

Université du Québec à Montréal
Maîtrise en génie logiciel
Rapport de projet final

**MESURER LA TAILLE FONCTIONNELLE ET BÂTIR UN MODÈLE DE PRODUCTIVITÉ POUR LES
ACTIVITÉS DE MAINTENANCE ADAPTATIVE EN INDUSTRIE**

RAPPORT DE PROJET

Présenté comme exigence partielle de la maîtrise en génie logiciel

Réalisé par

ILIONAR SILVA

Direction :

ALAIN ABRAN

Le 28 juillet 2000

Mis à jour : 2 décembre 2000

Table de matières

TABLE DE MATIÈRES	2
LISTE DES FIGURES	4
LISTE DES TABLEAUX	4
LISTE DES ACRONYMES ET ABRÉVIATIONS	5
RÉSUMÉ	6
INTRODUCTION.....	7
CHAPITRE I.....	9
1 CONTEXTE DU PROJET	9
1.1 DESCRIPTION SOMMAIRE DE L'ORGANISATION ET DE SES PRODUITS.....	9
1.2 LE GROUPE D'AMÉLIORATION DE PROCESSUS ET UN DE SES OBJECTIFS.....	9
1.3 SITUATION ACTUELLE.....	10
1.4 BESOINS À COMBLER PAR LE PROJET	10
1.5 BESOINS DU LABORATOIRE DE RECHERCHE EN GESTION DES LOGICIELS (LRGL)	11
1.6 DÉFINITION DU MANDAT	12
1.7 LIMITES DU MANDAT.....	12
1.8 OBJECTIFS DE L'ENTREPRISE ET MESURES PRÉVUES POUR LEURS ATTEINTES	13
1.9 L'APPORT DU PROJET AU GÉNIE LOGICIEL.....	13
CHAPITRE II	16
2 APPROCHE DE RECHERCHE	16
2.1 APPROCHE GÉNÉRALE.....	16
2.2 HYPOTHÈSE DE RECHERCHE.....	17
2.3 LE MODÈLE DE RECHERCHE	18
2.4 CRITÈRES DE FIABILITÉ POUR UN MODÈLE DE PRODUCTIVITÉ.....	19
CHAPITRE III.....	21
3 DESCRIPTION DES DONNÉES	21
3.1 DESCRIPTION DE LA COLLECTE DE DONNÉES	21
3.2 ANALYSE DES VARIABLES.....	22
3.2.1 <i>La banque de données disponibles</i>	22
CHAPITRE IV	25
4 ANALYSE STATISTIQUE DE LA BANQUE DE DONNÉES DISPONIBLES	25
4.1 TAILLE FONCTIONNELLE ÉTENDUE TOTALE (FFP_TOTAL) ET TAILLE FONCTIONNELLE ÉTENDUE MODIFIÉE (FFP_MODIFICATION).....	26
4.2 EFFORT SANS ANALYSE	29
4.3 COMPLEXITÉ FONCTIONNELLE.....	31

CHAPITRE V	34
5 MODÈLE D'ESTIMATION DE PRODUCTIVITÉ	34
5.1 MODÈLE DU COÛT UNITAIRE MOYEN.....	34
5.1.1 <i>Qualité du modèle de productivité du coût unitaire</i>	35
5.1.2 <i>Modèle du coût unitaire par la complexité basse</i>	35
5.1.3 <i>Modèle du coût unitaire par la complexité haute</i>	36
5.2 MODÈLE DE RÉGRESSION LINÉAIRE	37
5.2.1 <i>Estimation de l'effort en fonction de la taille FFP modifiée, une seule variable</i>	38
5.3 MODÈLE À PLUSIEURS VARIABLES INDÉPENDANTES.....	39
5.3.1 <i>Modèle avec la quantité des programmes affectés</i>	40
5.3.2 <i>Modèle additif avec la complexité</i>	41
5.3.3 <i>Modèle multiplicatif avec la complexité</i>	43
5.3.4 <i>Qualité du modèle</i>	47
CHAPITRE VI	48
6 SOMMAIRE DES RÉSULTATS	48
7 BIBLIOGRAPHIE	51
ANNEXE A – BANQUE DE DONNÉES	53
ANNEXE B : RÉSULTATS DES ANALYSES STATISTIQUES	54

Liste des figures

FIGURE 4.1 : FFP_MODIFICATION : DISTRIBUTION CUMULATIVE DE FRÉQUENCES	27
FIGURE 4.2 : EFFORT_SANALYSE: DISTRIBUTION CUMULATIVE DE FRÉQUENCES.....	30
FIGURE 5.1 : COMPARAISON GRAPHIQUE ENTRE LA DROITE DU COÛT UNITAIRE ET LES DONNÉS RÉELLES)	35
FIGURE 5.2 : COMPARAISON GRAPHIQUE DE LA DROITE DU COÛT UNITAIRE ET LES DONNÉS RÉELLES PAR PROJETS AVEC COMPLEXITÉ BASE.....	36
FIGURE 5.3 : COMPARAISON GRAPHIQUE DE LA DROITE DU COÛT UNITAIRE ET LES DONNÉS RÉELLES PAR PROJETS AVEC COMPLEXITÉ HAUTE.....	36
FIGURE 5.4 : RÉGRESSION LINÉAIRE SIMPLE SANS LES DEUX OUTLIERS (19 PROJETS).....	38
FIGURE 5.5 : REPRÉSENTATION GRAPHIQUE DE L'ÉQUATION 5.4 ET 5.5.....	42
FIGURE 5.6 : REPRÉSENTATION GRAPHIQUE DE L'ÉQUATION 5.8 ET 5.9.....	44

Liste des tableaux

TABLE 3.1 : BANQUE DES DONNÉES AVEC LES VARIABLES PLUS IMPORTANTES.....	23
TABLEAU 3.2 – NOM ET DESCRIPTION DES DONNÉES RECUEILLIES.....	24
TABLE 4.1 : TEST DE DISTRIBUTION NORMALE POUR VARIABLES CONTINUES - KOLMOGOROV-SMIRNOV	28
TABLE 4.2. OUTLIER	28
TABLE 4.3 : TEST DE DISTRIBUTION NORMALE POUR VARIABLES CONTINUES - KOLMOGOROV-SMIRNOV	28
TABLE 4.4. OUTLIER	29
TABLE 4.5 : TEST DE DISTRIBUTION NORMALE POUR VARIABLES CONTINUES - KOLMOGOROV-SMIRNOV	30
TABLE 4.6 : RÈGLE DE TRANSFORMATION POUR LA COMPLEXITÉ.....	33
TABLE 5.1 : RELATION STATISTIQUE ENTRE L'EFFORT ET LE FFP SELON DIFFÉRENTES APPROCHES.....	39
TABLE 5.2 : TEST DE « TURNING POINTS » ET DE « RANK » POUR VALIDER L'ASPECT ALÉATOIRE DE RÉSIDUS.....	45
TABLE 5.3 : COMPARAISON DE LA QUALITÉ DU MODÈLE PAR L'ÉQUATION 5.6 ET 5.10.....	47

Liste des acronymes et abréviations

Cette section définit les différents acronymes et abréviations utilisés dans ce document.

Acronymes et Abréviations	Description
ABC	Defence System ABC
FFP	Full Function Point, Abréviation anglaise pour la méthode des Points de fonction étendus
FPA	Function Point Analysis
IFPUG	International Function Point Users Group
LRGL	Laboratoire de recherche en gestion des logiciels
IA Inc.	Industrial Organization Inc.
PF	Point de fonction
PFE	Points fonction étendus
SEPG	Software Engineering Process Group
SLOC	Lignes de code
UQAM	Université du Québec à Montréal

Résumé

Souvent plusieurs produits informatiques ne correspondent pas aux exigences ou aux besoins exprimés par le client, dépassant les prévisions de coûts, d'efforts et d'échéanciers avec un faible niveau de qualité. Pour cet aspect, un grand nombre d'organisations, entre autres la compagnie IA du secteur privé canadien de la Défense, ont besoin de mettre en place des méthodes pour une mesure uniforme afin de pouvoir contrôler et améliorer leur processus de développement et surtout d'entretien.

À cet effet, ce projet examine l'influence de divers facteurs en utilisant la méthode des points de fonction étendus pour mesurer la taille fonctionnelle des projets sur l'effort requis lors des activités de maintenance adaptative chez IA Inc. Ceci nous a permis d'établir une relation entre l'information disponible (effort, ligne de code modifiée, complexité etc.) avec les points de fonction étendus mesurés. Cette relation était considérée comme base de divers modèles de productivité pour les projets d'entretien de l'une des applications les plus importantes dans le domaine en temps réel nommée ABC, laquelle fait partie intégrante du système DEF. Ces modèles considèrent l'effort puisqu'il est un facteur primordial pour déterminer les coûts et les échéanciers d'un logiciel.

Pour réaliser ce projet, on a premièrement effectué le comptage de ce système ABC, qui jusque là avait 21 projets de maintenance adaptative au niveau du logiciel. Cette information, avec tout le reste de la documentation disponible sur les projets, constituaient notre banque de données. Notre démarche consistait alors à appliquer des outils statistiques sur cette information, ce qui nous permettait d'établir des modèles de productivité, lesquels identifiaient plusieurs caractéristiques propres des projets de maintenance adaptative du logiciel ABC.

Les modèles utilisés dans cette recherche révèlent que les points de fonction étendus à eux seuls n'expliquent pas l'effort requis pour réaliser un projet de maintenance adaptative. Ceci ne nous permettait pas d'utiliser la relation de la taille fonctionnelle et l'effort comme modèle de productivité. La qualité de cette relation était très faible et non adéquate dans les modèles de productivité à une seule variable indépendante avec l'analyse de régression et le coût unitaire moyen.

Toutefois, une analyse plus détaillée des facteurs pouvant affecter les activités de maintenance nous a permis de voir que la qualité de cette relation entre la taille fonctionnelle et l'effort s'était améliorée au niveau de l'influence du facteur de complexité fonctionnelle des projets. Ce facteur était significatif comme variable indépendante dans le modèle de productivité retenu, affectant l'effort dans les activités d'entretien.

Il a été également constaté que la méthode de points de fonction étendus est adéquate pour mesurer des projets de maintenance dans le domaine en temps réel. Cette recherche donne donc des résultats satisfaisants, ce qui contribue à l'utilisation de cette méthode dans ce domaine pour mesurer des logiciels et augmente ainsi les possibilités de généralisation de ces résultats dans l'entreprise IA Inc., pour d'autres entreprises et en général pour la communauté scientifique.

Introduction

Depuis les deux dernières décennies, il existe en génie logiciel une méthode qui permet de mesurer la taille fonctionnelle des logiciels: la méthode des points de fonction. Cette méthode est indépendante de la technologie et applicable à une vaste gamme d'environnements de développement de logiciels. Cela offre l'opportunité de mesurer la productivité, d'estimer les coûts des projets, de contrôler les changements de la taille fonctionnelle d'un logiciel et d'améliorer la communication avec les utilisateurs en se concentrant sur des exigences fonctionnelles correspondant à leurs perspectives lors de la phase de spécification des exigences [HEL99]. Cette méthode de mesure a été conçue par Allan Albrecht en 1979 et améliorée postérieurement par IFPUG à partir de 1984 avec l'ajustement des standards et des normes qui ont aidé à promouvoir son évolution et son utilisation comme une mesure basée sur les fonctionnalités livrées aux usagers [ABR96]. Cette méthode a été conçue et mise au point pour des applications de gestion où son efficacité a été prouvée.

Toutefois, « Function Point Analysis » (FPA) n'a pas atteint le même niveau d'acceptation pour les logiciels en temps réel où elle n'est pas perçue comme une mesure adéquate de la fonctionnalité [KEC98]. FPA utilise la quantité de données emmagasinées comme un facteur déterminant de la taille fonctionnelle d'un logiciel sans considérer complètement le processus d'emmagasiner ces valeurs, pour les applications en temps réel. Ces applications ont des caractéristiques spécifiques pour les données et transactions: "Un grand nombre de données transitoires ou temporaires et un grand nombre de traitements internes, c'est-à-dire, un grand nombre de processus composés de plusieurs sous processus ou étapes" [DES97]. En conséquence, ces caractéristiques intrinsèques des systèmes en temps réel ne sont considérées que d'une façon superficielle, ce qui occasionne une sous-estimation de la taille fonctionnelle pour ce type de systèmes [ABR97].

Conséquemment, une nouvelle méthode a été développée par le professeur Alain Abran et ses collègues pour rendre la mesure plus adéquate, principalement pour les systèmes en temps réel: la méthode des points de fonction étendus.

Le présent document constitue le rapport final du projet en entreprise réalisé dans le cadre du programme de maîtrise en génie logiciel à l'UQAM. Ce projet intitulé « Mesurer la taille fonctionnelle du logiciel ABC et bâtir un modèle de productivité pour ces activités de maintenance adaptative chez IA Inc. a été réalisé au sein du groupe d'amélioration de processus de génie logiciel d'IA (SEPG), lequel contribue aux améliorations des processus de développement et d'entretien des logiciels.

Ce projet, réalisé sur un période de 15 mois, du 1 avril 1999 au 1 juillet 2000, avait comme objectif principal de mesurer, chez IA Inc. (IA), la taille fonctionnelle du système ABC ainsi que de déterminer, en tenant compte du contexte organisationnel, un modèle d'estimation de productivité pour des projets de maintenance adaptative ("modifications pour adapter un logiciel aux changements de ses exigences et de son environnement" [ABR93]) de ce logiciel en temps réel. Un tel modèle s'inscrit parmi les initiatives proposées par l'industrie du logiciel pour mesurer la taille d'un logiciel en se concentrant sur des exigences fonctionnelles correspondant aux perspectives de l'utilisateur. Ceci est une exigence lors de l'analyse de productivité d'un logiciel d'après les modèles classiques de gestion [ABR96].

Ce rapport présente en détail, dans une première partie, la proposition du projet de recherche. Cette partie décrit le contexte global du projet, la problématique, les objectifs et l'approche employée pour les rencontrer.

La deuxième partie porte sur l'approche de recherche préconisée, approche qui utilise la méthode de points de fonction étendus pour mesurer les projets de maintenance adaptative ainsi que les données historiques. On présente divers modèles mathématiques qui peuvent exprimer convenablement la relation quantifiée entre l'effort et la taille fonctionnelle ainsi que entre l'effort et d'autres facteurs.

La troisième partie est dédiée à la présentation et description des données de mesure disponibles. On y décrit, pour la réalisation du projet, le processus employé pour la collecte de données et sa validation.

La quatrième partie du rapport se concentre sur des analyses statistiques de la banque de données disponible et des résultats. Dans cette partie, on présente le contexte théorique dans lequel s'inscrit l'analyse statistique de la banque de données. On y aborde l'hypothèse standard d'une distribution normale des variables ainsi que l'analyse des paramètres statistiques à considérer.

Finalement la cinquième et dernière partie du rapport présente un sommaire des résultats par rapport aux objectifs définis. Il s'agit plus particulièrement de présenter les résultats d'une recherche et les interprétations qui en découlent.

CHAPITRE I

1 Contexte du projet

1.1 Description sommaire de l'organisation et de ses produits

IA Inc. est un intégrateur de systèmes spécialisé dans l'assemblage, l'intégration de systèmes de communications et des systèmes de défense. Ses services de production incluent des assemblages et du matériel pour calibrer les appareils et les systèmes électroniques et mécaniques, un véhicule extérieur incluant un système pour suivre l'essai, un assemblage modulaire et des compartiments d'essai « test bays » pour l'intégration de systèmes et de sous-systèmes.

IA Inc. est avant tout l'intégrateur d'un système de missiles de défense aérienne. Plus de 120 ingénieurs sont impliqués dans le développement et l'entretien du système.

Ce système de missiles DEF, composé de plusieurs sous-systèmes, a été conçu pour assurer la protection de points vitaux tels que des ports, des zones pétrolières ou des centres de transmissions. Au début, il était produit par IA pour les forces armées canadiennes.

1.2 Le Groupe d'amélioration de processus et un de ses objectifs

L'objectif de la compagnie IA Inc. est d'obtenir continuellement des améliorations dans les processus de développement et d'entretien de ses logiciels en utilisant des méthodes et des mesures existantes dans le domaine du génie logiciel qui lui permettent de gérer efficacement ses logiciels.

Compte tenu de l'importance du système DEF, la compagnie vise, grâce à ce projet à [TRU99]:

- 1.- Mesurer premièrement un de ses sous-systèmes plus importants nommé ABC;
- 2.- Obtenir une mesure financière \$/mesure qui facilite l'intégration de ces ratios aux actifs comptables de la compagnie;
- 3.- Construire des modèles de productivité plus fiables et applicables pour leur future maintenance adaptative afin de répondre efficacement aux exigences de ses clients à des coûts raisonnables et avec des efforts appropriés lors de l'estimation des modifications.

Le secteur de la compagnie visé par le projet est le groupe d'amélioration de processus de génie logiciel d'IA Inc. (SEPG), lequel contribue aux améliorations des processus de développement et d'entretien des logiciels en se basant sur les méthodes et les techniques pratiquées en génie logiciel.

1.3 Situation actuelle

La compagnie IA Inc. a présentement besoin de mettre en place des améliorations dans ses processus internes pour standardiser et uniformiser les différentes pratiques utilisées pour l'estimation de leurs projets d'entretien.

Principalement, la maintenance adaptative de son logiciel "ABC" introduit des limitations inhérentes aux méthodes de gestion pratiquées par les différents intervenants lors de l'estimation des coûts et des efforts.

Les divers facteurs ayant contribué à cette problématique sont :

- **Carence de procédures standards à suivre lors de l'estimation**

Au cours des diverses modifications du logiciel "ABC", les différents gestionnaires ont utilisé des procédures et des mécanismes correspondants à leurs expertises, qui n'ont pas été également suivis ni documentés lors de l'estimation. Ceci entraîne alors des faiblesses dans l'atteinte d'un niveau d'estimation approprié et uniforme pour toutes les modifications au cours du cycle d'entretien du logiciel ou du matériel.

- **Duplication de l'effort**

Bien que l'emphase n'aie pas été mise sur l'évaluation adéquate de la documentation existante de la part des gestionnaires successeurs, leurs efforts ont été dupliqués. Les différents intervenants dans la maintenance adaptative du logiciel ABC ont développé des grilles et des documents selon leur interprétation, sans mettre l'accent vraiment sur les méthodes ou les mesures utilisées précédemment pour l'estimation de productivité.

1.4 Besoins à combler par le projet

La compagnie a besoin d'une méthode uniforme pour mesurer la taille pour l'ensemble de ses logiciels développés ou maintenus afin de lui permettre principalement de calculer des mesures de performance organisationnelle de développement et maintenance, d'améliorer son processus d'estimation des efforts requis pour le développement et la maintenance ainsi que de comparer, par

rapport à la taille d'un logiciel, les divers efforts fournis dans leur processus de développement et d'entretien.

À cet effet, la réalisation de ce projet permettra à la compagnie IA d'appliquer la méthode des points de fonction étendus proposée par le Laboratoire de Recherche en gestion des logiciels, pour mesurer fonctionnellement son logiciel ABC en temps réel ainsi que pour bâtir des modèles de productivité qui répondent aux exigences demandées par sa maintenance adaptative au sein du département concerné. Ceci permettra de développer un ou des modèles appropriés de productivité qui pourront être utilisés pour les estimations futures des efforts et des coûts requis lors des modifications.

Ce modèle devra avoir un niveau explicatif de la productivité suffisant pour sa mise en oeuvre lors de l'estimation future des projets d'entretien [TRU99].

1.5 Besoins du Laboratoire de recherche en gestion des logiciels (LRGL)

Le laboratoire LRGL a comme mission de développer des modèles et des instruments analytiques de mesures pour permettre aux industries impliquées en génie logiciel d'améliorer leurs processus décisionnels afin de répondre efficacement à leurs objectifs d'affaires.

Étant donné que la méthode des points de fonction existe depuis quelques décennies, mais qu'elle ne permet pas de capturer adéquatement les caractéristiques fonctionnelles intrinsèques des systèmes en temps réels, le laboratoire a donc développé une extension de cette méthode : la méthode des points de fonction étendus qui permet de mesurer non seulement des applications en temps réel, des logiciels embarqués mais également des applications de gestion incluant des caractéristiques d'affaires. [MOR98]

Le laboratoire recherche continuellement des entreprises intéressées à collaborer pour utiliser la méthode des points de fonction étendus lors de la mesure de la taille fonctionnelle d'un logiciel en temps réel ainsi que lors de l'élaboration de modèles de productivité en se basant sur leurs données historiques. Cela lui permet d'augmenter la gamme de contextes applicables afin d'offrir de l'étalonnage « Benchmarking » aux entreprises qui souhaitent se situer par rapport à d'autres entreprises qui développent et maintiennent des logiciels dans des domaines similaires. De façon plus spécifique, le laboratoire vise à vérifier que l'utilisation de la méthode des points de fonction étendus peut être adaptée aux divers contextes industriels.

1.6 Définition du mandat

L'objectif du projet est d'offrir à la compagnie IA Inc., une méthode de mesure, les points de fonctions étendus, qui lui permettra de mesurer la taille fonctionnelle de son système ABC en temps réel afin de pouvoir estimer les efforts requis pour leurs activités de maintenance adaptative.

De plus, ce projet permettra de construire un modèle de productivité basé sur cette mesure fonctionnelle ainsi que sur toute l'information disponible découlant des activités de la maintenance adaptative de ce logiciel.

Dans le domaine du génie logiciel, tel que mentionné précédemment, les mesures de la taille fonctionnelle sont considérées comme essentielles pour quantifier le produit dérivé ainsi que son processus de développement ou d'entretien [ABR97]. Par exemple, une mesure de la taille fonctionnelle d'un logiciel, mesurée indépendante des technologies, est reconnue comme un aspect clé dans l'étalonnage et l'utilisation des modèles de productivité en génie logiciel. [OLI99]

À cet effet, la compagnie IA Inc. vise avec ce projet à ce que ces mesures puissent être utilisées pour construire un ou des modèles de productivité. Ces modèles pourraient être utilisés pour l'amélioration du processus d'entretien de ce logiciel, ainsi que pour comparer la productivité des diverses techniques et technologies mises en place chez IA Inc. lors de la maintenance adaptative de leurs logiciels.

D'autre part, ce projet permettra au LRGL d'appuyer les résultats des recherches sur la validation de la méthode des points de fonction étendus pour mesurer la relation entre l'effort et la taille fonctionnelle de logiciels embarqués en temps réel.

Par ailleurs, l'étudiant vise à répondre à une des exigences requises à l'UQAM pour l'obtention de son diplôme dans le programme de maîtrise en génie logiciel, soit le projet final à travers ce mandat.

1.7 Limites du mandat

L'étude se limitera à construire un modèle de productivité de maintenance adaptative du système ABC appliquant la méthode des points de fonction étendus et sur les données disponibles et mesurées chez IA Inc..

Les modèles de productivité à établir dans ce projet seront applicables uniquement aux activités de maintenance adaptative du logiciel ABC provenant de la compagnie IA inc., parce que les données s'appliquent à des activités de modifications fonctionnelles de ce logiciel, ce qui est réalisable seulement dans la maintenance adaptative.

Bien que ces modèles ne soient vérifiés que pour le contexte analysé, ils pourraient être comparés et utilisés par d'autres activités de recherche de la méthode des points de fonction étendus dans la même compagnie ou dans d'autres entreprises.

1.8 Objectifs de l'entreprise et mesures prévues pour leurs atteintes

Dans le cadre de l'amélioration de ses processus, le groupe de processus de génie logiciel chez IA Inc. (SEGP) s'est fixé comme objectif l'implantation d'une méthode de mesure des logiciels qui lui permet de répondre aux exigences de la maintenance adaptative des logiciels réalisés et maintenus chez IA Inc.

Cette méthode de mesure devra fournir à la compagnie une structure uniforme de mesure pour ses projets de maintenance adaptative ainsi que des mécanismes d'estimations de modèles de productivité pour l'évaluation de futures modifications fonctionnelles de ses logiciels.

Afin d'atteindre cet objectif, la compagnie a sélectionné le logiciel ABC en temps réel comme un projet pilote, auquel sera appliquée la méthode des points de fonction étendus pour mesurer sa taille fonctionnelle. Ceci est dans le but de prévoir la maintenance adaptative de ce logiciel en termes de sa taille fonctionnelle et de l'effort requis pour le produire.

Avec l'aide de l'étudiant, la compagnie IA Inc. effectuera le comptage de PFE et fournira toutes les données de mesure pertinentes pour que l'étudiant puisse construire les modèles de productivité; ceci permettra de déterminer la qualité de la relation entre les PFE et l'effort pour les activités de maintenance adaptative.

1.9 L'apport du projet au génie logiciel

Le génie logiciel est reconnu comme une discipline dont les origines remontent à la conférence de l'OTAN « North Atlantic Treaty Organization » en 1968. À cette conférence, un des principaux organisateurs, M. Bauer, a utilisé le terme « Génie logiciel ». Ce fut la première fois que le terme

Génie logiciel fut employé pour mettre en valeur la nécessité de l'approche d'ingénierie au sein du développement du logiciel. [MIL98].

Plus tard, M. Bauer a défini le génie logiciel comme « L'établissement et l'utilisation des principes solides d'ingénierie (méthodes) afin d'obtenir économiquement un logiciel fiable et qui travaille sur de vraies machines » (Bauer, F. L. Software Engineering. Information Processing 71, 1972).

Au cours des trente années qui ont suivi cette première conférence de génie logiciel, des tentatives significatives ont été faites pour améliorer l'efficacité globale du processus de développement du logiciel et pour réduire ainsi la fréquence et la sévérité des défaillances dans les projets de logiciel.

Une partie de cet effort d'amélioration a consisté à développer des mesures quantitatives, lesquelles peuvent être utilisées pour décrire, mieux comprendre et contrôler plus efficacement le cycle de vie de développement du logiciel. Ainsi, plusieurs mesures de logiciel et divers modèles ont été présentés pendant cette dernière période dans le domaine du génie logiciel.

Selon Everaldo E. Mills, de l'Université de Seattle [MIL98], la gestion efficace et pertinente de n'importe quel processus exige la quantification, la mesure et l'utilisation de modèles. La mesure du logiciel fournit alors une base quantitative pour le développement et la validation de la qualité des modèles du processus de développement de logiciel ainsi que pour l'amélioration de la productivité et de la qualité. Actuellement, la mesure du logiciel est plus généralement utilisée pour construire des modèles de productivité et des processus de développement de logiciel.

Au cours des dernières années, un certain nombre d'études et de commentaires ont été faits sur l'état du génie logiciel, lesquels tendent à démontrer que cette profession n'a pas atteint encore sa maturité. Il existe actuellement une nécessité spéciale pour cerner et classer les diverses sous-disciplines qui composent ce domaine afin de le rendre plus stable [HIL99][BOU99].

Comme les mesures et les modèles actuels ne sont pas pleinement satisfaisants, un certain nombre d'organismes et de laboratoires sont en train de les améliorer. Alors, l'apport de ce projet de maîtrise dans le domaine du génie logiciel est de vérifier l'utilité et l'efficacité de la méthode des points de fonction étendus pour mesurer la taille fonctionnelle d'un logiciel en temps réel.

C'est le premier ensemble de mesures avec FFP en maintenance adaptative avec une collecte de données qui permettra de construire des modèles d'estimation et de les faire évoluer. Au fur et à mesure que les chercheurs du domaine acquerront plus d'expérience et une maturité reconnue avec

les diverses mesures et modèles proposés, les résultats devraient s'améliorer et permettre des comparaisons plus significatives.

CHAPITRE II

2 Approche de recherche

2.1 Approche générale

Dans la littérature, diverses études décrivent l'amélioration que pourrait avoir la qualité de la relation entre la taille fonctionnelle d'un logiciel et l'effort dans un modèle de productivité pour les activités de maintenance adaptative en termes d'autres facteurs différents [MAY95]. Cependant, ceci est une mesure controversée. Boehm définit la productivité comme le rapport entre des livrables (sorties) produits par les processus et des entrées consommées par le processus, soit :

$$\text{Productivité} = \text{Sorties produites par le processus} / \text{Entrées consommées par le processus.}$$

La difficulté en utilisant ce taux est la controverse de ce qui constitue les entrées et les sorties du processus de développement d'un logiciel, et plus particulièrement pour le processus d'entretien.

Dans cette étude, la productivité des activités de maintenance adaptative pour le logiciel ABC dans l'entreprise IA Inc. est définie en fonction de la taille fonctionnelle du logiciel et de l'effort. La méthode de points de fonction étendus est utilisée afin de mesurer la taille fonctionnelle des projets des activités de maintenance adaptative du logiciel ABC ainsi que l'utilisation de l'information concernant l'effort pour développer chaque projet.

Toutefois, cette recherche examine en plus certains facteurs qui pourraient affecter la productivité de ce type d'activité tels que mentionné précédemment, mais principalement en termes de l'effort. Il s'agit donc de déterminer ces facteurs qui ont été mesurés qui affectent l'effort dans le processus d'entretien d'un logiciel; ceux qui viennent plus particulièrement des caractéristiques du produit, complexité, quantité de sous-programmes, lignes de code modifiées.

De façon systématique, cette recherche vise à l'évaluation des effets que peuvent occasionner la modification de la variable indépendante (la taille fonctionnelle) sur la variable dépendante (l'effort) lors de l'addition d'autres facteurs au modèle de productivité.

2.2 Hypothèse de recherche

L'hypothèse nulle qui énonce formellement les relations attendues entre la taille fonctionnelle d'un logiciel et son effort est la suivante :

L'effort requis pour produire un projet de maintenance adaptative est supérieur lorsque ses points de fonctions étendus augmentent; une contribution positive.

L'approche recueille et analyse les données pour quantifier des facteurs qui affectent l'effort dans les activités de maintenance. Cette quantification détermine l'étendue de l'effet de la taille fonctionnelle sur l'effort et montre ainsi la relation quantifiée entre l'effort et d'autres facteurs.

Les modèles proposés pour exprimer la relation entre la taille fonctionnelle et l'effort sont:

- Le modèle du *coût unitaire moyen*, lequel est défini comme le rapport entre l'effort et le nombre de points de fonction étendus. :

$$\text{Coût unitaire moyen} = \frac{1}{n} \sum_{i=1}^n \frac{\text{Effort}_i}{\text{FFPi}}$$

Il donne comme résultat l'effort moyen nécessaire pour produire un point de fonction étendu (dans notre cas le nombre moyen d'heures). Un modèle de productivité construit avec le coût unitaire moyen a donc la forme suivante :

$$\text{l'effort calculé} = \text{nombre de PFE} * \text{coût unitaire moyen}$$

- Le modèle de *Régression linéaire* (simple et multiple), avec les points de fonction étendus et les divers facteurs influençant l'effort comme variable indépendante et l'effort comme variable dépendante.

Dans une première section, un modèle mathématique est présenté en utilisant le modèle du *coût unitaire moyen* pour analyser l'effort moyen nécessaire pour produire un point de fonction étendu. Par la suite, l'analyse de régression linéaire simple est proposée pour analyser la relation entre la variable indépendante X (la taille fonctionnelle) et la variable dépendante Y (l'effort). Cela pourrait fournir une courbe standard pour prédire de nouvelles valeurs pour X à partir de Y et vice-versa.

Dans une deuxième section, un autre modèle mathématique est offert à partir de l'analyse de régression linéaire multiple pour quantifier les influences que divers facteurs ont sur la productivité des activités de maintenance adaptative.

La sortie de chaque modèle est une valeur prédictive de l'effort requis pour développer un produit final du processus d'entretien. La position des facteurs et leurs coefficients et exposants correspondants dans le modèle fournissent une base pour comprendre l'effet qu'a un facteur sur la sortie du modèle. Le modèle rend également explicite le degré d'influence relatif d'un facteur parmi les autres facteurs dans le modèle. La qualité du modèle est déterminée par certains critères de fiabilité, plus principalement par l'erreur relative moyenne (ERM) et le coefficient de détermination R^2 selon le cas. Ces critères seront décrits postérieurement dans la section 2.4

2.3 Le modèle de recherche

Afin de s'assurer que le modèle d'estimation de productivité fournira des résultats satisfaisants, il existe plusieurs exigences qu'un modèle d'estimation doit adresser pour soutenir (supporter) cette recherche :

- Le modèle doit utiliser la taille fonctionnelle comme entrée. Ceci montrera que la taille fonctionnelle peut être mesurée et son importance pour expliquer la variation de l'effort.
- Le modèle doit être précis. La forme et les coefficients du modèle doivent être d'un bons représentants du monde réel.
- Le modèle doit être explicable. L'effet que produira le changement de chaque paramètre du modèle sur l'effort doit être compréhensible.
- Le modèle devrait utiliser juste assez de facteurs afin que la variation de l'effort soit expliquée et chaque facteur soit significatif.
- Le modèle devrait utiliser des facteurs qui sont indépendants entre eux, mais qui sont liés à l'effort. Ceci empêche le double comptage et rend le modèle plus stable.
- Il doit être possible de calibrer numériquement les facteurs dans le modèle en utilisant l'analyse de régression multiple. Ce type d'analyse examine les corrélations entre les variables indépendantes, précise la qualité de l'ajustement du modèle aux données et enregistre la

signification que chaque variable indépendante a sur l'effort prédictif. En plus, des variables indépendantes faibles peuvent être retirées du modèle, permettant ainsi de déterminer la force des variables restantes.

- On propose un modèle de recherche basé sur le modèle de régression linéaire multiple avec des variables quantitatives et qualitatives; ceci sera discuté dans la section 5.3. Ce modèle identifiera les indicateurs quantitatifs pour les classes de variables qualitatives afin que celles-ci puissent être utilisées dans ce modèle. Une variable qualitative avec c classes sera représentée par $c-1$ variables indicatrices, donc chacune aura les valeurs 0 et 1. Les variables indicatrices sont souvent nommées « dummy variables or binary variables ». [JOH93]

2.4 Critères de fiabilité pour un modèle de productivité

Il existe des critères statistiques employés pour juger à quel point un modèle reflète les données et à quel point il permet de les prédire. R^2 , $\text{Adj-}R^2$ et l'écart-type (standard error) donnent une indication de la qualité d'ajustement du modèle aux données. $\text{Pred}(X)$ donnent une indication de l'exactitude du modèle.

Afin de déterminer la qualité des modèles de productivité à proposer pour les activités de maintenance adaptative dans cette étude, on a sélectionné les indicateurs statistiques suivants :

L'erreur relative moyenne (ERM), qui signale l'importance de l'écart en pourcentage, entre les valeurs estimées par un modèle et les données réelles. Des valeurs estimées parfaites auraient une ERM de zéro. [MAY95]

Qualité de la prédiction du modèle : Selon l'approche classique du génie logiciel un modèle de productivité est considéré bon s'il est capable de rencontrer le critère de l'erreur relative moyenne (Magnitude Relative Error) de +/- 25% pour 75% des observations, ceci est aussi connu comme la fiabilité de la prédiction [CON86], [VER81]. Ces chercheurs emploient cette notion pour définir une mesure de qualité de la prédiction du modèle en définissant qu'une évaluation technique peut être acceptable si $\text{PRED}(0.25)$ est au moins 0.75.

En plus, d'autres auteurs (Kemerer 1987, Miyazaki 1985, Gulezian 1991, Helm 1992) considèrent comme acceptable une erreur relative moyenne de +/- 20% dans 70% des observations [ABR93-2].

Coefficient de détermination (R^2) : Une autre mesure globale de la qualité d'un modèle donnée est son coefficient de détermination (R^2), qui décrit le pourcentage de variance expliqué par la variable prédictive utilisée dans le modèle de régression linéaire. La valeur de ce coefficient varie entre 0 et 1; donc un R^2 de près de 1 indique que presque toute la variabilité de la réponse de la variable prédictive est expliquée par le modèle, c'est-à-dire, qu'il existe une forte relation entre la variable indépendante et la variable dépendante.

R^2 Ajusté, Adj- R^2 est un indicateur plus réaliste de la qualité d'ajustement du modèle parce qu'il est ajusté par le nombre de paramètres dans le modèle. R^2 ajusté est toujours moins que R^2 [WEI85]. Adj- R^2 est représenté par l'équation 2.1: n est la quantité des variables et n - k - 1. Le degré de liberté est le nombre d'observations moins le nombre de variables indépendantes moins une :

$$\text{Adj-}R^2 = 1 - \frac{n-1}{n-k-1} * (1- R^2) \quad \text{équation 2.1}$$

En outre, les coefficients des variables indépendantes dans les modèles de productivité sont évalués en utilisant le paramètre statistique P value. Ceci afin de déterminer leur influence significative dans le modèle lorsqu'ils sont inférieurs à 0.05.

CHAPITRE III

3 Description des données

Afin d'atteindre l'objectif du projet (établir un modèle de productivité qui permet à la compagnie IA d'obtenir une estimation de l'effort au moyen des FFPs ainsi que de pouvoir inférer le coût de maintenance de ses projets), nous avons réalisé un processus systématique de collecte de données historiques afin de disposer d'une source importante de données industrielles lors de l'analyse.

Les différentes données recueillies serviront principalement à déterminer leur influence sur la relation entre la taille fonctionnelle et l'effort des activités de maintenance adaptative chez IA.

Il faut mentionner qu'on se sert de la *taille fonctionnelle mesurée* au moyen des FFPs pour chacun des programmes qui constituent le logiciel ABC et de *l'effort dépensé* (information fournie à partir du système de comptabilisation des efforts). L'effort dépensé dans les diverses phases du cycle de développement d'un projet est enregistré en heures par personne. Cette information nous a permis de mesurer la productivité des 21 projets en étude en utilisant la technique de coût unitaire moyen et de régression linéaire, ainsi que de mieux exprimer la relation existante entre la taille fonctionnelle et l'effort de développement et d'entretien.

3.1 Description de la collecte de données

Il existe trois grandes sources de données: l'utilisation de documents, l'observation par le chercheur (mandataire) et l'information fournie par les principaux intervenants de l'entreprise dans le projet. Au tout début, les données disponibles comprenaient l'information enregistrée concernant les projets de maintenance adaptative au niveau logiciel et matériel. Cependant, ces données ont été validées et approuvées en considérant leur importance et leur influence dans la productivité par un analyste fonctionnel des projets d'entretien du logiciel ABC et par un représentatif du groupe d'amélioration de processus de génie logiciel d'IA Inc.

Les projets de maintenance adaptative au niveau du matériel (Hardware) n'ont pas été sélectionnés à cause de notre intérêt d'étude dans ce mandat mais surtout pour la carence de documentation et l'absence de prise de contact avec les divers intervenants de ces projets. On a dû se concentrer sur le contenu fonctionnel de chaque projet de maintenance adaptative au niveau logiciel, c.-à-d. sur

des modifications et des améliorations fonctionnelles des programmes qui constituent le logiciel ABC plutôt que sur les difficultés ou les améliorations matérielles pour celui.

Le processus de validation des données dans ce projet est basé sur les principes suivants:

- Validation sommaire des modifications ou additions fonctionnelles dans chacun des projets d'entretien.
- Niveau de complexité fonctionnelle à l'égard des spécifications fonctionnelles au niveau du logiciel.
- Validation des lignes de code modifiées ou additionnées avec et sans commentaires lors de la maintenance adaptative.
- Validation de l'effort dépensé tout au long du cycle de vie de développement des projets de maintenance adaptative.

Les diverses étapes de ce processus de validation nous ont permis d'assurer la qualité et la pertinence des données à étudier, cela grâce à l'obtention d'un feed-back direct de la part des utilisateurs potentiels du logiciel ABC.

3.2 Analyse des variables

Dans cette section, les données seront groupées en variables quantitatives et/ou qualitatives. Elles proviennent de la phase de mesure, d'observation et d'inspection technique de l'équipe de travail faisant partie intégrale du processus d'analyse des données. Chaque variable obtenue permettra de trouver leur influence dans notre analyse de productivité et de modéliser les différentes relations empiriques observées. Le but est donc de présenter l'ensemble des données qui serviront à l'analyse descriptive de nos variables dépendantes et indépendantes telles que l'effort et la taille fonctionnelle, ainsi qu'à la construction des modèles de productivité qui représentent cette relation avec d'autres facteurs influents.

3.2.1 La banque de données disponibles

L'ensemble des données disponibles pour l'étude comprend les 21 projets de maintenance au niveau du logiciel, tel que mentionné précédemment. Chaque projet est constitué d'un ou plusieurs modules, lesquels représentent des additions ou des modifications fonctionnelles effectuées lors de la maintenance adaptative. Chaque module utilise le même langage de programmation. La table 3.1 présente le numéro d'identification des projets de maintenance adaptative (STR) avec le nombre correspondant de modules affectés et les variables les plus

importantes utilisées. Plus de détails sur l'ensemble des variables disponibles sont présentés dans l'annexe A.

ID	# Prog. Modified	Effort	FFP_Modification	Complexity
0004	14	88	216	0
0016	53	956	618	0
0020	7	148	89	0
0059	1	66	3	1
0061	1	83	3	1
0065	3	34	7	0
0092	3	96	21	0
0093	4	84	25	0
0094	8	31	42	0
0150	6	409	46	1
0153	1	30	2	0
0158	2	140	2	1
0172	4	308	67	1
0174	7	244	173	0
0192	4	188	25	1
0202	5	34	1	0
0203	1	73	1	1
0207	1	27	1	0
0208	1	91	8	1
0210	1	13	19	0
0213	18	724	157	1

TABLE 3.1 : BANQUE DES DONNÉES AVEC LES VARIABLES LES PLUS IMPORTANTES.

Il est à noter que non seulement le nombre total de points de fonction étendus de chacun des projets de maintenance adaptative et le nombre total d'effort dépensé ont été collectés, mais aussi d'autres types de données ont été recueillies afin d'identifier certains facteurs qui pourraient affecter la productivité des activités d'entretien, selon la littérature (Voir tableau 3.2).

Toutefois, l'information disponible pour les divers projets de maintenance en étude, nous a permis seulement d'obtenir les facteurs suivants :

- Complexité fonctionnelle de l'application
- Taille de l'application exprimée en FFP
- Taille des programmes modifiés en FFP et en lignes de code

ID	Type de données	Description
1	Effort	Quantité des heures dépensées pour réaliser un projet de maintenance adaptative.
2	Effort_Sanalyse	Quantité des heures dépensées pour réaliser un projet de maintenance adaptative sans compter les heures de la phase d'analyse.
3	FFP_Total	Nombre total de FFP pour chaque projet de maintenance adaptative. Il s'agit de l'addition de tous les FFPs correspondants aux divers modules affectés lors de la réalisation de chaque projet.
4	FFP_Modification	Quantité réelle des FFP modifiés ou additionnés lors de la réalisation d'un projet de maintenance adaptative. Ceci en considérant seulement la partie affectée de chaque module lors d'une modification. Il s'agit d'une relation parmi le nombre total de FFPs, le nombre total des lignes de code et le nombre total des lignes de code modifiées.
5	#_Programmes	Quantité des modules ou des programmes affectés lors de la réalisation d'un projet de maintenance adaptative.
6	Complexité fonctionnelle	Variable qualitative qu'indique le niveau de complexité fonctionnelle de chaque projet de maintenance.
7	SLOC modifiées	Quantité de lignes de code modifiées, c'est-à-dire l'addition seulement des lignes de code ajoutées ou effacées dans un module.
8	SLOC totales	Quantité de lignes de code totales. Il s'agit de l'addition de toutes les lignes de code de chaque module intervenant lors de la réalisation d'un projet de maintenance adaptative.

TABLEAU 3.2 – NOM ET DESCRIPTION DES DONNÉES RECUEILLIES

CHAPITRE IV

4 Analyse statistique de la banque de données disponibles

L'analyse statistique des divers types de données recueillies à l'aide des modèles mathématiques a permis d'identifier et de modéliser les principales caractéristiques des facteurs intervenant dans la productivité pour les activités de maintenance chez IA, selon l'échantillon des 21 projets analysés.

En génie logiciel, tel que mentionné précédemment, plusieurs chercheurs ont défini la productivité comme le ratio entre les extrants produits par le processus en termes de la taille fonctionnelle du logiciel et les intrants (ressources) utilisés par le processus en termes de l'effort en heures-personne. Cette étude a donc pour but d'identifier, premièrement, la relation entre la taille fonctionnelle et l'effort dépensé pendant le processus de maintenance adaptative et ensuite d'observer la variation de cette relation lors de la considération d'autres facteurs ayant un effet sur la productivité au moyen de l'analyse statistique.

Quant à cette considération, seulement un des facteurs (qui pourraient augmenter ou diminuer l'effort requis pour réaliser une activité de maintenance) soulignés par les chercheurs tels que Lientz et Swanson (1981), Martin et McClure (1983), Jones (1986) et McClure (1992), a été identifié dans la banque de données disponibles. Ce facteur est la complexité fonctionnelle et technique du logiciel [MAY95].

La fiabilité de notre analyse statistique pour les diverses données recueillies sera sa capacité à reproduire des résultats de façon consistante dans la population en étude car la non-fiabilité des résultats réduirait la possibilité d'observer un effet significatif dans les modèles de productivité à établir. Pour cette raison, il est important de s'assurer que la population de l'échantillon en étude suit une distribution normale approximative afin d'utiliser des essais statistiques qui permettent de faire des inférences sur la moyenne et d'autres propriétés de la population des données recueillies. En plus, la distribution normale est une caractéristique importante car elle donne la validité à l'application statistique des méthodes paramétriques et à la dérivation des équations associant la taille fonctionnelle avec d'autres mesures de projet [ISBSG5].

Pour supporter cette hypothèse standard que des variables en étude ont une distribution normale, une méthode statistique a été utilisée au moyen de l'outil Winstat. Le test de Kolmogorov-Smirnov

lequel permet de calculer la distance maximale entre la courbe de fréquence cumulative des données et la meilleure courbe normale ajustée et détermine la signification de cette distance. En plus, le paramètre *p-valeur* est observé afin d'identifier les résultats non significatifs lorsque le *p-valeur* est supérieur à 0.05. Ainsi, nous concluons que la différence entre les deux courbes n'est pas significative et que les données sont distribuées normalement.

Il est nécessaire aussi de valider la pertinence de certaines valeurs qui se trouvent plus loin que les autres ainsi que de la moyenne de la population. Ces valeurs sont connues comme Outliers. Pour des analyses statistiques, il faut déterminer ces valeurs et les enlever afin d'obtenir une meilleure distribution gaussienne. Dans certains cas, il est possible que ces outliers ne soient pas représentatifs de la population mesurée.

Pour calculer les outliers, on a utilisé la méthode de Grubbs [BAR94]. Cette méthode s'appelle également la méthode ESD (Extreme Studentized Deviate).

En outre, un autre paramètre statistique est analysé, *l'écart type* (standard deviation) qui permet de mesurer la variabilité dans les données. Celui-ci assure que si les données suivent une distribution gaussienne, alors 68% des valeurs se trouvent à l'intérieur d'un écart-type de la moyenne (de chaque côté) et 95% des valeurs se trouvent à l'intérieur de deux écart-type de la moyenne.

4.1 Taille fonctionnelle étendue totale (FFP_Total) et taille fonctionnelle étendue modifiée (FFP_Modification)

La variable **FFP total** représente la taille fonctionnelle totale d'un projet d'entretien en termes des points de fonction étendus. Chaque projet comprend divers programmes du logiciel ABC, lesquels ont été mesurés au moyen de la méthode de FFP, et par conséquent, une valeur de FFP total a été assignée pour chacun d'eux.

Afin de calculer des FFPs correspondants aux additions ou aux modifications effectuées dans des projets de maintenance **FFP_Modification**, une relation équivalente a été utilisée parmi l'information de la taille fonctionnelle totale (FFP total), des lignes de code totales et des lignes de code modifiées. La relation équivalente des FFPs modifiés a la forme suivante :

$$\text{FFP_Modification} = \text{SLOC modifiées} * \text{FFP_Total} / \text{SLOC totales} \quad (\text{équation 4.1})$$

Cette information nous a permis de comparer la qualité de la relation entre les points de fonctions étendus et l'effort avec celle des lignes de code et l'effort.

Dans un premier temps, pour chaque projet de maintenance, la relation a été calculée en faisant la somme algébrique de toutes les données recueillies pour les variables impliquées dans l'équation précédente et ainsi obtenir le nombre total de FFP_Modification. On a également cherché à voir si d'autres calculs donnaient des résultats plus significatifs que la relation pour l'ensemble des observations de chaque projet. Une meilleure précision sur cette relation a été déterminée en appliquant séparément l'équation 4.1 pour toutes les données de chaque module d'un projet de maintenance et ensuite, faisant la somme algébrique de leurs FFP_Modification. En conséquence, le total de FFP modifié est l'addition des FFP_Modification de chaque module qu'intègre un projet de maintenance adaptative.

Les résultats des analyses statistiques, réalisées pour la variable FFP_Modification au moyen de divers outils, nous ont permis d'évaluer l'hypothèse standard de la distribution normale de la variable ainsi que la présence des Outliers dans la population.

La figure 4.1 montre la distribution cumulative de fréquences pour les points de fonction étendus lors de la modification. Cette distribution statistique employée par l'outil Winstat superpose la meilleure courbe de distribution normale adaptée, afin qu'on puisse également faire une comparaison visuelle et observer que la différence est significative selon des références statistiques.

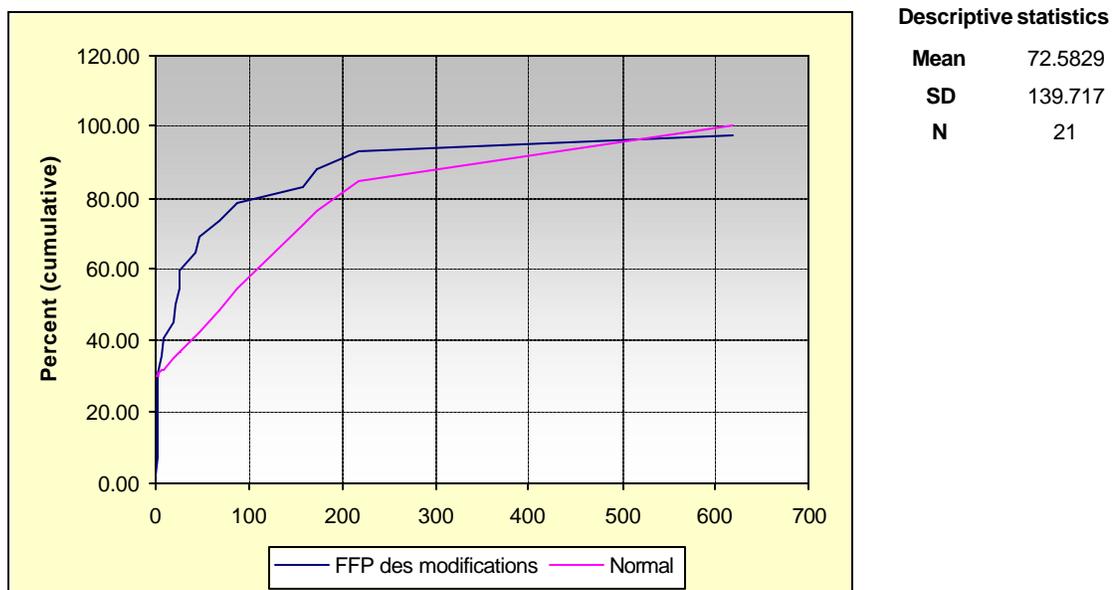


FIGURE 4.1 : FFP_MODIFICATION : DISTRIBUTION CUMULATIVE DE FRÉQUENCES

En plus, le test de Kolmogorov-Smirnov a été utilisé pour corroborer la distribution normale de cette variable. Le tableau 4.1 montre le résultat où D est la distance maximale mesurée entre la courbe de la distribution actuelle de la variable et la meilleure courbe adaptée à la distribution normale. P est la probabilité que la valeur D pourrait prendre par fluctuation aléatoire dans un échantillon d'une population normalement distribuée. Ainsi, un résultat de l'analyse avec un *p-valeur* non significatif (élevée), c.-à-d., un $P > 0.05$ nous permet de supposer que la variable est distribuée normalement. [WinStat] Ce qui n'est pas le cas dans notre échantillon pour la variable FFP_modification.

Variable	N	D	P
FFP_Modification	21	0.3025235609	0.04282111999

TABLE 4.1 : TEST DE DISTRIBUTION NORMALE POUR VARIABLES CONTINUES - KOLMOGOROV-SMIRNOV

Parallèlement, le test de Grubbs a été utilisé pour déterminer la présence des Outliers. Celui-ci indique que le projet WS0016 est plus loin que l'ensemble des données avec une valeur plus grande que trois écart-type. Le tableau 4.2 présente aussi la distance qui existe entre l'Outlier et la moyenne en termes de sigma, c'est-à-dire en termes de l'écart-type de la variable. Cependant, un autre paramètre est utilisé afin de valider la pertinence de ce résultat, P-valeur, lequel indique la probabilité de trouver au moins une valeur à cette distance de la moyenne dans un échantillon distribué normal. Ainsi, notre p-value est significatif indiquant qu'il n'est pas aussi probable de trouver cette valeur.

Variable	Project	Valeur	N * Sigma > 2	P < 0.05
FFP_Modification	WS0016	617.8387112	3.902564357	0.000998903

TABLE 4.2. OUTLIER

On a remarqué que l'exclusion du projet WS0016 donne une distribution plus près de la distribution normale selon le test de Kolmogorov-Smirnov. Celui-ci a présenté un *p-valeur* non significatif (élevé) après une nouvelle analyse, ce qui nous permet de supposer que la variable est distribuée normalement.

Variable	N	D	P
FFP_Modification	20	0.2737860752	0.0997259408

TABLE 4.3 : TEST DE DISTRIBUTION NORMALE POUR VARIABLES CONTINUES - KOLMOGOROV-SMIRNOV

4.2 Effort sans analyse

En ce qui concerne l'information sur l'effort dépensé par projet dans les différentes activités de maintenance, cela n'était pas disponible au complet. La documentation disponible pour la comptabilisation des efforts ne conserve pas l'effort dépensé dans la phase d'analyse par certains projets, ce qui empêchait d'établir leur effort approximatif pour le modèle de productivité. On a donc décidé de considérer l'effort dépensé par projets tout au long de cycle de développement sans inclure leur phase d'analyse. C'est ainsi qu'on a obtenu l'information concernant ce facteur pour l'ensemble des observations.

De même que pour la taille fonctionnelle, la figure 4.2 montre aussi la courbe d'une distribution normale pour la variable **Effort sans Analyse** (Effort_Sanalyse). Selon le test de Kolmogorov-Smirnov, la distribution de fréquence de l'effort sans analyse suit une variable distribuée normale. Cependant, le test de Grubbs indique que les projets WS0016 et WS0213 sont plus loin que l'ensemble des données. L'analyse de ces deux projets, à partir des données du tableau 4.4, a permis de constater que leur distance par rapport à la moyenne de la population représentée en multiple de sigma, (de l'écart-type de la variable) est plus grande. C'est ainsi que les projets WS0016 et WS0213 sont éloignés de la moyenne de plus de deux écart-type, ce qui veut dire que ces projets sont considérés comme Outliers pour la population en étude.

Variable	Project	Valeur	N * Sigma > 2	P < 0.05
Effort_Sanalyse	WS0213	724	2.223142217	0.241933288
Effort_Sanalyse	WS0016	956.25	3.179805388	0.015360774

TABLE 4.4. OUTLIER

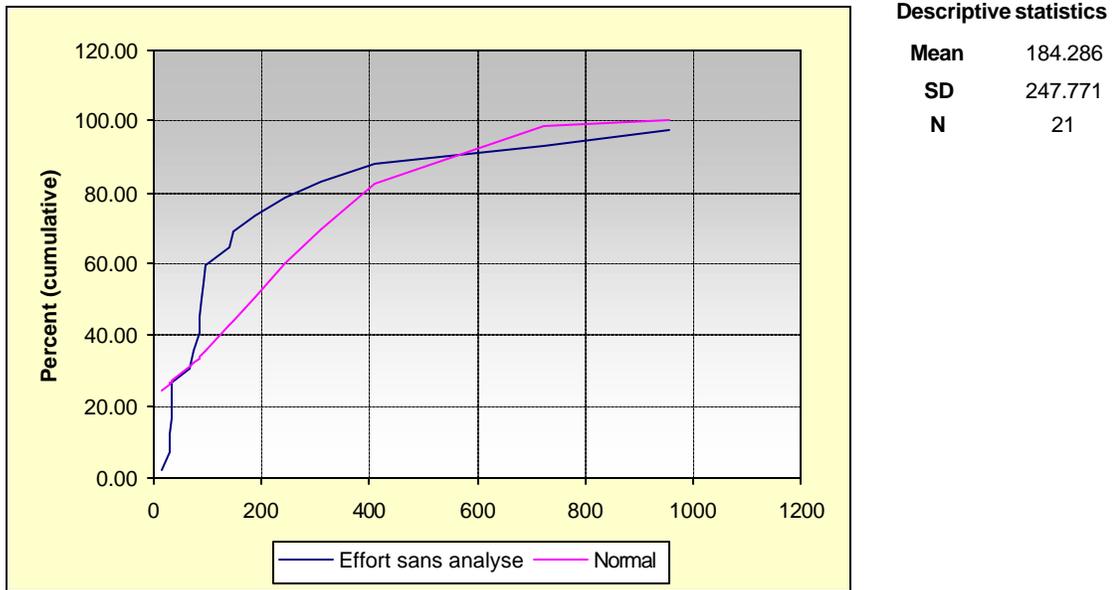


FIGURE 4.2 : EFFORT_SANALYSE: DISTRIBUTION CUMULATIVE DE FRÉQUENCES

En plus, il a été observé que l'exclusion des projets WS0016 et WS0213 a fourni une distribution plus près de la distribution normale selon le test de Kolmogorov-Smirnov. Il a présenté un p-valeur non significatif (élevé) qui nous a permis de supposer que la variable est distribuée normalement. Le tableau suivant montre les résultats du test de Kolmogorov-Smirnov pour les deux variables principales :

Variable	N	D	P
Effort_Sanalyse	19	0.25648788	0.164100618
FFP_Modification	19	0.278598083	0.104722809

TABLE 4.5 : TEST DE DISTRIBUTION NORMALE POUR VARIABLES CONTINUES - KOLMOGOROV-SMIRNOV

Le même processus de valider des Outliers et de corroborer la supposition standard de distribution normale pour les variables Effort_Sanalyse et FFP_Modification lors de l'analyse statistique a été également reproduit pour le reste des variables quantitatives recueillies dans l'étape de mesure du logiciel. Les résultats ont été similaires, par conséquent ils sont ajoutés dans l'annexe B afin de ne pas alourdir la lecture.

4.3 Complexité fonctionnelle

Un des objectifs du projet est de voir si certains facteurs qui, selon la littérature, affectent la productivité des activités de maintenance pourraient améliorer le modèle de productivité à une seule variable indépendante (les points de fonction étendus). C'est ainsi que nous avons obtenu l'information concernant la complexité fonctionnelle des projets, laquelle est représentée comme une variable qualitative.

On a fait une analyse de la complexité des projets en termes de types de fonctionnalités à additionner ou à modifier dans les programmes du logiciel ABC. Ces nouvelles caractéristiques étaient décrites dans les spécifications fonctionnelles de chacun des projets d'entretien. L'évaluation de la complexité par projet a été également basée sur le nombre total de FFP de l'ensemble de programmes affectés, et par conséquent, un nombre de complexité a été assigné selon des critères subjectifs des analystes chez IA Inc. et en fonction de leur expérience.

Calcul de la complexité fonctionnelle

Afin d'assigner un niveau de complexité fonctionnelle pour chacun des projets d'entretien, des facteurs qualitatifs ont été utilisés. Ces facteurs sont décrits comme des variables explicatives dans l'analyse de régression, permettant ainsi d'observer leur influence dans des modèles statistiques. Généralement, les facteurs comme la complexité sont représentés par des variables qualitatives ou indicatrices, lesquelles sont des valeurs numériques qui n'ont pas l'intention d'exprimer une valeur quantitative. Cependant, l'utilisation des variables indicatrices sert pour identifier des groupes ou des catégories dans les données à analyser. Ces variables sont utilisées chaque fois que des facteurs qualitatifs affectent la relation positive entre les variables lors de la régression linéaire. [CHA91]

La complexité dans notre cas d'étude est une variable qualitative, laquelle a été ajoutée subjectivement à chaque programme par une personne qualifiée chez IA Inc. pendant la phase de comptage et après révision des dossiers correspondant à chaque projet. Ainsi, en ce qui concerne la complexité fonctionnelle de chaque projet, on a sélectionné la complexité la plus grande et avec le plus de répétitions parmi le groupe de programmes d'un projet. Ceci car on considérait que les programmes avec cette complexité sont ceux qui ont eu plus d'influence sur l'effort dépensé lors de la réalisation du projet.

Au moment du comptage, les quatre niveaux de complexité suivants ont été utilisés avec des notations respectives 1, 3, 5 et 7, c'est-à-dire, non complexe, complexe, très complexe et extrêmement complexe. Toutefois, ces divers niveaux de complexité n'ont pas été efficaces lors de

l'analyse statistique à cause de la présence insuffisante de données dans certains de ces niveaux et du nombre de variables indépendantes générées par eux dans l'analyse de régression (trois variables indépendantes).

Une fois que l'échantillon d'observations fut passé de 21 projets à 19, après l'élimination de deux projets (considérés comme cas extrêmes selon des analyses précédentes à la section 4.1 et 4.2), il était intéressant de voir comme ces niveaux de complexité pourraient faire varier la solution statistique, c.-à-d., la rendre plus fiable pour notre étude. Cependant, cela n'était possible. Selon des références statistiques, une équation incluant quatre variables indépendantes (trois variables indépendantes pour représenter les quatre niveaux de complexité, plus la variable FFP) est considérée exagérée pour un échantillon de cette envergure (19 données), puisqu'il y avait des catégories de complexité avec seulement 1 ou 2 données.

Pour obtenir une équation plus simple et plus représentative en considérant l'échantillon de 19 données, seulement deux ou trois variables indépendantes étaient souhaitables. Dans ce but, des transformations ont été faites sur la complexité afin de considérer deux niveaux (complexité haute et complexité basse) au lieu de quatre niveaux en se basant sur le nombre total de FFP (l'ensemble de programmes affectés) par projets et leur nombre de complexité assignée.

Dans cette section, une telle analyse est présentée en considérant séparément les projets qui ont un nombre élevé de FFP et ceux qui ont un nombre mineur de FFP. Cette distinction se fait en utilisant comme référence l'écart-type de la population (Std. deviation) qui est de 60.15 dans l'échantillon de 19 projets. Cela est dû au fait qu'aucune donnée dans notre échantillon n'a eu un niveau de complexité égal à 5 ou à 7 avec un nombre total de FFP inférieur d'un écart-type, c.-à-d. inférieur à 60 FFP. On peut remarquer qu'il existe des données avec 360 et 964 FFP entre autres, ainsi on pourrait donc inférer qu'il existe une tendance de ne pas ajouter de complexité élevée aux projets si le nombre de FFP n'est pas assez considérable. C'est ainsi qu'on a obtenu l'information concernant ce facteur de complexité pour les divers projets. Lorsqu'un projet a eu plus de 60 FFP avec une valeur de complexité égale à 3, on a assigné à la variable de complexité la valeur de complexité basse. Par contre, si le projet a eu une valeur de complexité égale à 3 mais que le nombre de FFP est inférieur à 60, on a assigné à la variable de complexité la valeur de complexité haute. Le tableau 4.6 affiche la règle de transformation appliquée pour la complexité.

	FFP Total (1-60)	FFP Total 61+
non complexe (1)	B	B
Complexe (3)	H	B
très complexe (5)	H	H
Extrêmement complexe (7)	H	H

TABLE 4.6 : RÈGLE DE TRANSFORMATION POUR LA COMPLEXITÉ

Selon ce tableau, on pourrait dire, par exemple, qu'un programme ayant 25 FFP avec une valeur assignée de complexité égale à 3, la complexité qualitative est haute.

Cette analyse reste au niveau de l'expérience du mandataire au moyen d'essais empiriques lors de l'analyse statistique plutôt qu'au niveau scientifique, puisque aucune méthode scientifique ou systématique n'était utilisée pour valider l'hypothèse ou l'explication précédente. Ainsi, à partir d'essais empiriques, les chercheurs peuvent seulement confirmer ou infirmer des hypothèses sans leur donner une valeur scientifique. [NET83]

CHAPITRE V

5 Modèle d'estimation de productivité

Dans un premier temps, le but général est d'obtenir une relation simple entre la variation de la variable effort (mesurée en personnes-heures nécessaires pour la modification d'un projet) et la variable taille du projet (mesurée par le nombre de points de fonction étendus) ainsi que pouvoir inférer le coût de maintenance par modifier le système ABC.

Pour exprimer cette relation entre la taille fonctionnelle et l'effort, deux types de modèles ont été utilisés : le modèle du *coût unitaire moyen*, qui est décrit comme le rapport entre l'effort et le nombre des points de fonction ainsi que le *modèle statistique de la régression linéaire* avec les points de fonction étendus et les divers facteurs qui peuvent influencer la productivité des activités de maintenance.

5.1 Modèle du coût unitaire moyen

Le coût unitaire moyen est défini avec l'équation suivante:

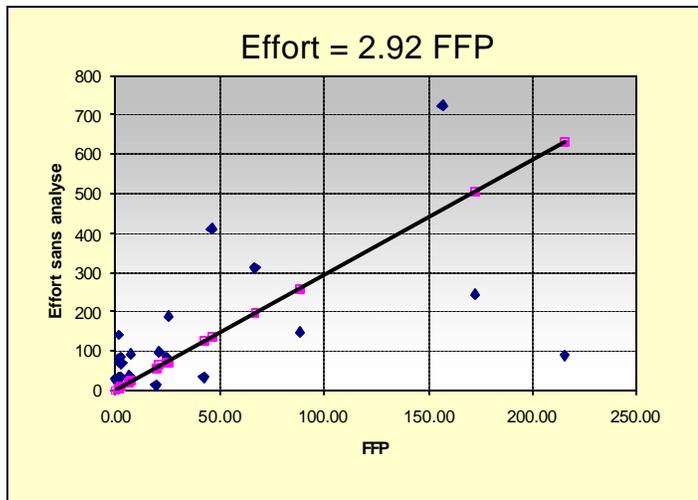
$$\text{Coût unitaire moyen} = \frac{1}{n} \sum_{i=1}^n \frac{\text{Effort}_i}{\text{FFP}_i} \quad (\text{équation 5.1})$$

Celui-ci donne comme résultat l'effort moyen nécessaire pour produire un point de fonction étendu, c.-à-d. le nombre moyen d'heures par FFP.

Un modèle de productivité construit à partir du coût unitaire moyen a donc la forme suivante :

$$\text{Effort} = \text{FFP} * \text{coût unitaire moyen}$$

Considérant notre échantillon d'observations (19 projets), le coût unitaire moyen est de 2.92 heures par point de fonction étendu (FFP) . Le modèle de productivité proposé à partir de ce coût unitaire moyen a donné le résultat suivant :



Descriptive statistics	
Équation	Effort = 2.92 * FFP
ERM(%)	575
PRED (+/-25%)	
# Projets à l'intérieur	1
%	5

FIGURE 5.1 : COMPARAISON GRAPHIQUE ENTRE LA DROITE DU COÛT UNITAIRE ET LES DONNÉES RÉELLES (EFFORT)

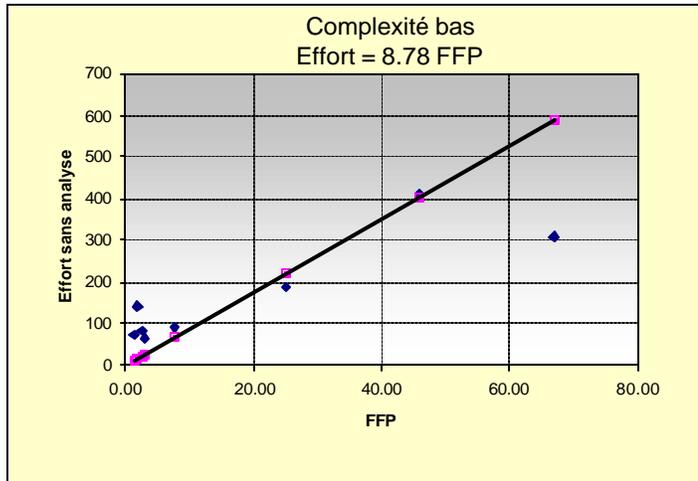
5.1.1 Qualité du modèle de productivité du coût unitaire

Si on analyse les résultats considérant les divers critères pour déterminer la qualité de ce modèle, on observe que : l'ERM (erreur relative moyenne) de 575% est extrêmement supérieure à 25%. Cette erreur est très élevée par rapport au critère proposé par Conte et al.(1986), lequel définit une ERM égale ou plus petite que 25% comme acceptable. En plus, seulement un projet (5% des efforts calculés) est à l'intérieur de l'écart de fiabilité des estimations de +/- 25% par rapport aux efforts réels. En conséquence, ce modèle, pour notre échantillon d'observations, ne donne pas des résultats significatifs.

On a donc décidé de séparer et d'observer l'échantillon en étude par rapport aux sous-groupes de projets qui ont des caractéristiques similaires afin de construire des modèles de productivité plus fiables et satisfaisants. Le critère sélectionné pour diviser l'échantillon concernait la complexité impliquée dans les projets.

5.1.2 Modèle du coût unitaire par la complexité basse

Il existe huit projets impliqués dans cette section. Le modèle du coût unitaire moyen par complexité basse est représenté de la forme suivante :



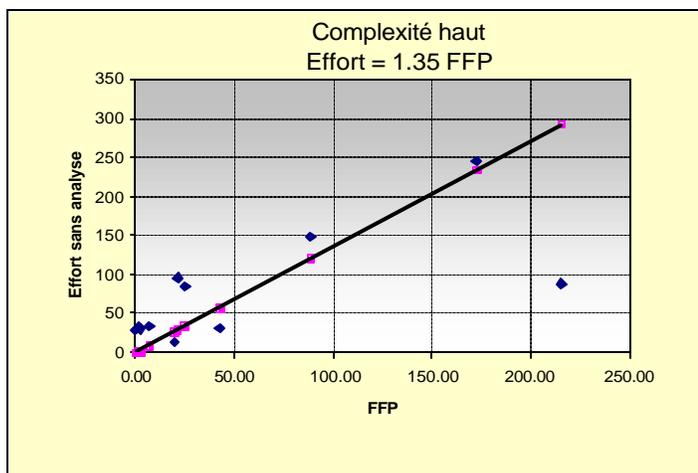
Descriptive statistics	
Équation	Effort = 8.78 * FFP
ERM(%)	220
PRED (+/-25%)	
# Projets à l'intérieur	2
%	25

FIGURE 5.2 : COMPARAISON GRAPHIQUE DE LA DROITE DU COÛT UNITAIRE ET LES DONNÉES RÉELLES (EFFORT) PAR PROJETS AVEC COMPLEXITÉ BASSE

Ce modèle de productivité présente une ERM égale à 220%, ce qui est extrêmement supérieur à 25% et seulement deux projets (25% des efforts calculés) sont à l'intérieur de l'écart de fiabilité des estimations de +/- 25% par rapport aux efforts réels. En conséquence, ce modèle ne donne pas des résultats significatifs.

5.1.3 Modèle du coût unitaire par la complexité haute

Il existe onze projets impliqués dans cette section. Le modèle du coût unitaire moyen par complexité haute est représenté de la forme suivante :



Descriptive statistics	
Équation	Effort = 1.35 * FFP
ERM(%)	880
PRED (+/-25%)	
# Projets à l'intérieur	2
%	18

FIGURE 5.3 : COMPARAISON GRAPHIQUE DE LA DROITE DU COÛT UNITAIRE ET LES DONNÉES RÉELLES (EFFORT) PAR PROJETS AVEC COMPLEXITÉ HAUTE

Ce modèle de productivité présente une ERM égale à 880%, ce qui est extrêmement supérieur à 25% et seulement 18% des efforts estimés sont à l'intérieur d'un intervalle de +/- 25% par rapport aux efforts réels. Ce modèle ne donne pas des résultats satisfaisants.

Dans cette étude, les résultats de ces modèles sont très semblables : des ERM très élevées et très peu de projets à l'intérieur de l'écart de fiabilité des estimations de +/-25%. On peut conclure que les modèles construits à partir du coût unitaire moyen n'atteignaient pas les critères de bons modèles de productivité.

On a donc décidé de poursuivre l'étude en utilisant l'analyse de régression, et ainsi définir une relation linéaire simple en tenant compte de l'ensemble des observations existantes. Si le résultat n'est pas encore significatif, on séparera et observera l'échantillon en étude par rapport aux sous-groupes de projets qui ont des caractéristiques similaires afin de construire des modèles de productivité plus fiables et satisfaisants.

5.2 Modèle de régression linéaire

Le modèle statistique le plus simple à utiliser et à comprendre après le coût unitaire est le modèle de régression linéaire simple. C'est le modèle qui permet de mesurer l'influence d'une variable quantitative sur une autre variable quantitative. On a utilisé la technique de la régression linéaire afin d'exprimer la relation qui existe entre la taille et l'effort de développement pour les projets d'entretien. Ainsi, la première sous-section présente les modèles construits avec les points de fonction étendus comme la seule variable indépendante. Ensuite, la deuxième sous-section présente les modèles de régression construits à partir des divers critères qui pourraient affecter la productivité des activités de maintenance adaptative en terme de leur taille fonctionnelle et de l'effort.

De même que pour les modèles construits à partir du coût unitaire moyen, on s'est servi de la taille fonctionnelle mesurée pour chacune des applications de référence et de la taille fonctionnelle modifiée représentée par la variable FFP_Modification. Toutefois, seulement les principaux schémas retenus ont été inclus dans cette section. Les résultats pour toutes les analyses sont indiqués sous forme de tableau à l'annexe B.

Dans un premier temps, et selon les recommandations statistiques [NET83] de travailler avec le modèle le plus simple possible, on commencera par établir une relation entre la taille du projet qui est mesurée par le nombre de FFP et l'effort sans analyse mesuré en personnes-heures pour la

maintenance du système. Pour la justification d'utiliser l'effort sans analyse, se référer au point 4.2. Une première analyse de régression avec toutes les données a produit le résultat suivant:

5.2.1 Estimation de l'effort en fonction de la taille FFP modifiée, une seule variable.

Dans un premier temps, le modèle de régression linéaire simple a été construit en combinant les deux variables suivantes : le total de points de fonction étendus de chaque projet comme variable indépendante et l'effort comme variable dépendante. Selon les résultats obtenus, en tenant compte des 19 projets de l'échantillon, le modèle de la régression linéaire simple est représenté par la droite suivante :

$$Y = A + B \cdot X \rightarrow Y = 0.61 \cdot X + 91.29 \quad (R^2 = 0.12 ; n = 19) \quad (\text{équation 5.2})$$

Le modèle fourni par l'équation (1) est représenté à la figure 5.4 par son diagramme de dispersion avec la ligne de régression correspondante. Ceci n'est évidemment pas le résultat d'une représentation linéaire fortement positive lorsqu'on analyse la qualité du modèle donnée par son coefficient de détermination, $R^2 = 0.12$. Seulement 12% de la variabilité totale de l'effort dépensé dans les projets est expliqué par la variation dans leur nombre de points de fonction étendus.

On pourrait penser qu'il n'existe pas une relation linéaire complète entre les deux variables analysées, donc le modèle linéaire simple est insuffisant.

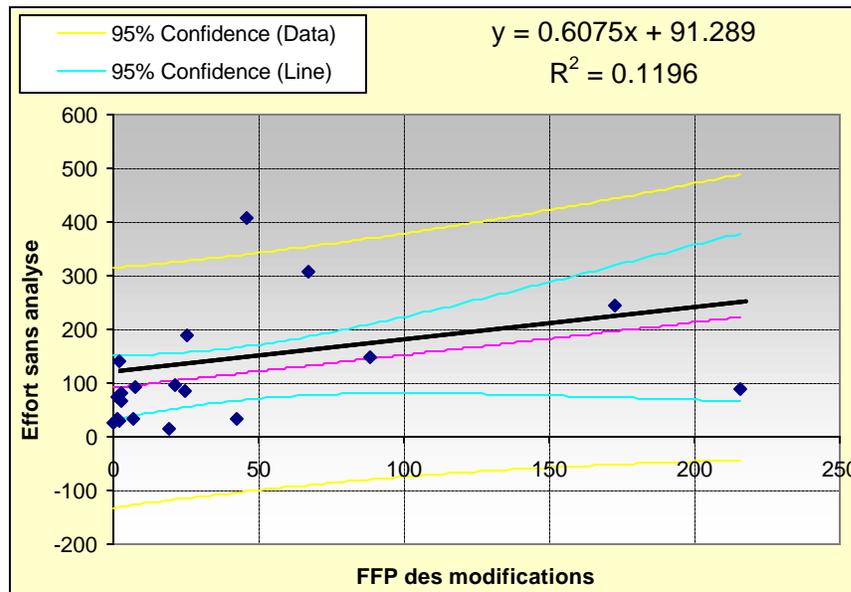


FIGURE 5.4 : RÉGRESSION LINÉAIRE SIMPLE SANS LES DEUX OUTLIERS (19 PROJETS)

Pour faire suite, nous avons plutôt cherché l'existence d'une relation non linéaire entre ces deux variables. Dans le tableau 5.1, nous avons présenté les résultats obtenus. R est le coefficient de corrélation entre les valeurs réelles de Y et les valeurs calculées par l'équation donnée. Les valeurs élevées de R (maximum = 1,0) indiquent un bon ajustement. R-Square, en termes statistiques, est le pourcentage de la variance dans la variable dépendante qui peut être expliquée par l'équation donnée.

X-Variable: FFP des modifications

Y-Variable: Effort sans analyse

		N	A	B	R	R-Square
$Y = A + B \cdot X$	Linear	19	91.289	0.607	0.35	0.12
$Y = A \cdot X^B$	Power	19	43.808	0.245	0.50	0.245
$Y = A \cdot e^{(B \cdot X)}$	Exponential	19	63.067	0.006	0.39	0.15
$Y = A + B \cdot \ln(X)$	Logarithmic	19	44.121	29.29	0.51	0.26
$Y = A + B/X$	Hyperbolic 1	19	132.463	-48.330	0.32	0.10
$Y = 1 / (A + B \cdot X)$	Hyperbolic 2	19	0.022	-8.8E-05	0.31	0.09
$Y = X / (A + B \cdot X)$	Hyperbolic 3	19	0.007	0.016	0.28	0.08

TABLE 5.1 : RELATION STATISTIQUE DÉFINIE ENTRE L'EFFORT ET LE FFP SELON DIFFÉRENTES FONCTIONS

Dans ces conditions, les différents modèles non linéaires ne présentaient aucun avantage significatif par rapport au modèle linéaire; malgré qu'il existe dans certains cas une amélioration mineure.

On a ensuite décidé d'ajouter des variables dans l'équation simple, afin d'identifier des facteurs pouvant affecter la productivité des activités de maintenance, lesquels pourraient améliorer le modèle de productivité. Le critère le plus important utilisé est la complexité fonctionnelle car il peut être un facteur qui affecte la productivité des activités de maintenance. Cependant, la pondération de la complexité dans chacun des projets a été établie par des jugements d'experts, donc de nature subjective, tel que mentionné précédemment.

5.3 Modèle à plusieurs variables indépendantes

Pour déterminer l'influence des divers facteurs, on a construit des modèles de régression à plusieurs variables indépendantes considérant les FFP et les différents critères mesurés qui peuvent influencer l'analyse. Le but est de voir s'il y a une amélioration de la qualité de la

relation entre la taille fonctionnelle et l'effort avec cette information, et par conséquent voir si ces variables ont un effet dans la relation de la productivité.

La régression linéaire multiple a été réalisée au moyen du logiciel Winstat, lequel a donné pour chacune des variables dans l'équation, leur coefficient estimé ainsi que l'intervalle de confiance pour le coefficient. On peut assurer à 95% que la valeur réelle du paramètre en étude se trouve à l'intérieur des limites indiquées pour l'intervalle de confiance lorsque l'échantillon est sélectionné à l'intérieur d'une population qui suit une distribution normale. P indique la signification du terme donné dans l'équation.

5.3.1 Modèle avec la quantité des programmes affectés

Chaque projet d'entretien est composé d'un ou plusieurs modules ou programmes qui sont affectés lors de sa réalisation. Cette différence sur le nombre de modules dans chaque projet pourrait affecter notre relation. Il s'agit donc de construire un modèle de régression qui tient compte des deux variables indépendantes suivantes : le FFP des modifications et la quantité de programmes modifiés :

$$y = ax + bz + c \rightarrow Y = a \text{ FFP} + b \text{ \#Prog Affecté} + C$$

$$Y = 0.78 \text{ FFP} - 3.62 \text{ \#Prog Affecté} + 98.5$$

Multiple Regression [Subset]

X-variable: FFP des modifications
Prog. affectés

Y-Variable: Effort sans analyse

Method: Direct

Summary

	N	R	R-Square	Std.Error
normal	19	0.351097237	0.12326927	104.9424091
corrected		0.116952678	0.013677929	

Equation

	Coefficient	95% Conf. (±)	Std.Error	T	P
Constant	98.50825392	85.26637866	40.22037206	2.449212896	0.026214014
FFP des modifications	0.782042726	1.672197646	0.788779968	0.991458654	0.336219748
# Prog. affectés	-3.621568004	29.60459571	13.96456461	-0.259339844	0.79868026

Analysis of variance

	Sum of Squares	Degrees of Freedom	Mean Square	F	P
Regression	24774.8274	2	12387.4137	1.124808481	0.349083911
Residue	176206.5476	16	11012.90923		
Total	200981.375	18	11165.63194		

Ce modèle n'offre pas une amélioration significative en comparaison à celui du modèle général lors de l'analyse de régression simple. Par contre, le même résultat du R2 a été obtenu (12%).

5.3.2 Modèle additif avec la complexité

Après avoir essayé plusieurs modèles avec le FFP total, le nombre de programmes modifiées et la quantité de lignes de code, la variable de complexité a été analysée. Donc, un deuxième modèle de régression qui tient compte de cette variable qualitative indépendante a été construit.

Selon la section 4.3, on a déjà deux niveaux de complexité qui sont acceptables pour un échantillon de 19 projets :

Complexité = 1 -> Si la complexité fonctionnelle du projet est « Haute »
 Complexité = 0 -> Si la complexité fonctionnelle du projet est « Basse »

On veut obtenir une relation avec une équation similaire à la suivante:

$$y = ax + bz + c \text{ (équation 5.3)}$$

$$\text{où si } z = 0 \rightarrow y = ax + c$$

$$\text{ou si } z = 1 \rightarrow y = ax + (b + c)$$

Alors, notre équation doit avoir la représentation suivante en utilisant la complexité:

$$Y = aX + bZ + c$$

$$\text{Effort} = a \text{ FFP} + b \text{ Complexité} + c$$

$$\text{Si complexité} = 0 \rightarrow \text{Effort} = a \text{ FFP} + c$$

$$\text{Si complexité} = 1 \rightarrow \text{Effort} = a \text{ FFP} + (c+b)$$

Le résultat obtenu de la régression linéaire multiple selon l'outil Winstat est :

Multiple Regression [Subset]

X-variable: FFP des modifications
complexité

Y-Variable: Effort sans analyse

Method: Direct

Summary

	N	R	R-Square	Std.Error
normal	19	0.674291063	0.454668438	82.7652883
corrected		0.621692844	0.386501993	

Equation	Coefficient	95% Conf. (±)	Std.Error	T	P
Constant	25.89693714	65.65485305	30.96956454	0.836206047	0.41535328
FFP des modifications	0.91905104	0.719087236	0.339195315	2.709503934	0.01546786
complexité	126.1175416	85.27080101	40.22245809	3.135500602	0.006386534

Analysis of variance	Sum of Squares	Degrees of Freedom	Mean Square	F	P
Regression	91379.88784	2	45689.94392	6.669974301	0.00782141
Residue	109601.4872	16	6850.092948		
Total	200981.375	18	11165.63194		

L'équation 5.3 représente l'équation générale de la représentation de l'effort selon le FFP des modifications et la complexité pour un échantillon de 19 projets. Nous avons donc construit un modèle qui considère de façon séparée les niveaux de complexité. Cela veut dire que l'on donne la même importance à chaque niveau de complexité dans l'explication de l'effort requis pour réaliser un projet de maintenance. L'équation générale peut être transformée aux équations 5.4 et 5.5 dépendant du niveau de complexité du projet à mesurer :

$$\text{Effort} = 0.92 \text{ FFP} + 126.12 \text{ Complexity} + 25.9 \text{ and } R^2 : 0.46 \text{ and } n = 19 \quad (\text{équation 5.3})$$

$$\text{If complexity} = 0 \rightarrow \text{Effort} = 0.92 \text{ FFP} + 25.9$$

$$\text{IF complexity} = 1 \rightarrow \text{Effort} = 0.92 \text{ FFP} + 152.02$$

Nous avons dans la figure 5.5 la représentation de ces équations. Le modèle obtenu pour le facteur « complexité » présente un coefficient de détermination $R^2 = 0.46$, supérieur à celui du modèle de régression linéaire simple, mais qui n'est pas suffisant pour notre étude.

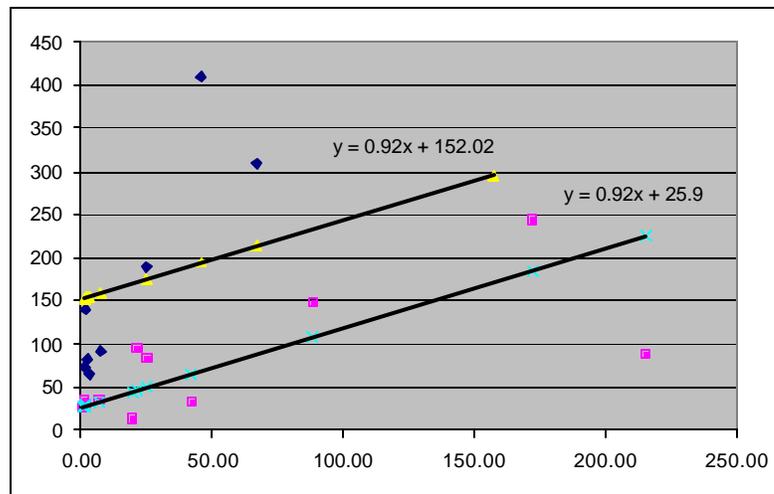


FIGURE 5.5 : REPRÉSENTATION GRAPHIQUE DE L'ÉQUATION 5.4 ET 5.5

5.3.3 Modèle multiplicatif avec la complexité

Selon le modèle représenté par l'équation 5.3, l'effet du FFP n'a pas été influencé par les valeurs de la complexité. Alors, l'effet d'augmentation en relation aux deux variables est déterminé par des constantes additives. Cela veut dire que l'effet du FFP est mesuré sans dépendance de la complexité. Pour améliorer notre modèle, on a additionné une nouvelle variable selon la recommandation du livre [NET83], laquelle est le produit de la complexité et le FFP (complexité * FFP). L'inclusion de cette variable conduit le modèle à reconnaître l'effet multiplicatif de ces deux variables dans la relation positive de la taille fonctionnelle et l'effort. Ce qui produit aussi l'élimination de l'effet de parallélisme représenté dans le modèle additif.

Le modèle qui considère l'explication précédente à la représentation suivante :

$$Y = \alpha X + \beta Z + \gamma (X * Z) + \mu$$

$$\text{Effort} = \alpha \text{ FFP} + \beta \text{ Complexité} + \gamma (\text{FFP} * \text{Complexité}) + \mu \quad (\text{équation 5.6})$$

Si complexité = 0 → Effort = α FFP + μ
 Si complexité = 1 → Effort = $(\alpha + \gamma)$ FFP + $(\mu + \beta)$

On peut donc observer que cette nouvelle représentation montre l'effet de la complexité sur la taille fonctionnelle. La complexité représentée par γ a une influence sur le comportement de la variable FFP, occasionnant une modification sur la pente et la constante de la droite lors de l'analyse de ses valeurs 0 et 1.

Le résultat obtenu de la régression linéaire multiple selon l'outil Winstat est :

$$\text{Effort} = 0.64 \text{ FFP} + 41.94 \text{ Complexité} + 3.85 (\text{C} * \text{FFP}) + 41 \text{ avec un } R^2 \text{ de } 0.75 \text{ et } n = 19 \quad (\text{équation 5.7})$$

Multiple Regression
[Subset]

X-variable: FFP des modifications
 Complexité
 FFP*comp
Y-Variable: Effort sans analyse
Method: Direct

Summary

	N	R	R-Square	Std.Error
normal	19	0.86624454	0.750379603	57.83254595
corrected		0.83693221	0.700455524	

Equation

Coefficient	95% Conf. (±)	Std.Error	T	P
-------------	---------------	-----------	---	---

Constant	41.00474734	46.75417723	21.93486835	1.869386526	0.081218825
FFP des modifications	0.639643938	0.524578486	0.246107636	2.599041417	0.020137389
complexité	41.94470183	73.48704949	34.4766789	1.216610856	0.242549768
FFP*comp	3.85204763	1.947769581	0.913802185	4.215406455	0.000749229

Analysis of variance	Sum of Squares	Degrees of Freedom	Mean Square	F	P
Regression	150812.3244	3	50270.77481	15.03041444	8.63574E-05
Residue	50169.05057	15	3344.603372		
Total	200981.375	18	11165.63194		

Le modèle au complet présente un coefficient de détermination $R^2 = 0.75$, ce qui est supérieur à celui du modèle à une seule variable. En plus, le coefficient de la variable qualitative lié à l'ajustement de la complexité et celui de la variable FFP sont statistiquement significatifs, p value < 0.05 .

Le résultat avec la décomposition de la complexité est :

Si complexité = 0 \rightarrow Effort = 0.64 FFP + 41 avec un R^2 de 0.47 et $n = 8$ (équation 5.8)
 Si complexité = 1 \rightarrow Effort = 4.49 FFP + 82.94 avec un R^2 de 0.78 et $n = 11$ (équation 5.9)

La figure 5.7 révèle que l'effort des projets dépend de leur taille fonctionnelle avec la variable complexité comme un facteur important lors de l'estimation.

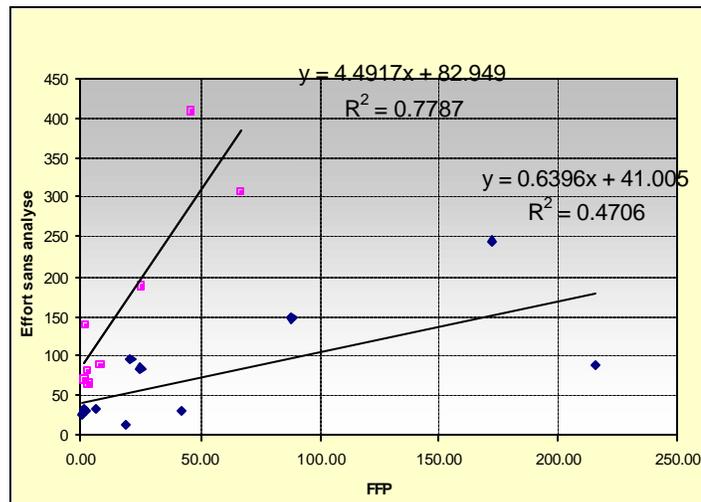


FIGURE 5.6 : REPRÉSENTATION GRAPHIQUE DE L'ÉQUATION 5.8 ET 5.9

De plus, une analyse des résidus a été faite pour donner plus d'informations sur la qualité du modèle construit. Généralement, dans l'analyse de régression statistique nous essayons de

modéliser le comportement d'une variable basée sur les valeurs des autres. Si le modèle est considéré bon, alors la différence entre la valeur prévue par le modèle et la valeur réelle de la variable en étude ne devrait pas dépendre de la position de la valeur dans la série globale [WINSTAT]. Ainsi, nous sommes intéressés à démontrer la normalité du résiduel et l'aspect aléatoire des résidus.

Dans notre exemple, la p-valeur de Kolmogorov-Smirnov de 0,49 indique la normalité du résidu, malgré que le projet plus extrême (WS0004) soit situé à plus de trois écart-type du reste de l'échantillon.

En plus, pour valider l'aspect aléatoire de la distribution des données lors de l'analyse de résidus, nous devons produire un diagramme de dispersion « scatterplot » de la variable indépendante contre les résidus. Cela veut dire, une représentation graphique de la variable quantitative indépendante (taille fonctionnelle) contre des résidus afin d'obtenir un ensemble de points distribués approximativement égal par les deux côtés de l'axe des abscisses, pour l'intervalle entier des valeurs de la variable indépendante. [WINSTAT] Cela peut être corroboré au moyen de deux tests : Le test de « Turning Points » et le test « Rank ». Dans les deux cas, la différence n'a pas été significative (Tableau 5.2), ainsi nous pouvions dire que les résidus étaient organisés aléatoirement.

Randomness

Test of turning points

	N	Turning pts.	Expected	Z	P
Residuo	19	10	11.3333	-0.7628	0.4456

Pairwise rank test

	N	Pairs with (X2 > X1)	Expected	Z	P
Residuo	19	80	85.5	-0.1283	0.8979

TABLE 5.2 : TEST DE « TURNING POINTS » ET DE « RANK » POUR VALIDER L'ASPECT ALÉATOIRE DE RÉSIDUS

Cependant, l'étendue de l'échantillon analysé pour la complexité basse n'est pas assez grand, c.-à-d., il existe moins de variations dans les observations de la variable dépendante à l'intérieur de l'étendue de la variable indépendante. Cela occasionne que le modèle n'offre qu'un pourcentage d'explication mineur pour la complexité basse, donc des résultats également non significatifs.

D'ailleurs, si nous regardons attentivement les données avec complexité basse, nous pouvons observer que le projet avec le plus grand FFP (WS0004) dépensait un effort assez bas lors de sa réalisation, ce qui n'est pas cohérent avec le reste de notre échantillon. Par conséquent, une analyse a été faite avec l'exclusion de ce projet dans le but de voir son influence dans notre modèle.

Ainsi, le résultat obtenu de la régression linéaire multiple représentée par l'équation 5.6 ($Y = \alpha X + \beta Z + \gamma (X * Z) + \mu$) avec l'élimination du projet WS0004 qui semble un projet différent, est représenté par l'équation suivante :

$$\text{Effort} = 1.25 \text{ FFP} + 56 \text{ Complexité} + 3.24 (C * \text{FFP}) + 26.94 \text{ avec un } R^2 \text{ de } 0.845 \text{ et } n = 18 \quad (\text{équation 5.10})$$

Le résultat avec la décomposition de la complexité est :

$$\text{Si complexité} = 0 \rightarrow \text{Effort} = 1.25 \text{ FFP} + 26.94 \text{ avec un } R^2 \text{ de } 0.47 \text{ et } n = 8 \quad (\text{équation 5.11})$$

$$\text{Si complexité} = 1 \rightarrow \text{Effort} = 4.49 \text{ FFP} + 82.94 \text{ avec un } R^2 \text{ de } 0.84 \text{ et } n = 10 \quad (\text{équation 5.12})$$

Multiple Regression [Subset]

X-variable: FFP des modifications

Complexité

FFP*comp

Y-Variable: Effort sans analyse

Method: Direct

Summary

	N	R	R-Square	Std.Error
normal	18	0.918990305	0.844543181	47.14705688
corrected		0.900683632	0.811231006	

Equation

	Coefficient	95% Conf. (±)	Std.Error	T	P
Constant	26.94735467	39.71317695	18.51558186	1.455387947	0.167617322
FFP des modifications	1.250881498	0.621085747	0.289570487	4.319782418	0.000706013
complexité	56.0020945	61.15790948	28.51381749	1.964033561	0.069699009
FFP*comp	3.24081007	1.659408153	0.77367035	4.188877176	0.000910015

Analysis of variance

	Sum of Squares	Degrees of Freedom	Mean Square	F	P
Regression	169063.2815	3	56354.42717	25.35238753	6.40791E-06
Residue	31119.82961	14	2222.844972		
Total	200183.1111	17	11775.47712		

Le modèle au complet présente un coefficient de détermination $R^2 = 0.845$, ce qui est supérieur à celui du modèle antérieur (équation 5.6). La variable FFP est statistiquement significative avec ce p value < 0.05.

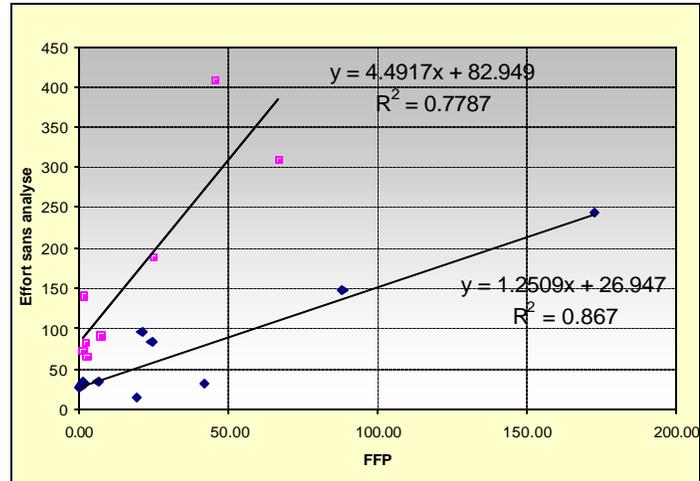


Figure 5.6 : Représentation graphique de l'équation 5.11 et 5.12

5.3.4 Qualité du modèle

Si on analyse les résultats considérant les divers critères pour déterminer la qualité de ce modèle, on observe que : Le modèle au complet présente un coefficient de détermination $R^2 = 0.845$, ce qui est supérieur à celui du modèle antérieur (équation 5.6) et plus proche de la valeur 1. Quant à l'ERM (erreur relative moyenne) de 40%, celle-ci est acceptable malgré qu'elle est supérieure à 25% (critère proposé par Conte et al.(1986)). De plus, selon la fiabilité de prédiction (PRED), dix projets (55.56% des efforts calculés) sont à l'intérieur de l'écart de fiabilité des estimations de +/- 25% (critère proposé par Conte et al.(1986)) par rapport aux efforts réels. Cependant, il existe des valeurs très proches à 25% ce qui fait qu'on atteint le 77.78% des observations avec 35% de fiabilité de prédiction. En conséquence, ce modèle donne des résultats plus significatifs pour notre échantillon d'observations.

Équation	R^2	ERM(%)	PRED (+/- 25%)		PRED (+/- 30%)		PRED (+/- 35%)	
			# Projets	%	# Projets	%	# Projets	%
5.6	75	0.51	10	52.63	12	63.16	14	73.68
5.10	84	0.40	10	55.56	12	66.67	14	77.78

TABLE 5.3 : COMPARAISON DE LA QUALITÉ DES MODÈLES DES ÉQUATIONS 5.6 ET 5.10

CHAPITRE VI

6 Sommaire des résultats

On peut affirmer que les objectifs du projet ont été atteints: toutes les activités planifiées ont été réalisées permettant ainsi la construction et la validation de plusieurs modèles de productivité pour des activités de maintenance adaptative du logiciel ABC chez IA.

La validation des modèles construits a permis d'analyser la qualité de la relation entre les points de fonction étendus et l'effort ainsi que l'analyse de l'intervention d'autres facteurs, telle que la complexité fonctionnelle des projets.

Cette relation dans certains modèles où les points de fonction étendus étaient retenus comme variables indépendantes, et l'effort comme variable dépendante, a été très faible; ce qui ne nous permettait pas de la considérer pour être utilisée comme modèle de productivité. Toutefois, la qualité de cette relation était améliorée lorsqu'on a considéré le facteur de complexité comme une autre variable indépendante, facteur reconnu comme ayant une influence sur la productivité selon la littérature.

Dans un premier temps, une hypothèse nulle, qui a énoncé formellement les relations attendues entre la taille fonctionnelle d'un logiciel et son effort, a été faite au début pour la valider et ainsi atteindre les objectifs du projet.

À cet effet, trois types de modèles ont été construits: des modèles avec le coût unitaire moyen, des modèles de régression simple avec le point de fonction étendu comme la seule variable indépendante et des modèles de régression multiples avec le point de fonction étendu et d'autres variables qui peuvent influencer le modèle. Les résultats des modèles construits à partir du coût unitaire moyen ont révélé que l'écart entre l'effort estimé et l'effort réel est très élevé et que la fiabilité de ces modèles est très basse. Ceci veut dire que, pour l'ensemble de données analysées, le rapport entre l'effort et les FFP n'est pas adéquat pour être utilisé comme modèle de productivité.

D'un autre côté, pour les modèles de régression simple, les résultats des analyses statistiques ont indiqué que, pour l'ensemble des observations analysées, il n'existe pas une relation entre l'accroissement de la taille fonctionnelle des projets de maintenance adaptative et l'accroissement

de l'effort requis pour les produire, quelque soit la forme supposée de la relation (Linear, Power, Exponential, Logarithmic, Hyperbolic).

Dans ces conditions, la banque de données a été divisée en sous-ensembles avec des caractéristiques similaires, afin de voir si, à l'intérieur de chaque sous-ensemble, la relation entre la taille fonctionnelle et l'effort donne une amélioration significative. Plusieurs modèles ont été construits et analysés. Cependant, le seul modèle fiable et significatif était le modèle de régression multiple avec le point de fonction étendu et la complexité comme variable indépendante. Ce modèle a permis d'identifier pour l'ensemble des observations lors de l'analyse statistique, plusieurs caractéristiques des projets de maintenance.

Pour atteindre ce résultat, tout d'abord, un modèle a été produit en utilisant un modèle de régression multiple additif. Celui-ci a amélioré notre résultat de l'analyse de régression simple, malgré que le niveau de fiabilité n'ait pas été atteint selon les références statistiques. Avec les résultats obtenus, on peut donc penser que le modèle varie selon la représentation des niveaux de complexité au niveau de l'intersection de leur droite et non au niveau de leur pente (relation de parallélisme). Au niveau de la taille fonctionnelle, la plupart des projets ont obtenu entre 0.33 et 617.84 points de fonction étendus et la moyenne pour l'ensemble des observations était de 72.58. Quant à l'effort, la moyenne d'heures requises pour réaliser un projet était de 184.28 (27.5 - 956.25).

Ensuite, un modèle de régression multiplicatif a été construit afin de déterminer s'il n'existe pas une relation d'interaction entre les variables plutôt qu'une relation de parallélisme pour les niveaux de complexité identifiés. Tandis que les projets de maintenance de notre étude se composent généralement de deux types de complexité (haute et basse), ce modèle a satisfait la représentation de seulement un d'eux, c.-à-d., la complexité haute, avec un niveau de fiabilité acceptable pour l'ensemble des observations. Cette représentation pourrait peut-être s'expliquer par le fait que l'ensemble des projets avec complexité basse a l'influence d'un projet qui semble un cas extrême. Ce projet est celui qui a la taille la plus grande parmi l'échantillon avec un effort dépensé exagérément bas. Ainsi, cette situation nous a permis de supposer que l'information concernant ce projet n'était pas assez fiable et nous avons procédé à son élimination.

De cette façon, le dernier modèle présenté nous a permis de valider notre hypothèse et ainsi d'atteindre nos objectifs. On a construit un modèle de productivité pour les activités de maintenance adaptative qui permet d'estimer l'effort selon la quantité des points de fonction étendus et le niveau de complexité du projet. Le niveau de qualité du modèle a été assez acceptable avec un coefficient de détermination $R^2 = 0.845$. De plus, même si notre modèle n'est pas dans l'intervalle (écart) de

fiabilité de prédiction (PRED) et d'erreur relative moyenne (ERM) proposé par Conte et al. (1986), le modèle est assez proche de ces limites.

Toutefois, un aspect important qu'il faut remarquer dans cette activité est celui de l'analyse de la complexité fonctionnelle des projets. On pourrait penser que l'assignation subjective de cette variable affectera la qualité du modèle de productivité proposé lors de son utilisation, puisque la valeur de cette variable dépend du niveau d'expertise de l'analyste. Il est donc recommandé d'appliquer aussi une mesure standard pour ce facteur afin d'obtenir des améliorations dans les modèles de productivité construits, ce qui emmènera à un processus plus cohérent et uniforme pour l'analyse de productivité pour les activités de maintenance adaptative.

Finalement, la méthode des points de fonction étendus nous a permis de quantifier la taille fonctionnelle d'un logiciel en temps réel pour les activités de maintenance adaptative chez IA avec un niveau de fiabilité acceptable. Cela permettra à la compagnie d'améliorer leur processus d'estimations d'efforts requis pour ces activités ainsi qu'au laboratoire LRGL pour suivre leurs études avec d'autres ensembles de données pour confirmer des comportements similaires lors de l'utilisation de la méthode augmentant ainsi l'efficacité des analyses de productivité dans la maintenance adaptative.

7 Bibliographie

- [ABR93] Alain Abran et Nguyenkim H. Measurement of the Maintenance Process from a Demand based Perspective. Software maintenance: research and Practice, Vol. 5 1993, p 63-90
- [ABR93-2] Alain Abran et Pierre N. Robillard. Analyse comparative de la fiabilité des points de fonction comme modèle de productivité. Rapports de recherche. ICO Janvier 1993. p. 16-24
- [ABR96] Alain Abran et Pierre N. Robillard. Function Point Analysis : An empirical study of its Measurement Processes. IEEE Transactions on Software Engineering. Vol.22, No. 12. Décembre 1996.
- [ABR97] Alain Abran, Marcela Maya, Jean-Marc Desharnais et Denis St-Pierre. Adapting Function Points to Real-time Software. American Programmer, Novembre 1997.
- [BAR94] V. Barnett and T. Lewis, Wiley. Outliers in Statistical Data. 3rd edition, 1994.
- [CON86] Conte S.D., Dunsmsore D.E., Shen V.Y., 1986, Software Engineering Metrics and Models, Benjamin/Cummings Publishing.
- [BOU99] P. Bourque, R. Dupuis, A. Abran, J. W. Moore, and L. L. Tripp. The Guide to the Software Engineering Body of Knowledge. IEEE Software, vol. 16, pp. 35-44, November/December 1999.
- [BEL80] David A. Belsley, Edwin Kuh, Roy E. Welsch. Regression Diagnostics. Identifying Influential data and Sources of Collinearity. Wiley series in probability and mathematical statistics. 1980.
- [DES97] Desharnais, Jean-Marc. Mesure de la taille fonctionnelle des logiciels temps réel. Montréal. 1997
- [HEL99] Roger Heller, An Introduction to Function point Analysis. Q/P Management Group, 1999
- [HIL99] Thomas B.Hilburn, Iraj Hirmanpour, Soheil Khajenoori, Richard Turner, Abir Qasem. A Software Engineering Body of Knowledge Version 1.0. TECHNICAL REPORT, CMU/SEI-99-TR-004, ESC-TR-99-004, April 1999
- [ISBSG5] International Software Benchmarking Standards Group Ltd, Release 5. March 1998.
- [NET83] Jhon Neter, Willian Wasserman, Michael H. Kutnen. Applied Linear Regression Models. ISBN 0-256-02547-9. 1983
- [KEC98] Nihal Kececi, Ming Li et Carol Smidts. Function Point Analysis: An Application to a Nuclear Reactor Protection System. 1998.
- [MAY95] Maya M. La technique étendue des points de fonction dans la construction des modèles de productivité en maintenance adaptative. UQAM – Rapport d'activité de synthèse. décembre 1995

- [MIL98] Everald E. Mills, Metrics in the software engineering curriculum. Computer Science/Software Engineering Department, Seattle University. Annals of Software Engineering 6, 1998.
- [MOR98] Pam Morris et Jean-Marc Desharnais, Measuring all the Software not just what the Business Uses. IFPUG Fall Conference, Orlando, Florida, September 21-25, 1998
- [CHA91] Samprit Chatterjee, Bertran Price. Regression Analysis by Example. Second edition. Wiley series in probability and mathematical statistics. 1991
- [OLI99] Serge Oligny, Alain Abran et Denis St-Pierre. Improving Software functional size Measurement. UQAM Software Engineering Management Research Laboratory. 1999
- [TRU99] Sylvie Trudel, Sommaire sur les PFE et le projet de mesure du logiciel ABC. Mai 1999
- [VER92] Verner J., Graham T., 1992, A Software Size Model, IEEE Transaction on software Engineering, vol. 18, no. 4, Avril, pp, 265-278.
- [WEI85] S. Weisberg, Applied Linear Regression, John Wiley & Sons, New York, NY, 1985.

Annexe A – Banque de données

# STR	# Prog. affectés	Effort	Effort SAnalyse	# FFP total	FFP Modification	FFP Modification (OLD)	SLOCC totales	SLOCC modifiées	Complexité (4 niveaux)	Complexité (2 niveaux)
PG0004	14	288.5	87.75	360	215.63	177.69	3748	1850	1	0
PG0016	53	965.25	956.25	984	617.84	482.30	13796	6762	1	0
PG0020	7	148	148	123	88.58	61.08	2044	1015	3	0
PG0059	1	108.5	65.75	40	3.06	3.06	327	25	3	1
PG0061	1	115.5	82.75	16	2.59	2.59	222	36	3	1
PG0065	3	34	34	18	6.78	6.54	575	209	1	0
PG0092	3	96	96	120	21.14	23.58	1206	237	3	0
PG0093	4	84.25	84.25	88	24.82	24.53	1137	317	3	0
PG0094	8	31.5	31.5	151	42.23	21.49	2227	317	3	0
PG0150	6	439	409.25	75	45.84	28.29	1604	605	5	1
PG0153	1	30.5	30.5	36	2.09	2.09	431	25	1	0
PG0158	2	141.5	140	7	1.83	0.61	575	50	3	1
PG0172	4	371.25	308.5	125	67.05	78.97	1393	880	7	1
PG0174	7	247	244.5	232	172.54	141.10	2103	1279	3	0
PG0192	4	188.5	188.5	53	25.08	18.77	1022	362	3	1
PG0202	5	33.75	33.75	44	1.37	1.03	1148	27	1	0
PG0203	1	75.75	72.75	22	1.42	1.42	202	13	3	1
PG0207	1	32.25	27.5	6	0.33	0.33	110	6	1	0
PG0208	1	91.75	90.75	54	7.78	7.78	444	64	3	1
PG0210	1	13.75	13.75	37	19.28	19.28	495	258	3	0
PG0213	18	890.25	724	248	156.97	117.08	4819	2275	7	1

Annexe B : Résultats des analyses statistiques

B1 : Modèles du coût unitaire moyen

Description	# Projets	Coût Unitaire	ERM(%)	PRED (+25%)	
				# Projets à l'intérieur	%
Ensemble d'observation	19	2.92	575	1	5
Complexité base	8	8.78	220	2	25
Complexité haute	11	1.35	880	2	18
# Prog. Affectés < 9	18	19.7	537	1	5.6
# Prog. Affectés < 6	14	24.4	416.6	2	14.2

B2 : Modèles de régression à une seule variable indépendante (FFP)

Description	Type	Type nom	# Projets	A	B	R	R ²
Ensemble d'observation	$Y = A + B \cdot X$	Linear	21	82.13	1.41	0.81	0.66
	$Y = A \cdot X^B$	Power	21	37.28	0.36	0.65	0.42
	$Y = A \cdot e^{(B \cdot X)}$	Exponential	21	69.52	0.01	0.62	0.38
	$Y = A + B \cdot \ln(X)$	Logarithmic	21	-26.89	76.95	0.64	0.41
	$Y = A + B/X$	Hyperbolic 1	21	216.05	-98.43	0.27	0.07
	$Y = 1 / (A + B \cdot X)$	Hyperbolic 2	21	0.02	0	0.35	0.12
	$Y = X / (A + B \cdot X)$	Hyperbolic 3	21	0.01	0.05	0.31	0.09
Observation avec 19 projets	$Y = A + B \cdot X$	Linear	19	91.29	0.61	0.35	0.12
	$Y = A \cdot X^B$	Power	19	43.81	0.25	0.49	0.25
	$Y = A \cdot e^{(B \cdot X)}$	Exponential	19	63.07	0.01	0.39	0.15
	$Y = A + B \cdot \ln(X)$	Logarithmic	19	44.12	29.29	0.51	0.26
	$Y = A + B/X$	Hyperbolic 1	19	132.463	-48.33	0.32	0.10
	$Y = 1 / (A + B \cdot X)$	Hyperbolic 2	19	0.022	0	0.31	0.09
	$Y = X / (A + B \cdot X)$	Hyperbolic 3	19	0.01	0.02	0.28	0.08
Observation avec 18 projets	$Y = A + B \cdot X$	Linear	18	93.29	0.46	0.23	0.05

	$Y = A * X^B$	Power	18	45.18	0.22	0.44	0.19
	$Y = A * e^{(B*X)}$	Exponential	18	63.97	0.01	0.28	0.08
	$Y = A + B*\ln(X)$	Logarithmic	18	47.47	26.60	0.45	0.21
	$Y = A + B/X$	Hyperbolic 1	18	124.42	-43.53	0.3	0.09
	$Y = 1 / (A + B*X)$	Hyperbolic 2	18	0.02	0	0.24	0.06
	$Y = X / (A + B*X)$	Hyperbolic 3	18	0.01	0.02	0.26	0.07

]

B3 : Modèles de régression à plusieurs variables indépendantes

Description	Type	# Projets	R ²	F	P-valeur	A	B	C	D
FFP des modifications, # Prog. affectés	$Y = A X + B Y + C$	19	0.12	1.12	0.35	98.5	0.78	-3.62	
FFP des modifications, Complexité	$Y = A X + B Y + C$	19	0.46	6.67	0.01	0.92	126	25.9	
FFP des modifications, Complexité, FFP*comp	$Y = A X + B Y + C Z + D$	19	0.75	15	8.64E-05	0.64	41.9	3.85	41
FFP des modifications, Complexité, FFP*comp	$Y = A X + B Y + C Z + D$	18	0.84	25.4	6.41E-06	1.25	56	3.24	26.9