques proposées, plusieurs sont catégorielles. Nous faisons références aux valeurs de couleur JAUNE dans les exemples précédents (Annexe C, p.126). Afin de rendre ces variables numériques, il faut effectuer un score pour chaque possibilité (chaque classe) de la caractéristique.

Pour chacun de ces champs, voyons ensemble la méthode utilisée pour numériser la valeur de la caractéristique.

Table 30 - Conversion du numéro de série

Engine Serial Number	
XX0205	→ 0205

Dans ce premier cas, à la table 30, le numéro de série sert uniquement à savoir de quel moteur nous parlons. A l'aide de ces numéros de série, nous avons pu trouver (en triant les données par ordre croissant de numéros) le délai entre les événements. A noter que ce champ a été nommé « *Days Between Events* ».

Les 2 premiers caractères ont été retirés et le numéro alphanumérique a ainsi été converti en une valeur strictement numérique, sans perte d'information intrinsèque, puisque chaque modèle de moteur a ses propres caractères

d'identification. Dans la même famille, tous les moteurs ont les mêmes lettres. (XX dans le cas du moteur XX306B).

**Table 31 - Conversion du Pays d'appartenance**

Political Belonging	
U.S.A.	→ 0.4380

Afin de rendre numérique ce champ, le score a été obtenu en classant les pays par ordre de richesse, sous entendant ainsi que les pays les plus riches sauraient davantage comment maintenir un moteur en bon état de vol. Ainsi, les moteurs détenus par la Chine, par exemple, obtiennent un score plus faible et ceci reflète la réalité puisque les opérateurs chinois ont plus tendance à être délinquants comparativement aux opérateurs américains. Les scores obtenus sont présentés dans la table 32.

**Table 32 - Résultats de la conversion des pays d'appartenance**

Nigeria	0.0150
China	0.0780
French Polynesia	0.1750
Italy	0.3020
United Kingdom	0.3180
Germany	0.3190
Austria	0.3470
Canada	0.3570
U.S.A.	0.4380

**Table 33 - Conversion du numéro de série de l'aéronef**

Aircraft Serial Number	→ #####
DOR####	

Dans ce cas, le numéro de série sert uniquement à savoir de quel aéronef nous parlons. Les 3 premiers caractères ont été retirés et le numéro alphanumérique a ainsi été converti en une valeur strictement numérique, avec possible perte d'information intrinsèque, puisque chaque modèle d'aéronef a ses propres caractères d'identification, mais il se peut que les nombres reviennent d'un aéronef



à l'autre. Ceci risque de causer de la confusion, alors il risque de ne pas être considéré dans le modèle.

**Table 34 - Conversion de la classe de l'opérateur**

Operator Class
COMMUTER

Cette caractéristique est importante (table 34), puisqu'elle peut nous éclairer sur le type de mission de l'opérateur, elle nous aide à estimer le genre de vols que l'opérateur risque d'effectuer. Habituellement, les lignes aériennes sont celles qui effectuent les vols les plus long (donc, il y a moins d'usure sur le moteur, puisque la puissance maximale est utilisée seulement lors du décollage et c'est ce qui endommage le plus le moteur). En revanche, les commutateurs (commuter) sont ceux dont les vols risquent d'être les plus courts, alors il y a beaucoup plus de chances d'endommager le moteur, puisque la demande maximale de puissance est requise beaucoup plus fréquemment qu'une ligne aérienne pourrait le faire. Les scores obtenus sont présentés à la table 35.

**Table 35 - Résultats de la conversion des classes d'opérateurs**

AIRLINE	0.3
EQUIPMENT	
MANUFACTURER	0.5
CORPORATE/UTILITY	0.7
COMMUTER	0.9

{	Operator Name	
	ATLANTIC COAST AIRLINES	
	Operator Short Name	Ces trois sont considérés comme un ensemble
ATLANTIC COAST		
	Owner Name	

Ces trois derniers champs sont difficiles à catégoriser. Ils ont donc été regroupés dans un quatrième champ appelé « *Max proba owner and operator name* ». Nous avons tenté de trouver la possibilité que l'opérateur et le détenteur

soient dans des pays riches et ne soient pas délinquants. Les scores obtenus sont présentés à la table 36. Plus le score est élevé, plus le pays est délinquant et plus les moteurs sont usés lorsqu'ils entrent en atelier pour la maintenance:

**Table 36 - Scores des pays riches et non délinquants**

ATLANTIC COAST AIRLINES	ATLANTIC COAST		0
BRISTOW HELICOPTERS GROUP LTD.	BRISTOW HELI		0
AIR VALLEE S.P.A.	AIR VALLEE		0
ALTRIA GROUP, INC.	ALTRIA GROUP		0
AERO-DIENST GMBH	AERO-DIENST		0
HAINAN AIRLINES CO LTD	HAINAN AIRLINES	CANADA LEASING INC.	0.5
UNKNOWN	UNKNOWN	CANADA LEASING INC.	0.5
CHANG AN AIRLINES CO., LTD	CHANGAN A/L	CANADA LEASING INC.	0.5
SKYWAY AIRLINES	SKYWAY A/L	CANADA LEASING INC.	0.5
CHANG AN AIRLINES CO., LTD	CHANGAN A/L		0.5
CANADA LEASING INC.	P&WC LEASING IN		0.5
TYROL AIR AMBULANCE		WELCOME AIR LUFTFAHRT GMBH & CO KG	0.5
SHELL PETROLEUM DEVELOPMENT CO. NIGERIA	SHELL PETROLEUM	CANADA LEASING INC.	0.5
ULTIMATE JET CHARTER	AVCRAFT JET CHA	ULTIMATE JET CHARTER	0.5
GANDALF AIRLINES	GANDALF		0.5
ATLANTIC COAST AIRLINES	ATLANTIC COAST	CANADA LEASING INC.	4
BRISTOW HELICOPTERS GROUP LTD.	BRISTOW HELI	CANADA LEASING INC.	4
AVCRAFT AVIATION LLC	AVCRAFT AVIA.	AVCRAFT AVIATION LLC	4
AIR VALLEE S.P.A.	AIR VALLEE	AIR VALLEE S.P.A.	6
CHANG AN AIRLINES CO., LTD	CHANGAN A/L	HAINAN	6
CHANG AN AIRLINES CO., LTD	CHANGAN A/L	CHANG AN AIRLINES CO., LTD	6
BRISTOW HELICOPTERS GROUP LTD.	BRISTOW HELI	SHELL PETROLEUM DEVELOPMENT CO. NIGERIA	12
ATLANTIC COAST AIRLINES	ATLANTIC COAST	DORNIER LUFTFAHRT GMBH	60
CHANG AN AIRLINES CO., LTD	CHANGAN A/L	HAINAN AIRLINES	161

**Table 37 - Pays**

Country
U.S.A.

Ce champ (table 37) peut être ignoré, à moins que l'événement n'ait pas lieu dans la même région que celle où il appartient (« *Political Belogning* »). Dans ce dernier cas, ce champ peut nous aider à retrouver les valeurs environnementales pour la région où l'événement a eu lieu, plutôt que de viser la région d'appartenance.

**Table 38 - État/Province**

State/Province
Virginia

La province (Table 38) est surtout utilisée pour effectuer la recherche verticale afin trouver les facteurs environnementaux, lorsque le pays est trop vaste pour avoir une seule valeur unique, alors nous utilisons la province (comme par exemple dans les Etats-Unis, chaque province a ses taux environnementaux).

**Table 39 - Classification d'événement sur 3 caractères**

Event Classification 3
BUR

Ceci est l'un des champs clés (table 39). Il s'agit d'une partie de la prédiction (« *Target* ») à fournir. Il indique le type d'événement qui s'est produit pour que le moteur ait besoin d'une visite en usine. La première lettre signifie s'il s'agit d'un événement « *B-Basic* » ou « *N-Non Basic* ». Les événements de base (« *basic* ») sont moins graves que les événements qui ne sont pas de base (« *Non Basic* »). La deuxième lettre signifie si l'événement est prévu (« *Planned* ») ou non-prévu (« *Unplanned* »). Habituellement, les événements non-prévus sont plus graves que les prévus. Finalement, la troisième lettre nous renseigne sur ce qui a été fait comme maintenance sur ce moteur en relation avec cet événement. Par exemple, un

« O » signifie un « *overhaul* » (on doit défaire le moteur en morceaux), alors qu'un « R » signifie « *Removal* » (on a du enlever le moteur). Les scores obtenus sont listés dans la table 40. :

**Table 40 – Scores des types d'événements**

BPR	4
BPO	5
NUR	6
BUR	8
BUO	9

## Annexe E – Jeux d'observation sur le SVM.

### ***E.1 Premier jeu d'essai incluant TOUTES les caractéristiques***

Nous avons utilisé les caractéristiques suivantes :

Numéro, Normalized SN, Days between events, Political Belonging Normalized, Engine Time Since New, Engine TSN Recorded Date, Engine Cycles Since New, Time Between Overhaul, Time Since Overhaul, Cycles Since Overhaul, Normalized A/c SN, Opr clas normalized, Max proba owner and operator name, Event C 3 Rating, Engine age in days, Engine Age, Sand, Salt, Volcano, Pollution, Elevation, Crude Oil Price USD, S&P500, Corp Profits US, Interest Rates US, W GDP, USDCAD=X, Nb of Engines, MRT, TRT, AMU, Target(La classe cible: 1=Planned 2=BasicUnplanned 3=Non-BasicUnplanned). On peut consulter les paramètres du SVM après le jeu d'entraînement à la table 41.

**Table 41 - Paramètres du SVM**

MSE	C	epsilon	nu	kernel		Avec 100
4.06E-08	10000	1.00E-06		type radial	gamma 10	« runs ».

Pour le jeu de validation, nous avons utilisé l'en-tête suivant:

```
# svm example set
```

```
@examples
dimension 31 (nombre de caractéristiques)
number 617 (nombre d'exemples, d'enregistrements)
b -1.051360415765215
```

Les résultats obtenus n'étaient pas satisfaisants (pour prédire les 154 enregistrements):

```
# examples ys
-306.243
-182.517
-96.0364
-312.546
-171.809
-37.5184 [...]
```

Tellement que l'erreur moyenne était de 37.6841485.

## ***E.2 Deuxième jeu d'essai incluant CERTAINES des caractéristiques choisies***

Nous avons utilisé les caractéristiques suivantes :

Days between events, Political Belonging Normalized, Engine Time Since New, Engine Cycles Since New, Time Between Overhaul, Opr clas normalized, Event C 3 Rating, Engine age in days, Sand, Salt, Volcano, Pollution, Elevation, Crude Oil Price USD, S&P500, Corp Profits US, Interest Rates US, W GDP, USDCAD=X, Nb of Engines, MRT, TRT, AMU, Target(La classe cible: 1=Planned 2=BasicUnplanned 3=Non-BasicUnplanned).

Nous avons utilisé les paramètres « *DOT standard* » du SVM, sur cent exécutions (« *runs* »). Pour le jeu de validation, nous avons utilisé l'en-tête suivant:

```
# svm example set
@examples
dimension 23 (nombre de caractéristiques)
number 100 (nombre d'exemples, d'enregistrements)
```

b -0.1671717133712994

Les résultats obtenus (table 42) sont beaucoup plus satisfaisants (pour prédire les 25 enregistrements).

**Table 42 – Résultats obtenus sur le 2<sup>ième</sup> jeu d'essai**

Target	result	delta
1	1.00054	-0.00054
1	0.999512	0.000488
1	0.999606	0.000394
1	1.00017	-0.00017
1	1.00025	-0.00025
1	1.00174	-0.00174
1	1.00015	-0.00015
1	1.00061	-0.00061
1	1.00029	-0.00029
1	0.999846	0.000154
1	1.00002	-2E-05
1	1.00029	-0.00029
1	1.00048	-0.00048
1	1.00014	-0.00014
2	2.00006	-6E-05
2	2.00217	-0.00217
2	1.99986	0.00014
2	2.0012	-0.0012
2	1.99949	0.00051
2	1.99978	0.00022
2	1.99974	0.00026
3	1.5005	1.4995
3	1.50133	1.49867
3	1.49991	1.50009
3	1.49995	1.50005

Ici, l'erreur moyenne était de 0.23969464, mais en considérant la 3<sup>ième</sup> classe. Si nous ne calculons que les deux premières classes (séparabilité binaire), alors le résultat est encore meilleur : -0.000283.

### ***E.3 Troisième jeu d'essai incluant CERTAINES des caractéristiques choisies***

Comme les résultats précédents étaient satisfaisants, pour séparer les événements de type Prévus, comparativement aux types imprévus, nous avons tenté d'effectuer une classification/prédiction sur un plus grand ensemble de données.

Nous avons utilisés les mêmes caractéristiques et la même configuration SVM, avec l'en-tête suivante:

```
# svm example set
@examples
dimension 23
number 771
b -0.1664645680990583
```

Cette fois-ci, les résultats (table 43) étaient tout aussi impressionnants, sur 129 enregistrements à prédire, voici les résultats :

**Table 43 - Résultats obtenus sur le 3<sup>ième</sup> jeu d'essai**

Target	Results	Delta
1	1.00004	0%
1	1.00014	0%
1	0.999718	0%
1	1.00038	0%
1	0.999903	0%
1	0.999891	0%
1	0.999757	0%
1	0.999984	0%
1	0.99998	0%
1	0.999979	0%
1	1.00003	0%
1	0.99959	0%
1	1.00001	0%
1	0.999685	0%
1	0.999934	0%
1	1.00018	0%
1	1.00006	0%
1	1.00001	0%
1	1.0001	0%
1	0.999981	0%



Target	Results	Delta
1	1.00018	0%
1	0.999986	0%
1	1.00011	0%
1	0.99984	0%
1	1.00011	0%
1	1.00003	0%
1	1	0%
1	0.999948	0%
1	0.999747	0%
1	1.00045	0%
1	1.00025	0%
1	1.00046	0%
1	0.999935	0%
1	0.999799	0%
1	1.00007	0%
1	0.999911	0%
1	1.00015	0%
1	0.99988	0%
1	0.999973	0%
1	0.999882	0%
1	1.00001	0%
1	0.999969	0%
1	1	0%
1	1.00017	0%
1	1.00002	0%
1	1.0002	0%
1	0.999708	0%
1	0.99985	0%
1	0.999822	0%
1	0.999873	0%
1	1.00002	0%
1	0.999587	0%
1	1.00003	0%
1	1.00005	0%
1	0.999952	0%
1	0.999932	0%
1	0.999972	0%
1	1.00045	0%
1	0.999906	0%
1	1.00043	0%
1	1.00008	0%
1	0.999992	0%
1	1.00007	0%
1	1.00001	0%
1	0.999799	0%
2	2.00053	0%
2	2.00003	0%

Target	Results	Delta
2	2.00015	0%
2	2.00002	0%
2	2.00015	0%
2	1.99998	0%
2	1.99994	0%
2	2.00054	0%
2	2.00042	0%
2	1.99995	0%
2	1.99981	0%
2	2.00003	0%
2	1.9999	0%
2	2.00006	0%
2	2.00049	0%
2	1.99967	0%
2	2.00008	0%
2	2.00004	0%
2	2.0005	0%
2	2.00004	0%
2	2.00026	0%
2	1.9999	0%
2	2.00024	0%
2	2.00013	0%
2	1.99991	0%
2	2.00052	0%
2	2.00009	0%
2	2.00001	0%
2	2.00018	0%
2	2.00001	0%
2	1.99998	0%
2	2.00056	0%
2	2.00013	0%
2	1.99987	0%
2	1.99993	0%
2	2.00006	0%
2	2.00003	0%
2	1.99986	0%
2	2.00002	0%
2	2.00041	0%
2	2.00014	0%
2	1.99988	0%
2	2.00045	0%
2	2.00021	0%
2	1.99996	0%
2	2.00007	0%
2	2.00007	0%
2	1.9999	0%
2	1.99996	0%

Target	Results	Delta
2	1.99988	0%
2	2.00003	0%
3	1.50005	150%
3	1.49974	150%
3	1.50001	150%
3	1.49985	150%
3	1.5	150%
3	1.4997	150%
3	1.49998	150%
3	1.49996	150%
3	1.50005	150%
3	1.50066	150%
3	1.49944	150%
3	1.49994	150%
3	1.49978	150%

Seuls ceux de la classe 3 ont obtenus des erreurs, mais si on observe bien, le SVM leur a quand même assigné à tous la même classe (ici, la valeur 1.5, au lieu de 3).

#### ***E.4 Quatrième jeu d'essai incluant LA PLUPART des caractéristiques choisies***

Un nouveau jeu a été créé, cette fois-ci, avec toutes les caractéristiques, pour tenter de séparer les événements prévus, des événements non-prévus de base ou non de base, Ces caractéristiques ont été utilisées:

Numéro, Normalized SN, Days between events, Political Belonging Normalized, Engine Time Since New, Engine TSN Recorded Date, Engine Cycles Since New, Time Between Overhaul, Time Since Overhaul, Cycles Since Overhaul, Normalized A/c SN, Opr clas normalized, Max proba owner and operator name, Event C 3 Rating, Engine age in days, Engine Age, Sand, Salt, Volcano, Pollution, Elevation, Crude Oil Price USD, S&P500, Corp Profits US, Interest Rates US, W GDP, USDCAD=X, Nb of Engines, MRT, TRT, AMU, Target(La classe cible: 1=Planned 2=BasicUnplanned 3=Non-BasicUnplanned).

Nous avons utilisé les paramètres « *DOT standard* » du SVM, sur cent exécutions (« *runs* »). Pour le jeu de validation, nous avons utilisé l'en-tête suivant:

```
# svm example set
@examples
dimension 31 (nombre de caractéristiques)
number 617 (nombre d'exemples, d'enregistrements)
b -1.051360415765215
```

Les résultats obtenus (table 44) ne sont pas vraiment satisfaisants (pour prédire les 154 enregistrements):

**Table 44 - Résultats obtenus sur le 4<sup>ème</sup> jeu d'essai**

Target	Results	delta
2	577.53	-575.53
1	522.577	-521.577
2	501.703	-499.703
3	408.068	-405.068
2	369.588	-367.588
1	336.833	-335.833
3	325.688	-322.688
2	317.089	-315.089
1	287.737	-286.737
2	258.047	-256.047
2	235.345	-233.345
1	228.459	-227.459
[...]		

### ***E.5 Cinquième jeu d'essai incluant CERTAINES des caractéristiques choisies***

Cette fois-ci, en utilisant encore les mêmes 31 caractéristiques, pour tenter de séparer le nombre de jours avant le prochain événement: 1 si le nombre de jours est inférieur à 365 et 2 s'il est supérieur. Nous avons utilisé les paramètres « *DOT standard* » du SVM, sur cent exécutions (« *runs* »). Pour le jeu de validation, nous avons utilisé l'en-tête suivant:

```
# svm example set
@examples
```

dimension 31  
number 514  
b -0.03091074823736911

Tous les enregistrements obtiennent le même score de 1.49416. L'exemple n'est donc pas concluant.

### ***E.6 Sixième jeu d'essai incluant CERTAINES des caractéristiques choisies***

Toujours dans la même direction, en réduisant le nombre de caractéristiques, en tentant de prédire le nombre de jours supérieurs à 365 (classe 2) ou inférieurs (classe 1). Le même en-tête de validation que l'essai 5 est conservé. En utilisant ces caractéristiques:

Days between events, Political Belonging Normalized, Engine Time Since New, Engine Cycles Since New, Time Between Overhaul, Time Since Overhaul, Cycles Since Overhaul, Opr clas normalized, Max proba owner and operator name, Event C 3 Rating, Engine age in days, Sand, Salt, Volcano, Pollution, Elevation, Crude Oil Price USD, S&P500, Corp Profits US, Interest Rates US, W GDP, USDCAD=X, Nb of Engines, MRT, TRT, AMU, Target.

Les résultats contiennent encore beaucoup d'erreurs:

-26.5895  
-179.956  
-52.8884  
-39.6383  
-38.0967  
[...]

### ***E.7 Septième jeu d'essai incluant CERTAINES des caractéristiques choisies***

Un autre jeu, se concentrant cette fois-ci seulement sur les types d'événements BUR et tenter de les séparer classe 1=  $x < 365$  ou classe 2  $x > 365$ . Le même en-tête de validation que l'essai 5 est conservé. En utilisant ces caractéristiques:

Days between events, Political Belonging Normalized, Engine Time Since New, Engine TSN Recorded Date, Engine Cycles Since New, Time Between Overhaul, Time Since Overhaul, Cycles Since Overhaul, Opr clas normalized, Max proba owner and operator name, Engine age in days, Sand, Salt, Volcano, Pollution, Elevation, Crude Oil Price USD, S&P500, Corp Profits US, Interest Rates US, W GDP, USDCAD=X, Nb of Engines, MRT, TRT, AMU, Target.

La table 45 nous montre que les erreurs diminuent comparativement au 6<sup>ième</sup> jeu d'essai:

**Table 45 - Résultats obtenus sur le 7ième jeu d'essai**

Target	Results	delta
1	1.92	-92%
1	1.90	-90%
1	1.90	-90%
1	1.89	-89%
1	1.89	-89%
1	1.88	-88%
1	1.88	-88%
1	1.87	-87%
1	1.86	-86%
1	1.85	-85%
1	1.85	-85%
1	1.81	-81%
1	1.79	-79%
1	1.77	-77%
1	1.72	-72%
1	1.66	-66%
1	1.65	-65%
1	1.54	-54%
1	1.47	-47%
1	1.42	-42%
1	1.35	-35%
2	2.19	-19%

Target	Results	delta
2	2.15	-15%
2	2.13	-13%
2	2.13	-13%
2	2.08	-8%
1	1.08	-8%
1	1.07	-7%
2	2.05	-5%
2	2.04	-4%
2	2.02	-2%
2	1.89	11%
2	1.88	12%
2	1.88	12%
2	1.87	13%
2	1.83	17%
2	1.73	27%
2	1.72	28%
2	1.69	31%
2	1.68	32%
2	1.66	34%
2	1.66	34%
2	1.66	34%
2	1.63	37%
2	1.61	39%
2	1.61	39%
2	1.56	44%
2	1.55	45%
2	1.54	46%
2	1.53	47%
2	1.51	49%
2	1.48	52%
2	1.47	53%
2	1.39	61%
2	1.31	69%
2	1.20	80%
2	1.13	87%

### ***E.8 Huitième jeu d'essai incluant CERTAINES des caractéristiques choisies***

Comme les jeux précédents n'étaient pas concluants, d'autres tests sont effectués pour tenter de déterminer quelles sont les caractéristiques aidant à séparer les types d'événements. Nous avons utilisé les caractéristiques suivantes :

Days between events, Political Belonging Normalized, Engine Time Since New, Engine Cycles Since New, Time Between Overhaul, Opr clas normalized, Event C 3 Rating, Engine age in days, Sand, Salt, Pollution, Elevation, Crude Oil Price USD, S&P500, Corp Profits US, Interest Rates US, W GDP, USDCAD=X, Nb of Engines, MRT, TRT, AMU, Target(La classe cible: 1=Planned 2=BasicUnplanned 3=Non-BasicUnplanned).

Nous avons utilisé les paramètres « *DOT standard* » du SVM, sur cent exécutions (« *runs* »). Pour le jeu de validation, nous avons utilisé l'en-tête suivant:

```
# svm example set
@examples
dimension 23
number 100
b -0.1671717133712994
```

Nous obtenons sensiblement les mêmes résultats que l'essai 2.

### ***E.9 Neuvième jeu d'essai incluant CERTAINES des caractéristiques choisies***

Encore un nouvel essai pour le nombre de jours. Voici les caractéristiques utilisées :

Days between events, Political Belonging Normalized, Engine Time Since New, Engine TSN Recorded Date, Engine Cycles Since New, Time Between Overhaul, Opr clas normalized, Max proba owner and operator name, Event C 3 Rating, Engine age in days, Sand, Salt, Volcano, Pollution, Elevation, Crude Oil Price USD, S&P500, Corp Profits US, Interest Rates US, W GDP, USDCAD=X, Nb of Engines, MRT, TRT, AMU, Target (1= moins de 1 an, 2=plus d'un an).

Nous avons utilisé les paramètres « *DOT standard* » du SVM, sur cent exécutions (« *runs* »). Pour le jeu de validation, nous avons utilisé l'en-tête suivant:



```
# svm example set
@examples
dimension 25
number 513
b -1.229252783600563e-016
```

Les résultats ne sont pas satisfaisants. Toutes les valeurs se voient attribuer la même valeur : 1.69786.

### ***E.10 Dixième jeu d'essai incluant CERTAINES des caractéristiques choisies***

Un dernier jeu d'essai pour trouver le nombre de jours. Voici les caractéristiques utilisées :

Days between events, Political Belonging Normalized, Engine Time Since New, Engine Cycles Since New, Time Between Overhaul, Time Since Overhaul, Cycles Since Overhaul, Opr clas normalized, Max proba owner and operator name, Event C 3 Rating, Engine age in days, Sand, Salt, Volcano, Pollution, Elevation, Crude Oil Price USD, S&P500, Corp Profits US, Interest Rates US, W GDP, USDCAD=X, Nb of Engines, MRT, TRT, AMU, Target(1=moins d'un mois, 2=plus d'un mois).

Nous avons utilisé les paramètres « *DOT standard* » du SVM, sur cent exécutions (« *runs* »). Pour le jeu de validation, nous avons utilisé l'en-tête suivant:

```
# svm example set
@examples
dimension 26
number 615
b 1.567236681050511
```

Encore une fois, les résultats ne sont pas concluants: tous les enregistrements se voient attribuer la valeur 1.48164.

## E.11 Les exemples relatifs au temps sur SPSS 14.0

Table 46 - « Cluster Membership » Exemple 1 relatif au temps

Case	20 Clusters	16 Clusters	12 Clusters	8 Clusters	4 Clusters	3 Clusters	2 Clusters
1:BPOS	1	1	1	1	1	1	1
2:BUOH	2	2	2	1	1	1	1
3:BUOZ	3	2	2	1	1	1	1
4:NURL	4	3	3	1	1	1	1
5:BPOH	5	1	1	1	1	1	1
6:BPOL	6	4	1	1	1	1	1
7:BPOZ	1	1	1	1	1	1	1
8:BPRH	7	5	4	2	1	1	1
9:BPRL	8	3	3	1	1	1	1
10:BPRS	9	6	2	1	1	1	1
11:BPRZ	10	7	5	2	1	1	1
12:BUOL	11	8	6	3	1	1	1
13:BUOS	12	9	7	4	2	2	2
14:BURH	13	10	5	2	1	1	1
15:BUURL	14	11	8	5	1	1	1
16:BURS	15	12	4	2	1	1	1
17:BURZ	2	2	2	1	1	1	1
18:NUOH	16	4	1	1	1	1	1
19:NUOL	17	13	9	6	3	1	1
20:NUOS	18	14	10	4	2	2	2
21:NUOZ	19	15	11	7	3	1	1
22:NURH	2	2	2	1	1	1	1
23:NURL	6	4	1	1	1	1	1
24:NURS	20	16	12	8	4	3	2

Considérant les groupes « temps » Z, L, S, H (respectivement: Zéro, « Low », « Std », « High »).

**Table 47 - « Cluster Membership » Exemple 2 relatif au temps**

Case	2 Clusters
1:BPOS	1
2:BUOH	1
3:BUOZ	1
4:NURL	1
5:BPOH	1
6:BPOL	1
7:BPOZ	1
8:BPRH	1
9:BPRL	1
10:BPRS	1
11:BPRZ	1
12:BUOL	1
13:BUOS	2
14:BURH	1
15:BURL	1
16:BURS	1
17:BURZ	1
18:NUOH	1
19:NUOL	1
20:NUOS	2
21:NUOZ	1
22:NURH	1
23:NURL	1
24:NURS	2

**Table 48 - « Cluster Membership » Exemple 3 relatif au temps**

Case	20 Clusters
1:BPOS	1
2:BUOH	2
3:BUOZ	3
4:NURL	4
5:BPOH	5
6:BPOZ	6
7:BPRH	7
8:BPRL	8
9:BPRS	9
10:BPRZ	10
11:BUOL	11
12:BUOS	12
13:BURH	13
14:BURS	14
15:NUOH	15
16:NUOL	16
17:NUOS	17

Case	20 Clusters
18:NUOZ	18
19:NURH	2
20:NURL	19
21:NURS	20

Table 49 - « Cluster Membership » Exemple 4 relatif au temps

Case	23 Clusters
1:BPOS	1
2:BUOH	2
3:BUOZ	3
4:NURL	4
5:BPOH	5
6:BPOL	6
7:BPOZ	7
8:BPRH	8
9:BPRL	9
10:BPRS	10
11:BPRZ	11
12:BUOL	12
13:BUOS	13
14:BURH	14
15:BURL	15
16:BURS	16
17:BURZ	17
18:NUOH	18
19:NUOL	19
20:NUOS	20
21:NUOZ	21
22:NURH	17
23:NURL	22
24:NURS	23

Table 50 - « Cluster Membership » Exemple 5 relatif au temps

Case	24 Clusters	20 Clusters	16 Clusters	12 Clusters	8 Clusters	4 Clusters	3 Clusters	2 Clusters
1:BPOZ	1	1	1	1	1	1	1	1
2:BPOZ	2	2	1	1	1	1	1	1
3:BPOZ	2	2	1	1	1	1	1	1
4:BPOZ	3	1	1	1	1	1	1	1
5:BPOZ	3	1	1	1	1	1	1	1
19:BUOZ	4	3	2	2	2	1	1	1
20:BUOZ	4	3	2	2	2	1	1	1
21:BUOZ	5	4	3	2	2	1	1	1
22:BUOZ	4	3	2	2	2	1	1	1
23:BURS	1	1	1	1	1	1	1	1

Case	24 Clusters	20 Clusters	16 Clusters	12 Clusters	8 Clusters	4 Clusters	3 Clusters	2 Clusters
24:BUOS	4	3	2	2	2	1	1	1
92:NURS	17	14	9	8	5	1	1	1
93:BURS	18	15	13	10	7	3	1	1
94:NUOS	16	13	12	9	6	1	1	1
95:NURS	19	16	14	11	8	4	3	2
96:BURS	14	11	10	3	3	2	2	1
97:NURH	16	13	12	9	6	1	1	1
98:BPOH	20	17	15	12	6	1	1	1
99:BPOH	20	17	15	12	6	1	1	1
100:BURH	21	18	11	2	2	1	1	1
101:BUOH	5	4	3	2	2	1	1	1
102:BPRH	22	19	16	10	7	3	1	1
103:NUOH	21	18	11	2	2	1	1	1
104:BPOH	20	17	15	12	6	1	1	1
112:BPRH	24	20	16	10	7	3	1	1
113:BPOH	20	17	15	12	6	1	1	1
114:BUOH	24	20	16	10	7	3	1	1
115:BPRH	22	19	16	10	7	3	1	1
116:BPRH	22	19	16	10	7	3	1	1

## Annexe F - Un exemple sur un modèle spécifique.

Intrants du modèle XX300Y (305, 306, 307 & 308):

### F.1. Service Bulletins

Utilisation: Intervalles de Maintenance

Données pertinentes: les intervalles des O/H (« *Overhauls* ») et des HSI (« *Hot Section Inspection* »)

Voir en annexe A.1, p. 74, un exemple de Bulletin.

Il n'y a pas de Service Bulletin pour ce modèle spécifique, sauf, bien entendu, les définitions des intervalles de maintenances, communes à tous les modèles.

<b>N°</b>	<b>AREA</b>	<b>SubArea</b>	<b>Inspection-Task</b>	<b>Reference</b>
1	Overhaul	Engine	Dissassemble, clean and do a full inspection of the detail parts. Do the applicable repairs and reassemble the engine. thru 72-00-53 to separate modules 72-11-00 thru 72-53-00 to do the overhaul	72-00-11
2	Overhaul	Reduction Gearbox	Dissassemble, clean and do a full inspection of the detail parts. Do the applicable repairs and reassemble the engine. to remove gearbox 72-11-00 to do the overhaul	72-00-11
3	Heavy Maintenance		Hot Section Dissassemble the engine rear end up to the HP stubshaft. Do a full inspection on all parts and repair damaged parts. Reassemble the rear end of the engine. Inspection/Check-04 (Task 72-00-00-280-811)	Inspection 72-00-00
4	Heavy Maintenance		Specific Repair affected module, disassemble to get access to the detail parts affected and do the applicable inspections and repairs. Reassemble and install the module. thru 72-00-53 for removal/installation of affected parts, inspections, repairs and reassembly	Remove 72-00-11
5	Limited-life Parts	Limited-life Parts	Dissassemble the engine sufficiently to get access and replace the time-expired components. 72-00-00 Inspection/Check-06 (Task 72-00-00-280-830)	
6	Reduction Gearbox		Refurbish module 72-00-11 to remove module 72-11-00 to refurbish module.	If module time since refurbishment is > soft-time

- 7 Reduction Gearbox If module time since refurbishment is < soft-time  
Return to service. But, if the module is defective, partially disassemble module to get  
access to damaged parts and repair affected parts. 72-00-11  
to remove module 72-11-00 to disassemble and repair affected parts.
- 8 Turbomachinery Accessory Gearbox (FIC, RIC) If module  
time since refurbishment is more than soft-time, refurbish module. If module time since  
refurbishment is less than soft-time, return to service. But, if the module is defective, partially  
disassemble it to get access to damaged parts and repair 72-00-  
21/72-00-22 to remove front/rear inlet case 72-21-00 or 72-22-00 to refurbish module or to  
disassemble and repair affected parts.
- 9 Turbomachinery LP and HP Compressors (if MTSR > S-t) Refurbish  
module. But if the module is defective, partially disassemble module to get access to damaged  
parts and repair. If limited-life parts need replacement, partially disassemble module to get  
access, and replace parts that don't have enough life. 72-00-31  
to remove module 72-31-00 or 72-41-00 to refurbish module or to disassemble and repair  
affected parts or to replace limited-life parts.
- 10 Turbomachinery LP and HP Compressors (if MTSR < S-t) Return to  
service. But if the module is defective, partially disassemble module to get access to damaged  
parts and repair. If limited-life parts need replacement, partially disassemble module to get  
access, and replace parts that don't have enough life. 72-00-31  
to remove module 72-31-00 or 72-41-00 to refurbish module or to disassemble and repair  
affected parts or to replace limited-life parts.
- 11 Turbomachinery Hot Section (if MTSR > S-t: Refurbish, RTS) But if the  
module is defective, partially disassemble module to get access to damaged parts and repair.  
If limited-life parts need replacement, partially disassemble module to get access, and replace  
parts that don't have enough remaining life. 72-00-41/72-00-51/72-00-52 to remove  
combustion section/HP tub/LP turb 72-41-00/72-51-00/72-52-00 for CS/HP/LP repairs
- 12 Turbomachinery Power Turbine (if MTSR > S-t: Refurbish, RTS) But if the  
module is defective, partially disassemble module to get access to damaged parts and repair.  
If limited-life parts need replacement, partially disassemble module to get access, and replace  
parts that don't have enough remaining life. 72-00-53 to remove/install PT Shaft and turbines  
72-53-00 to refurbish module or to disassemble and repair affected parts or to replace limited  
life parts.

Ici, le modèle 306B est dans la catégorie des Turbo Fans (famille: XX300).  
La table 51 nous renseigne sur la situation de cette famille par rapport aux autres familles.

**Table 51 - Familles de moteurs**

	ENGINES CERTIFIED	AIRCRAFT CERTIFIED	OPERATORS	AIRCRAFT FLYING POPULATION	ENGINES PRODUCED	OPERATING TIME - HOURS
XX6	88	133	6580	15367	43438	325940274
XX100	30	15	358	2206	5773	105124714
XX200	10	6	305	647	1797	2033064
XX15D	13	16	1787	2805	6535	38554824
XX300	9	6	457	773	2033	5481296
XX500	5	3	571	922	2242	4069974
XXU	3	2	40	609	1147	20157931
XX600	2	2	n/a	n/a	98	n/a
<b>TOTAL</b>	<b>160</b>	<b>183</b>	<b>10098</b>	<b>23329</b>	<b>63063</b>	<b>501362077</b>

## **F.2. Maintenance Manuals**

Utilisation: Cycles de vie des pièces

Données pertinentes: le nombre d'heure limite de chaque pièce

Voir en annexe A.2, p. 75, un exemple de Manuel

**Table 52 - Manuel de maintenance avec chacune des caractéristiques du 306B**

N°	1
Engine	XX306B
RTBO	6000
RHSI	3000
Low Utilization	Inspection at 10yrs after EIS, 11yrs after manufacture
IBI	2400
SBI	1200
1st ST	2000
1st ST B I OP1 Pre-SB	1000 CSN HIS or OH
1st ST B S OP1Pre-SB	200
1st ST B I OP2 Pre-SB	1600 CSN HIS or OH



1st ST B S OP2 Pre-SB	400
1st ST TB post SB I	1600 CSN HIS or OH
1st ST TB post SB S	400
2nd ST TB post SB I	8500
HP TB & Vanes I	2400 FHSN HIS or OH
HP TB & Vanes S	1200
HP TL Seal I POST SB	1200 FH/CSN HIS or OH
HP TL Seal I PRE SB	1200 FH/CSN HIS or OH
HP TL Seal I PRE SB HARD TIME	300
HP TL Seal S	300FH/CSN
LP Turbines BII	2400FHSN HIS or OH
LP Turbines BSI	1200
Exhaust Mixer Initial	2000FHSN HIS or OH
Exhaust Mixer Subsequent	2000
Fuel filter Discard FH I&S	1200
External fuel tubes & fuel system cpnts FH I&S	2000
Emergency FSOV linkage mechanism FH I&S	4000
Emergency FSOV linkage mechanism FH I&S Hard Time	3000
Drain Mast FH I&S	2000
Primary Fuel Nozzles FH I&S	3000
Igniters FH	1200
Operational check for cont ignition	400
Check P1 Air tube I&S	400
Check P3 Air tube I&S Pre-SB	60
Check P3 Air tube I&S Post-SB	400
Bleed valves p25 & p28	5000FH or HIS
Bleed valves p25 & p28 Hard Time	3000FH or HIS
BOV Solenoids interval PRE SB	400-600
Starting BOV S Post SB	1400-1600
Handling BOV S Post SB	1000
BOV S Valve Cartridge screen Post SV Op1	1200
BOV S Valve Cartridge screen Post SV Op2	4000
Check P1/T1 Sensors, N1, N2 speed sensors, T45 probes & vibratio	800
ITT/T45 system Func check	1200
Fuel oil heat exchanger Pre SB Visual check	60
No3 Brg Scavenage oil tube visual check	1200
Chip detector opr check	2000
AGB Overpressure pop-up ind	40
Oil consumption monitoring and oil level check	40
Oil filter	1200
External oil tubes Incl no4 brg tubes, AGB cover and seals	800

### **F.3. Shop Visit Historical Data (TRAC, TRS)**

Utilisation: événements de type NUR, BPR, NPR (fréquence d'apparition)

Données pertinentes: [sur 6 mois] le nombre d'événements d'un type et le total des moyennes mensuelles d'utilisation (AMU).

Voici les données à obtenir de cette source :

Sur un total de 2410 événements pour la famille des XX300, 771 se rapportent au XX306B. (32%). L'allure des données est présentée à la table 53. :

**Table 53 - Historique d'un moteur de type 306B**

Module	ENGINE
Event Date	1-Jul-99
Political Belonging	[Austria = 7, Canada = 3, China = 285, French Polynesia = 4, Germany = 6, Italy = 41, Nigeria = 27, USA = 397, UK = 1]
Engine Time Since New	0
Engine TSN Recorded Date	1-Jul-99
Engine Cycles Since New	0
Engine CSN Recorded Date	1-Jul-99
Time Between Overhaul	0
Time Since Overhaul	0
Cycles Since Overhaul	0
Engine Type	Turbo Fan
Engine Model	XX306B
Aircraft Serial Number	DOR####
Engine Baseline TBO	9999
Engine Baseline HSI	9999
Operator Name	[ADAC LUFTRETTUNG GMBH AERO-DIENST GMBH AIR VALLEE S.P.A. ALTRIA GROUP, INC. ATLANTIC COAST AIRLINES AVCRAFT AVIATION LLC BRISTOW HELICOPTERS GROUP LTD. CHANG AN AIRLINES CO., LTD CIRRUS AIRLINES FAIRCHILD DORNIER GMBH GANDALF AIRLINES GREAT PLAINS AIRLINES HAINAN AIRLINES CO LTD ITALI AIRLINES SRL JOHNSON CONTROLS INC. OZARK MANAGEMENT INC. CANADA LEASING INC. SHELL PETROLEUM DEVELOPMENT CO. NIGERIA SKYWAY AIRLINES TYROL AIR AMBULANCE TYROLEAN AIRWAYS GMBH TYROLEAN JET SERVICE GMBH ULTIMATE JET CHARTER UNION PLANTERS NATIONAL BANK UNKNOWN WANAIR ]
Continent	[Africa = 27, Americas = 400, Asia = 285, Australasia = 4, Europe = 55]
Subcontinent	[Australasia = 4, North America = 400, Northern Africa = 27, Northern Europe = 7, Southern Asia = 285, Southern Europe

	= 48]
Region	[Canada = 3, Central Africa = 27, Central Asia = 285, Euronorth = 7, eurosouth = 48, South Pacific = 4, USA = 397]
Country	[Austria = 7, Canada = 3, China = 285, French Polynesia = 4, Germany = 6, Italy = 41, Nigeria = 27, USA = 397, UK = 1]
State/Province	[Bavaria = 3, Bergamo = 8, England = 1, Hainan Dao = 73, Illinois = 1, Kentucky = 1, Lagos = 3, Massachusetts = 4, Michigan = 4, Missouri = 7, New York = 1, North Carolina = 1, Ohio = 5, Oklahoma = 8, Pescara = 2, Quebec = 2, Saarland = 1, Shaanxi = 209, South Carolina = 19, Tahiti Archipelago = 4, Tennessee = 3, Tirol = 3, UNKNOWN = 68, Virginia = 230, Wisconsin = 110 over TOTAL = 771]
Event Category	[Basic = 716, Non-basic = 55]
Event Type	[Planned = 437, Unplanned = 334]
Engine Disposition	[Overhauled = 8, Remove to be repaired = 763]
Event Classification 5	[BPONC = 5, BPONS = 1, BPRN+ = 5, BPRNC = 262, BPRNS = 163, BPRNU = 1, BUONC = 2, BURAC = 14, BURAS = 11, BURCC = 1, BURIC = 7, BURIS = 5, BURNC = 179, BURNS = 56, BURNU = 3, BURPC = 1, NURAC = 5, NURAS = 3, NURAU = 1, NURDC = 1, NURIC = 1, NURIS = 1, NURNC = 30, NURNS = 12, NURNU = 1]
Event Classification 5 - Condition	[ACCEPTABLE/NOTHING FOUND = 11, BORESCOPE INSPECTION = 6, BURNED THRU OR OFF = 3, CRACKED = 1, CYCLE EXPIRATION = 1, ENGINE REMOVED, PLANNED MAINTENANCE = 3, ENGINE/MODULE REMOVED-FOR REPAIR = 141, ERROR, OPERATOR = 1, FOD-BIRD INGESTION, ONE ENGINE = 2, FOD-EXTERNAL OTHER OBJECTS = 1, HOT SECTION INSPECTION = 453, IMPROPER MAINTENANCE / PRESEVRATION = 1, INSTALLED-PART, MODULE, COMPONENT = 1, MAINTENANCE-FAULTY/CONTRARY TO PROCEDURE = 2, MAINTENANCE-PRECAUTIONARY = 1, MAJOR REFURBISHMENT = 13, MANUFACTURING ERROR = 3, METAL IN TAILPIPE = 1, NOISE-EXTERNAL, ENGINE = 1, OPERATEUR/FBO MAINTENANCE FAULTY = 2, OVERHAUL VS REPAIR = 8, OVERTEMP = 1, PART REMOVED-INVESTIGATION = 3, PLANNED (CONVINIENCE) REMOVAL-MAINT. REQ'D = 4, PLANNED CONVINIENCE REMOVAL-NO MAINT. REQ'D = 3, POWER LOSS/ROLLBACK = 1, SERVICE BULLETIN INCORPORATION = 71, SURGE, COMPR-ACCEL (NONRECOVERABLE) = 1, TEST PERFORMANCE = 1, TORQUE OIL PRESSURE ERRATIC = 1, TROUBLE SHOOTING-REMOVED IN = 4, TURBINE BLADE DISTRESS = 4, VIBRATION-EXCESSIVE-ENGINE = 2]
Zone/System - Condition	[ACCESSORY GEARBOX = 2, ACCESSORY UNITS = 4, AIR INLET SECTION = 1, AIR SYSTEM = 7, AIRFRAME RELATED = 4, BYPASS DUCTS = 1, COMBUSTION CHAMBER = 4, COMBUSTION SECTION = 1, COMPRESSOR ROTOR BALANCING ASSY = 5, COMPRESSOR SECTION = 22, COMPRESSOR TURBINE = 7, ELECTRICAL SYSTEM = 1, ENG. FUEL & CONTROL SYSTEMS = 1, ENGINE = 351, ENGINE SECTION = 1, EXHAUST DUCT ASSY = 1, FUEL SYSTEM = 2, GAS GENERATOR = 1, HIGH TURBINE = 346, OIL SECTIONS =

	1, OIL SYSTEM = 4, TURBOMACHINE MODULE = 3]
Engine Comment	
Engine Age	4.77
Module Age	[01DEC98, 01OCT02]
Engine Target BUR	0.1
Engine Target IFSD	0.005

**Table 54 - Détails à considérer**

Tables to consider:				
	WATOG		%age of shop visits that are:	# of shop visits
<b>Planned (Scheduled and Convenience)</b>				
Overhauls	BPO	842, 844	TO FIND	TO FIND
HSIs	BPO	842, 844	TO FIND	TO FIND
Repairs	BPR	143	TO FIND	TO FIND
<b>Unplanned</b>				
Overhauls	BUO	132	TO FIND	TO FIND
HSIs	BUO or BUR	842, 844, 143	TO FIND	TO FIND
Repairs	BUR	143, 509	TO FIND	TO FIND
<b>Convenience</b>				
Overhauls	NPO	842, 844	TO FIND	TO FIND
HSIs	NPO or NUR	143	TO FIND	TO FIND
Repairs	NPR	143	TO FIND	TO FIND

**Other:**

Les données se présentent de 2 façons:

a) Une liste exportable sur Excel (Provenant de TRAC-TRS) (SAS)

Figure 67 - Liste exportable de TRS-TRAC (SAS)

On peut ensuite faire les moyennes des « reliability rates »

	A	C	D	F	H	J	L	N	P	R	T
1	Removal Rates - From October 12, 2002 Until Today										
2	306B										
3	Engine Section - Military Included - Current Data ( October 11, 2007 )										
4	Rates: NIF BIF TIF NUR BUR BPR TSV BSV TUR - Rolling: 6 month(s)										
5											
6											
7	MONTH	NIF	NIF Rate	BIF Rate	TIF Rate	NUR / NU	BUR / BU	BPR Rate	TSV Rate	BSV	TUR Rate
68	6 last mo	Average	0.007	0.0095	0.016667	0.011667	0.046833	0.1375	0.1965	0.184167	0.058667
69	6 mo bef	Average	0	0.011833	0.011833	0.036	0.087667	0.121333	0.244833	0.209167	0.1235
70	6 mo bef2	Average	0	0	0	0.0085	0.043333	0.1265	0.1785	0.169667	0.052167
71	6 mo agair	Average	0.004667	0	0.004667	0.012833	0.0355	0.280833	0.3295	0.3165	0.048333
72											
73											

Figure 68 - Moyennes des taux

On utilise des tranches de 6 mois, on ne prend jamais les 3 derniers mois, car certains événements ne sont pas encore définitivement classés.

b) On peut également obtenir des taux en ligne, en utilisant TRS (TRAC Reporting System):

	305A			305B		
BUR	0.100	0.053	→	0.100	0.000	→
Basic IFSD	0.005	0.015	→	0.005	0.000	→
	306A			306B		
BUR	0.100	0.107	→	0.100	0.043	→
Basic IFSD	0.005	0.027	→	0.005	0.000	→

Figure 69 - Taux en ligne

#### **F.4. Warranty Data**

Utilisation: Phase 2 (Parts & Planning)

Données pertinentes: Aucune

#### **F.5. Part Sales Data**

Utilisation: Phase 2 (Parts & Planning)

Données pertinentes: Aucune

#### **F.6. Modèles de prévision de chaque département**

Utilisation: Les notes pertinentes s'appliquant à la prévision pour ce modèle

Données pertinentes: Il existe 2 documents.

1-Engine Configuration New (voir la figure 70)



## 2-Forecast version x 2006-05-03 (voir la figure 71):

The screenshot displays an Excel spreadsheet with a forecast application interface. The interface is organized into several functional areas:

- LOAD Customer Input:** Includes an "Input Data from customer" field and an "Input" button.
- FORECAST:** Contains instructions to enter the "number of Year" and load the "Input Sheet". It features a "Forecast (#year)" field with the value "5" and a "Forecast" button. Below it is a "BUR Rate" field with the value ".039" and a "Per 1000 Hours" label. There is also a "LEASCO (Included)" checkbox.
- Configuration Table:** Lists various configuration options with corresponding buttons: "Change LCF limit and disposition" (LCF Limits), "Operator general information" (Operator), "Workscope Definition" (Workscope), "Campaign Definition" (Campaign), and "Shop Visit Cost" (SIC Matrix).
- Campaign Review:** Includes a "Campaign Status" field and a "Campaign" button.
- Leasco Requirements:** Features a "Review Leasco requirements" field and a "Leasco" button.
- Service Center:** Contains a "Shop Workscope Decision" field and a "WS Decision" button.

The spreadsheet background shows columns A through O and rows 1 through 44. The status bar at the bottom indicates "contiennent cette information AUTRES VARIABLES À CONSIDÉRER".

Figure 71 - Forecast version x 2006-05-03

### F.7. Notes des administrateurs de système

Utilisation: Aucun administrateur système.

Données pertinentes: Aucune à ce modèle.

### F.8. Les facteurs macro-économiques

Utilisation: Les règles seules qui affectent ce modèle

Données pertinentes:

Une base de données expérimentale est en création et des fichiers MS Excel contiennent cette information.

**AUTRES VARIABLES À CONSIDÉRER**



**Les facteurs environnementaux – la liste utilisée est une mise à jour des données environnementales utilisée pour les XX6/XX6A (hélicos).**

Le type de mission du détenteur: à déterminer pour tous les détenteurs (457) des 773 avions en fonction.

Si le détenteur détient une extension de l'intervalle d'O/H? Entrée manuelle pour chaque?

## Annexe G – Technique statistique pour l'estimation de distributions.

On pose les variables suivantes:  $L$ ,  $i$ ,  $n$ ,  $x$  et  $\lambda$ .

$F(x)$  dénote la fonction de distribution.

$$L = \prod_{i=1}^n f(x_i) \quad \begin{array}{l} x_i = \text{Les valeurs de l'échantillon} \\ f = \text{La fonction de distribution} \end{array}$$

Prenons le cas d'une distribution de forme exponentielle.

Sa fonction de distribution sera la suivante:  $f(x) = \lambda e^{-\lambda x}$

Le système d'équation prendra la forme suivante:

$$\Lambda = \ln L = \ln \prod_{i=1}^n f(x_i) = \sum_{i=1}^n \ln f(x_i).$$

Considérant  $\max L \Leftrightarrow \max \Lambda$

On doit résoudre  $\frac{\partial \Lambda}{\partial \lambda} = 0$  pour  $\lambda$ .

Dans le cas d'une distribution normale:  $\frac{\partial \Lambda}{\partial \mu} = 0; \frac{\partial \Lambda}{\partial \sigma} = 0$

Prenons un exemple avec ce mini-échantillon  $\{0.2; 0.35; 0.51\}$

$$\Lambda = \ln f(0.2) + \ln f(0.35) + \ln f(0.51)$$

$$\ln f(x) = \ln(\lambda \cdot e^{-\lambda x}) = \ln \lambda + \ln e^{-\lambda x} = \ln \lambda + (-\lambda x)$$

Sachant que  $\ln e^a = a$

$$\ln f(x) = \ln \lambda - \lambda x$$

$$\Lambda = \ln \lambda - 0.2\lambda + \ln \lambda - 0.35\lambda + \ln \lambda - 0.51\lambda$$

Ainsi,  $\Lambda = 3 \ln \lambda - 1.06\lambda$

$$\frac{\partial \Lambda}{\partial \lambda} = \frac{3}{\lambda} - 1.06 = 0$$

$$\text{Et } \frac{3}{\lambda} = 1.06 \text{ donc } \lambda = \frac{3}{1.06} = 2.83.$$

Test avec une Distribution Normale :

$F(x)$  dénote la fonction de distribution.

$$f(x) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$$

Essayons de déduire le système d'équations :

$$\Lambda = \ln\left(\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\right) = \frac{\ln e^{-\frac{x^2}{2}}}{\ln \sqrt{2\pi}} = \frac{-\frac{x^2}{2}}{\ln \sqrt{2\pi}} = \left(\frac{-x^2}{2} \cdot \frac{1}{\ln \sqrt{2\pi}}\right) = \frac{1.837877x^2}{-3.6757541}$$

## Annexe H – Notes de la rencontre avec le groupe SAS et la solution proposée (Anglais)

SAS B.I. → Using Business Warehouse

Create report → Analyze Report → Deploy



Supply chain process

It brings efficiency, but your DATA must be **accurate** at the first place.

SAS can use flat files or get the information directly from SAP.

At Pratt, we have slow movers, not like one-week turn around time.

How to measure accuracy?

SAS do a great job, better than SAP APO?

Create time series of all events (key events), based on where those events are in this time series.

Forecasting the rates.

A lot of things seem random.

Forecast based on historical data.

Bring it in house and do the forecast.

There is a real need for prioritize.

Everything is in the calibration of the system.

It has a specific number of tools (looking at a time frame), some may be left behind.

After 1 cycle : re-calibrate.

Try to find out what affect, all factors that have an impact.

Ex.: gas price.

Any external factor.

⇒ Determine these factors.

Stuff not explained by major factors.

Then Highlight it.

Independent variables that affects the forecast.

Ex.: Titanium fluctuation price.

No need to create ABAP to access data.

- Some companies have a forecast analyst.
- Modifying the models?
- Override the forecast?
- Put into SAP?
- Then start a collaborative plan, using different inputs from:
  - Sales
  - Parts
  - ...

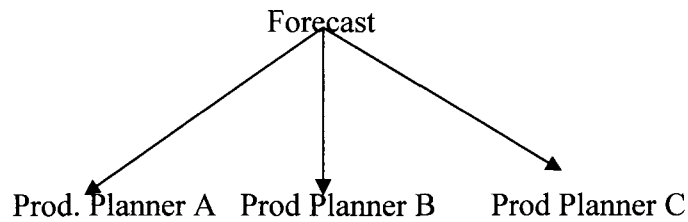
Come up with a business model and then analyze it using SMP, PPDS.

Build the capability to verify the accuracy.

Make sure the inputs are accurate.

It builds a cycle.

Kind of re-engineering.



What does the market?

BICC : competency centre, for large SAP consumers.

Other than manpower, It's for addressing your need for forecast.

The tools are scalable to enterprise level.

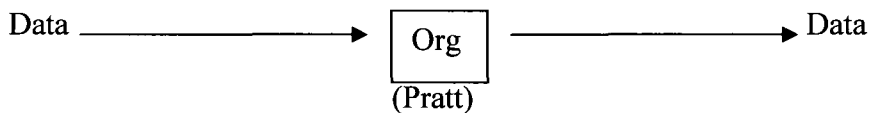
Canada Post, Canada Blood association, are using it.

We have ERP.

Next level : optimizing.

DM: - Segment customers  
 What we do with the data, with the analysis.

Workshop => Centre of excellence (1 hour)



Ex.: Government => Aging problem of population

It decreases over years; they will miss talent, qualified resources.

Variance : Ex.: Sobey's  
 Profitability

BICC → How to use it  
 It's all about this.

Flexibility

Ex.: bring a new variable?

Do we need to create a business case for that little change?

Able to create « *on the fly* ».

6 Steps process, create a project, save it, run it again.

Detect one-time events vs. cycles.

SAS website → Solutions → forecasting → success stories.

Nov 29<sup>th</sup> (webcast):

Presentation of the paper + demo. Q&A.

BICC book.

---

DEMO

SAS 9.

Data preparation = ERP pool.

Step 1

Name the project.

Step 2

Select data set(s)  
Ex.: Sales info (historical)

Step 3

Assign data roles

- a) Time period
- b) BV variables:      parts  
                                 City  
                                 Site #  
                                 Customer #
- c) Dependant variables
- d) Independent variables:      Demand  
   Price
- e) Reporting variables

Step 4

Configure hierarchy  
Project: forecast by plant, by city, if not there, just call the data.\

Step 5

Elements related to stats (forecast options)

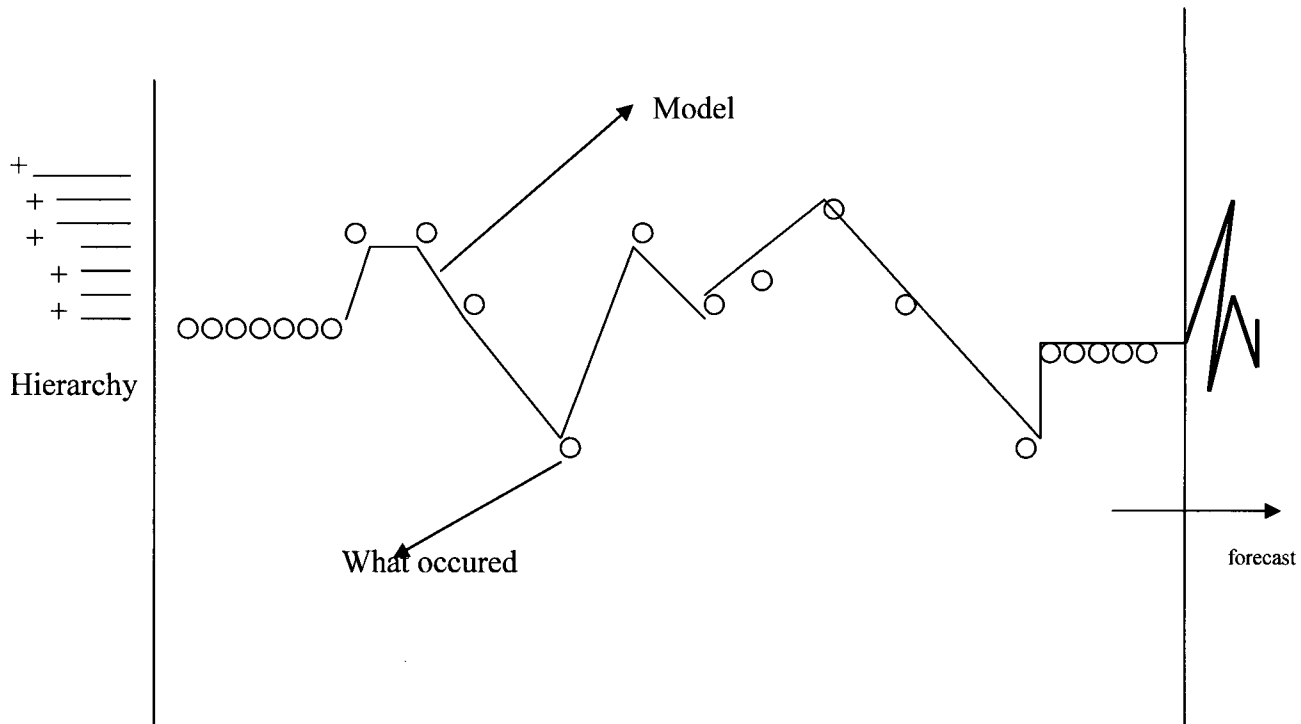
What to do with missing info (take the lowest, highest value? The average?)

To feed the holes in the data.

Incorporate the pool.

Step 6

Done. Look at the results.





## Bibliographie

- [1] LAROSE, DANIEL T., *"DES DONNEES A LA CONNAISSANCE Une introduction au data mining"*, 224p, 2005, Vuibert Informatique.
- [2] KATERINA TAŠKOVA, PANČE PANOVA, ANDREJ KOBLER, SAŠO DŽEROSKI, DANIELA STOJANOVA *"Predicting Forest Stand Properties from Satellite Images with Different Data Mining Techniques"*, 2006, Ljubljana, Slovenia.
- [3] JING PENG, DOUGLAS R. HEISTERKAMP, H. K. DAI *"LDA/SVM Driven Nearest Neighbor Classification"* cvpr, vol. 1, pp.58, 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'01) - Volume 1.
- [4] GE, E., NAYAK, R., XU, Y. AND LI, Y. *"Data Mining For Lifetime Prediction of Metallic Components"*, 2006 In Proc. Fifth Australasian Data Mining Conference (AusDM2006), Sydney, Australia. CRPIT, 61. Peter, C., Kennedy, P. J., Li, J., Simoff, S. J. and Williams, G. J., Eds. ACS. 75-81.
- [5] JORGE BERNAL-CHAVES, CARMEN PELÁEZ-MORENO, ASCENSIÓN GALLARDO-ANTOLÍN, & FERNANDO DÍAZ-DE-MARÍA, *"Multiclass SVM-Based Isolated-Digit Recognition using a HMM-Guided Segmentation"*, In ISCA Tutorial and Research Workshop on Non-Linear Speech Processing 2005 (NOLISP 05).
- [6] JIANG ZHI-QIANG, FU HAN-GUANG, LI LING-JUN, *"Support Vector Machine for mechanical faults classification"*, Journal of Zhejiang University SCIENCE A 2005, Vol. 6; No. 5 p. 433-439.
- [7] TAO SHAO-HUI, CHEN DE-ZHAO, HU WANG-MING, *"SVD-LSSVM and its application in chemical pattern classification"*, Journal of Zhejiang University SCIENCE A 2006, Vol. 7 No. 11 p. 1942-1947.
- [8] LONG-HUI WANG JUAN LIU, YAN-FU LI & HUAI-BEI ZHOU, *"Predicting Protein Secondary Structure by a Support Vector Machine Based on a New Coding Scheme"*, Genome Inform Ser Journal 2004, Vol. 15 No. 2 P.181-190.
- [9] Le Blogue DATAINTELLIGENCE 2008,  
<http://datalligence.blogspot.com/2008/11/fraud-prediction-decision-trees-svm.html> .
- [10] H. BLOCQUEEL, L. SCHIETGAT, J. STRUYF, A. CLARE, EN S. DZEROSKI, *"Hierarchical multilabel classification trees for gene function"*

- prediction*”, 2006 Probabilistic Modeling and Machine Learning in Structural and Systems Biology (Rousu, J. and Kaski, S. and Ukkonen, E., eds.), pp. 9-14.
- [11] H WILLENBROCK, A S JUNCKER, K SCHMIEGELOW, S KNUDSEN & L P RYDER, "*Prediction of immunophenotype, treatment response, and relapse in childhood acute lymphoblastic leukemia using DNA microarrays*", Journal Leukemia 18, 2004.
- [12] MOHAMADALLY, H. AND FOMANI, B. "*svm : Machines à vecteurs de support ou séparateurs à vastes marges*" BD Web 2006, ISTY3, Versailles St Quentin, France.
- [13] C.W. HSU, C.C. CHANG, AND C.J. LIN "*A practical guide to support vector machine,*" Available: <http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>.
- [14] Support Vector.net [www.support-vector.net/software.html](http://www.support-vector.net/software.html) .
- [15] Data Warehousing Review  
[http://www.dwreview.com/Data\\_mining/Closed\\_loop\\_DM.html](http://www.dwreview.com/Data_mining/Closed_loop_DM.html) .
- [16] Accentiv' Surfsgold [http://www.surfsgold.com/hub\\_articles\\_02dmstages.htm](http://www.surfsgold.com/hub_articles_02dmstages.htm) .
- [17] Overview of the KDD Process  
[http://www2.cs.uregina.ca/~dbd/cs831/notes/kdd/1\\_kdd.html](http://www2.cs.uregina.ca/~dbd/cs831/notes/kdd/1_kdd.html) .
- [18] ER MAMTA JUNEJA & ER NITIKA PHULL "Data Mining and its Scope"  
[http://www.pcc.qub.ac.uk/tec/courses/datamining/stu\\_notes/dm\\_book\\_2.html](http://www.pcc.qub.ac.uk/tec/courses/datamining/stu_notes/dm_book_2.html) .
- [19] Web Content Mining Tutorial  
<http://www.cs.uic.edu/~liub/WebContentMining.html> .
- [20] NEEL SUNDARESAN & JEONGHEE YI, "*Mining the Web for Relations*" Proceedings of the Third International Joint Conference on Natural Language Processing 2008.
- [21] FAN ZHAO, "*A Review of web Mining Research*" 2004, University of Nebraska, Lincoln.
- [22] CHAKRABARTI & RAMAKRISHNAN, "*Mining the web*", 2006,  
<http://www.cse.iitb.ac.in> .
- [23] BILL PALACE, "*Data Mining, Technology Note prepared for Management 274A*", Anderson Graduate School of Management at UCLA, June 1996.
- [24] Dr. KAMRAN PARASAYE, "*Data mining*" 1995 Hermosa Beach, CA.

- [25] CHRIS ERICKSON, "Quote Of The Week" HPCwire Oct. 13, 1995: Vol. 14, No. 41.
- [26] Wikipedia [http://fr.wikipedia.org/wiki/Apprentissage\\_automatique](http://fr.wikipedia.org/wiki/Apprentissage_automatique) .
- [27] Wikipedia [http://fr.wikipedia.org/wiki/Algorithme\\_esp%C3%A9rance-maximisation](http://fr.wikipedia.org/wiki/Algorithme_esp%C3%A9rance-maximisation) .
- [28] C.W. HSU, C.C. CHANG, AND C.J. LIN "A practical guide to support vector machine," Available: <http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>.
- [29] B.XIAOJUAN, S.QILONG, C.HAO, T.XUYAN, "Compression method based on training data set", 2008 Journal of. Systems Engineering and Electronics, vol. 19, pp. 198-201.
- [30] S. RUSSELL ET P. NORVIG, "Artificial Intelligence: A Modern Approach", 2ième édition, 2003, Prentice Hall.
- [31] <http://archive.ics.uci.edu/ml/> .
- [32] Wikipedia. [http://fr.wikipedia.org/wiki/Logique\\_floue](http://fr.wikipedia.org/wiki/Logique_floue) .
- [33] Dictionnaire informatique <http://dictionnaire.phpmyvisites.net/definition-logique-floue--11114.htm> .
- [34] <http://bmantel.club.fr/pages/logfloue/logfloue00.html> .
- [35] <http://www.4p8.com/eric.brasseur/lofo.html> .
- [36] ALI IDRI & ALAIN ABRAN, "La logique floue appliquée aux modèles d'estimation d'efforts de développements de logiciels Cas du modèle COCOMO'81" <http://www.gelog.etsmtl.ca/publications/pdf/560.pdf>.
- [37] VINCENT BARRA, VÉRONIQUE DELOUILLE et JEAN-FRANÇOIS HOCHEDÉZ, "Segmentation d'Images solaires en Extrême Ultraviolet par une Approche Classification floue Multi spectrale", Research Report 07-07, 2 avril 2007, Observatoire Royal de Belgique, BRUXELLES.
- [38] R.M. FAYE, S. SAWADOGO et F. MORA-CAMINO, "Logique floue appliquée à la gestion à long terme des ressources en eau" 2002, REVUE DES SCIENCES DE L'EAU, Rev. Sci. Eau 15/3(2002) 579-596.

- [39] Dr. RAY HOARE, *"Using CHAID for classification problems"*2004, A paper presented at the New Zealand Statistical Association 2004 conference, Wellington.
- [40] SOUMAYA CHAFFAR, PIERRE CHALFOUN, CLAUDE FRASSON, *"La prédiction de la réaction émotionnelle dans un environnement d'évaluation en ligne"* 2006, Département d'informatique et de recherche opérationnelle, Université de Montréal.
- [41] MIHAELA SCUTURICI, VASILE-MARIAN SCUTURICI, JÉRÉMY CLECH, DJAMEL ZIGHED "Navigation dans une base d'images à l'aide de graphes topologiques" 2004, Université Lumière Lyon, France.
- [42] JULIEN MAUPETIT, "Prédiction de la structure des protéines" Décembre 2007, Université Paris Diderot.
- [43] NARJÈS BOUFADEN, TRUONG LE HOANG, PIERRE DUMOUCHEL, *"Détection et prédiction de la satisfaction des usagers dans les dialogues Personne-Machine"* 2007, Conférence sur le Traitement Automatique des Langues Naturelles (TALN 2007). Toulouse, France.
- [44] ROMAIN VINO, NATALIA GRABAR, MATHIEU VALETTE, *"Application d'algorithmes de classification automatique pour la détection des contenus racistes sur l'Internet"* 2003, actes du colloque TALN 2003, Batz sur Mer, p. 257-284.
- [45] HYUNJUNG SHINA, SUNGZOO CHOB & TU-BINGEN, *"Response modeling with support vector machines"* 2006, Journal Expert Systems with Applications Vol. 30 No. 4.
- [46] CHI-HOON LEE, RUSSELL GREINER, AND MARK SCHMIDT, *"Support Vector Random Fields for Spatial Classification"*2005, Department of Computing Science, University of Alberta, Edmonton AB, Canada.
- [47] LAURIANNE SITBON, JENS GRIVOLLA, LAURENT GILLARD, PATRICE BELLOT, PHILIPPE BLACHE, *"Vers une prédiction automatique de la difficulté d'une question en langue naturelle"* 2006, Actes de la Conférence TALN 2006, Louvain, Belgique.
- [48] LCDR ERIC JAFAR, USN LCDR(Sel) TERENCE NOEL C. MEJOS, USN, LT CHIEH YANG, USN ADVISORS: Dr. KEEBOM KANG, Dr. UDAY M. APTE, AND DR. KENNETH H. DOERR , *"Naval Aviation Maintenance: A Case Study for Process Improvement"* 15 December 2006, Naval Postgraduate School. Monterey, California.

- [49] BO LIU, ZHIFENG HAO, XIAOWEI YANG and XUDONG LIN, "Chinese Question Classification with Support Vector Machine" July 2006, IJCSNS International Journal of Computer Science and Network Security, VOL.6 No.7A.
- [50] LOFTI ZADEH, "*Fuzzy Logic*", 1965, Information and Control 8 (3) 338–353.
- [51] Wikipedia <http://fr.wikipedia.org/wiki/M%C3%A9t%C3%A9orologie> .
- [52] Wikipedia <http://fr.wikipedia.org/wiki/Climatologie> .
- [53] Wikipedia <http://fr.wikipedia.org/wiki/Sismologie> .
- [54] tn.refer.org  
[http://www.tn.refer.org/hebergement/cours/logique\\_floue/we\\_hinfo/fonc.htm](http://www.tn.refer.org/hebergement/cours/logique_floue/we_hinfo/fonc.htm)  
 .
- [55] Wikipedia <http://en.wikipedia.org/wiki/Dendrogram> .
- [56] YANCONG ZHOU & HYUK CHO, "*Classification Algorithms on Text Documents*" 2001, University of Texas, USA.
- [57] LEVIN et al., 1986; SALAMOV AND SOLOVYEV, 1997; FIGUREAU ET AL., 2003; JIANG, 2003; KIM, 2004.
- [58] HUA and SUN, 2001; KIM and PARK, 2003; WARD et AL., 2003; HU et al., 2004.
- [59] LUDOVIC LE BIGOT - FRANCK PANAGET, ÉRIC JAMET, JEAN-FRANÇOIS ROUET, "*Études expérimentales sur le modèle de l'interlocuteur en dialogue personne-machine*" 2005, France.
- [60] <http://nlp.stanford.edu/IR-book/html/htmledition/rocchio-classification-1.html>.
- [61] R. RIFKIN, A. KLAUTAU, "*In Defense of One-Vs-All Classification*" 2004, Journal of Machine Learning Research, vol.5, pp. 101-141.

[62] SHIN & CHO, "*NPPS*" May 2005, Pattern Recognition Letters archive  
Volume 26, Issue 6.

[63] Wikipedia [http://fr.wikipedia.org/wiki/Th%C3%A9orie\\_des\\_contraintes](http://fr.wikipedia.org/wiki/Th%C3%A9orie_des_contraintes) .

[64] Wikipedia [http://en.wikipedia.org/wiki/Just\\_In\\_Time\\_\(business\)](http://en.wikipedia.org/wiki/Just_In_Time_(business)) .

[65] Wikipedia [http://en.wikipedia.org/wiki/Six\\_Sigma](http://en.wikipedia.org/wiki/Six_Sigma) .

[66] Wikidaimi [http://wiki.daimi.au.dk/cpntools-help/iid\\_data\\_values.wiki](http://wiki.daimi.au.dk/cpntools-help/iid_data_values.wiki) .