

Université de Montréal

Travaux pratique 2

PAR

Guillaume Pilon (1055643)  
Jérémie Tousignant (1038501)

Falculté des arts et sciences

Travail présenté à Michalis Famelis  
Dans le cadre du cours IFT3913  
Qualité logiciel

18 mars 2022

## Introduction

Nous avons effectué une étude de cas, afin d'étudier des métriques de la qualité du logiciel, tel que le NCLOC, DCP, NOCom, WMC. Toutes les données récoltées et calculées proviennent des classes Java appartenant au dossier "main" du projet JFreechart. Dans cette étude, nous avons exploré l'hypothèse que les classes modifiées plus de 10 fois sont mieux commentées que celles modifiées moins de 10 fois. De plus, nous avons tenté d'observer si le WMC avait une relation linéaire avec divers métrique mentionné précédemment.

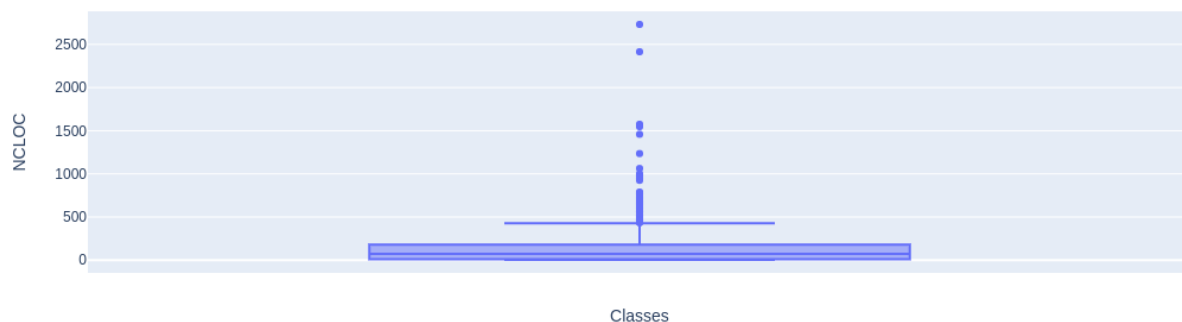
Toutes les analyses et les graphique sont aussi disponible en format interactif dans dans le Jupyter Notebook sur le dépôt git:

[https://github.com/GuiPil/TP\\_IFT3913\\_H22/tree/main/TP2/analyse](https://github.com/GuiPil/TP_IFT3913_H22/tree/main/TP2/analyse)

## T1 : Visualisez

Nous avons tout d'abord explorer chacune des métriques soit : NCLOC, DCP, NOCom et WMC.

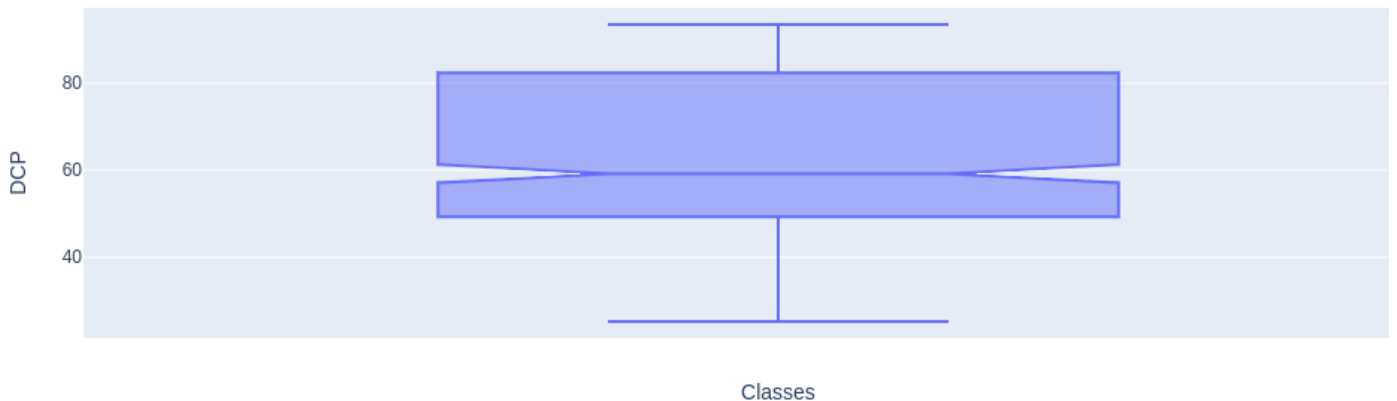
### NCLOC:



**Figure 1 :** Graphique à boîte à moustache des NCLOC en fonction de leurs classes respective et où le NCLOC est le nombre de lignes de codes non-vides et ne sont pas des commentaires.

La moyenne de ligne de code qui ne sont pas des lignes de commentaires par classe (N=632) est 145.04 avec un écart-type de 240.29. On observe sur le graphique que la moustache supérieure est à 429 et l'inférieure n'existe pas, avec un q3 de 180, un q1 de 12, un max de 2732 et un min de 4. On observe que la majorité des classes comporte moins de 180 lignes de code, avec plusieurs classes aberrantes (N=32) en lien avec cette métrique dépassant les 429 lignes de codes non commentées ou vides.

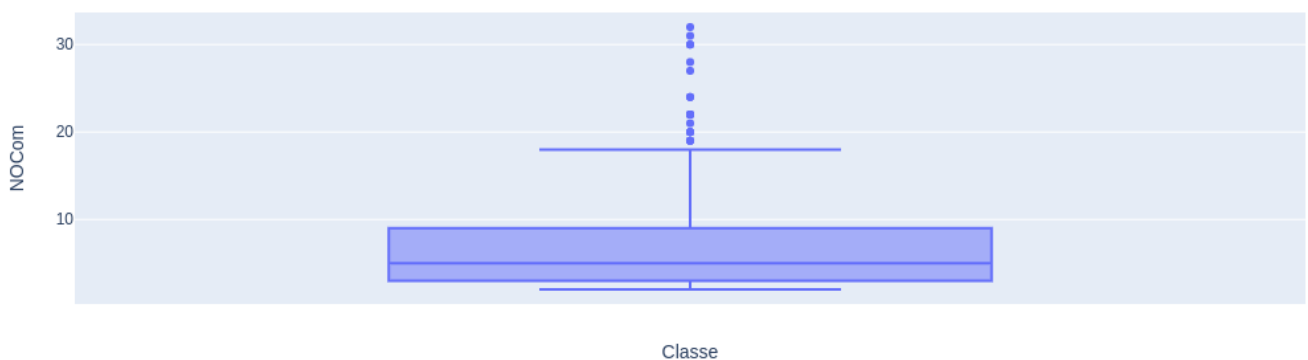
## DCP



**Figure 2 :** Graphique à boîte à moustache des DCP en fonction de de leurs classes respective et où le NCLOC est la densité de commentaire (CLOC/LOC) en pourcentage.

La moyenne de densité de commentaire par classe (N=632) est de 63.47 avec un écart-type de 17.87. On observe sur le graphique un max à 93.44 et un min à 25.24. On peut remarquer une distribution plus symétrique qu'avec le NCLOC, cependant, il faut prendre en considération que la données en pourcentage, donc contraint les données entre 0 et 100. On observe que 50% des données se situent entre 49.3 et 82.3.

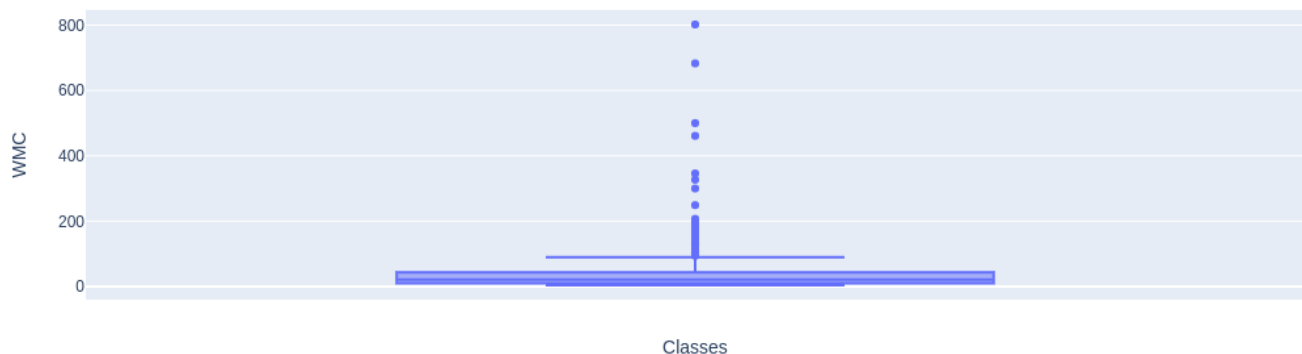
## NOCom



**Figure 3:** Graphique à boîte à moustache des NOCom en fonction de de leurs classes respective et où le NOCom est le nombre de commits dans l'historique git .

La moyenne de NOCom par classe (N=632) est de 6.74 avec un écart-type de 4.76. Les moustaches se situent à 2 et 18, mais il existe un max à 32. La distribution est asymétrique avec quelques données aberrantes (N=19) dépassant les 18 commits. On observe que 50% des données se situent entre 3.0 et 9.0.

## WMC



**Figure 4:** Graphique à boîte à moustache des WMC en fonction de de leurs classes respective et où le WMC est la mesure de complexité *Weighted Methods per Class* .

La moyenne de WMC est de 38.81 avec un écart-type de 62.71 pour les classes étudiées (N=632). Les moustaches sont à 90 et 4 avec un max à 802. La distribution est orientée vers le bas du graphique avec des données qui dépassent 100. On observe que 50% des données se situent à 10.0 et 43.0. On observe également plusieurs données extrêmes (N=48) dépassant une valeur WMC de 90.

## **T2 : Évaluer l'hypothèse**

### Choix de l'étude

Nous avons effectué un étude de cas sur les classes du projet jfreechart afin d'évaluer l'hypothèse selon laquelle les classes qui ont été modifiées **plus de 10 fois sont mieux documentées** que celles qui ont été modifiées moins de 10 fois.

Il est important de définir ce qu'on attend par une classe **mieux commentée**. Une classe qui est mieux commentée aura généralement un ratio élevé de commentaire par rapport aux nombres de ligne de code.

Cette question peut donc être répondu avec l'aide de la densité de commentaire (DCP), une métrique disponible dans notre jeu de données.

### Énoncé des hypothèses

On cherchera donc à rejeter l'hypothèse nulle:

$H_0$  : Il n'y a pas de différence concernant la densité de commentaires pour les classes avec plus de 10 commits et ceux avec moins de 10 commits

Et valider si on accepte l'hypothèse alternative :

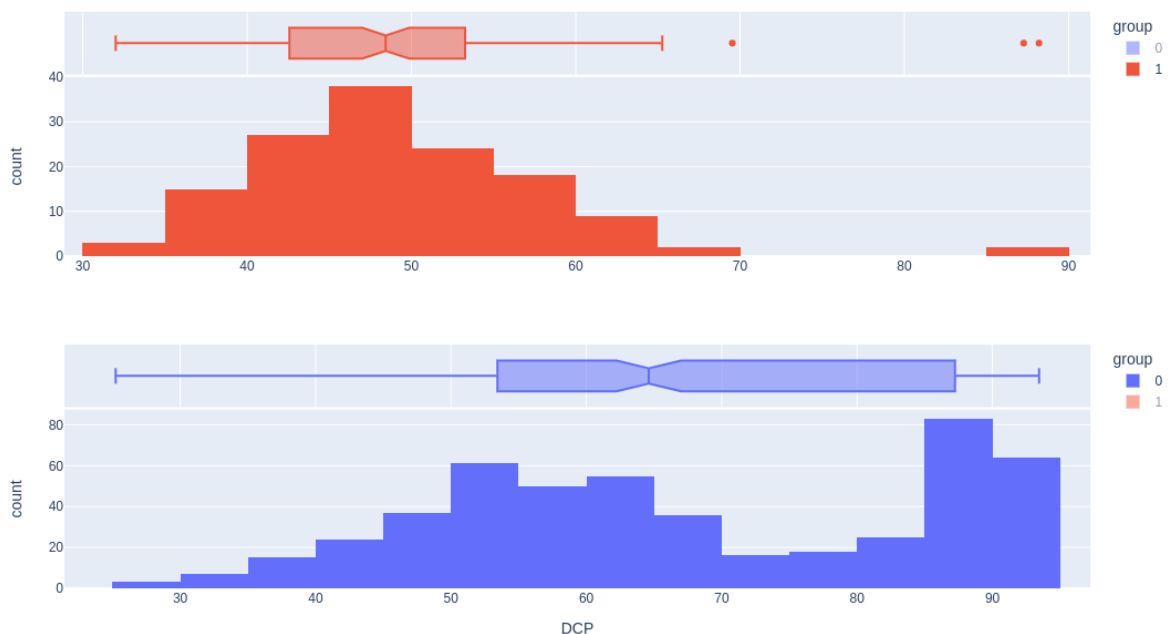
$H_A$  : Les classes avec plus de 10 commits ont une densité plus élevée de commentaire.

### Interprétation et généralisation des résultats

L'intégralité des analyses sont à l'intérieur du jupyter notebook fourni dans le zip.

Dans le but de tester les hypothèses, nous allons utiliser le nombre de commits comme indicateur de modification d'une classe et nous allons utiliser le DCP comme indicateur d'une classe qui est mieux commenté, donc plus celui-ci est élevé.

Afin de tester les hypothèses énoncées précédemment, nous avons séparé les classes en deux catégories. D'une part, les classes avec 10 commits et plus et de l'autre les classes avec moins de 10 commits. Nous avons décidé de visualiser la distribution des deux groupes et ensuite tester s'il y a une différence significative entre les deux groupes en utilisant un test-t à deux échantillons.



**Figure 5:** Histogrammes du DCP par classe, le groupe 1 (orange) sont les classes avec un nombre de commit < 10 et le groupe 0 (bleu) sont les classes avec  $\geq 10$  commits.

En observant la figure 5, on remarque que les deux distributions ne semblent pas normales. En observant davantage les relations (voir J.N.) et par des tests de Shapiro sur le groupe 0 ( $p = 2.30e-06$ ) et sur le groupe 1 ( $p =$

7.82e-15), avec de  $p < 0.05$  on peut pas rejeter l'hypothèse null, donc on ne peut conclure que les jeux de données sont ont une distribution normale.

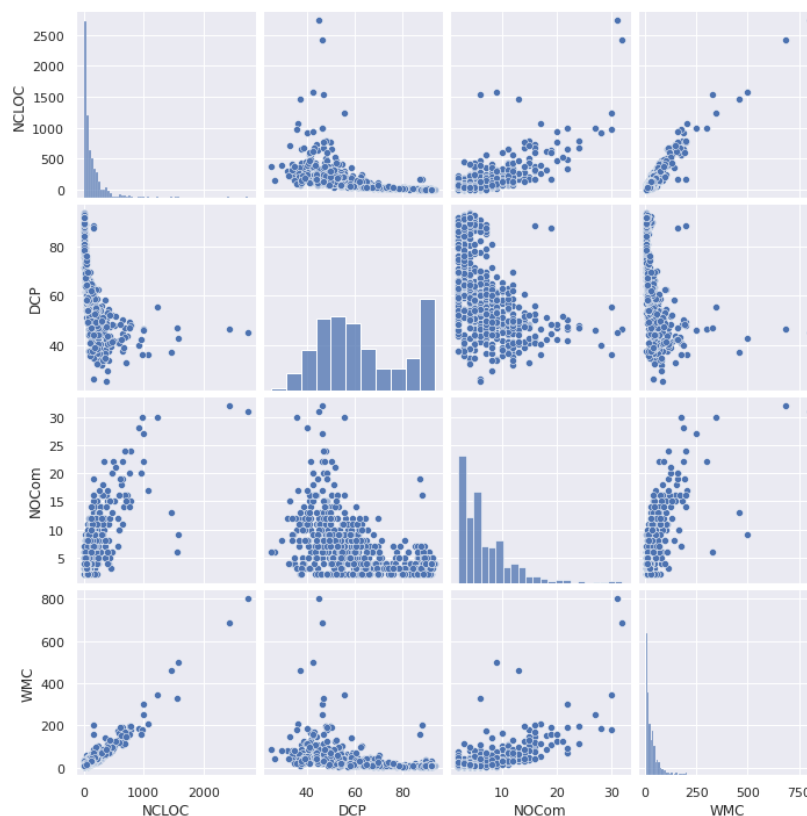
Ensuite nous avons testé les moyennes avec un test-t ( $p = 1.99e-29$ ), donc on peut rejeter l'hypothèse nulle du test-t. En observant les moyennes de DCP des deux groupes soit 67.52 pour le groupe 0 ( $\geq 10$ ) et 48.99 pour le groupe 1 ( $< 10$ ). On observe que la moyenne du groupe 0 est plus élevée.

### Conclusion et menaces à la validité

Le test d'hypothèse peut être rejeté, mais il faut faire attention, car les deux distributions n'étaient pas normales, ce qui vient biaiser le résultat. Les distributions semblent toutefois tendre vers une validation de notre hypothèse. Il s'agit simplement d'une étude de cas, donc il faut faire attention de ne pas généraliser les résultats à tous les projets. Pour ce projet, on peut quand même affirmer, avec une certaine prudence, que les classes avec plus de modifications semblent bénéficier d'une meilleure qualité de commentaire. En effet, les classes avec plus de modification bénéficient d'une meilleure densité de commentaire ( $p < 0.001$ ).

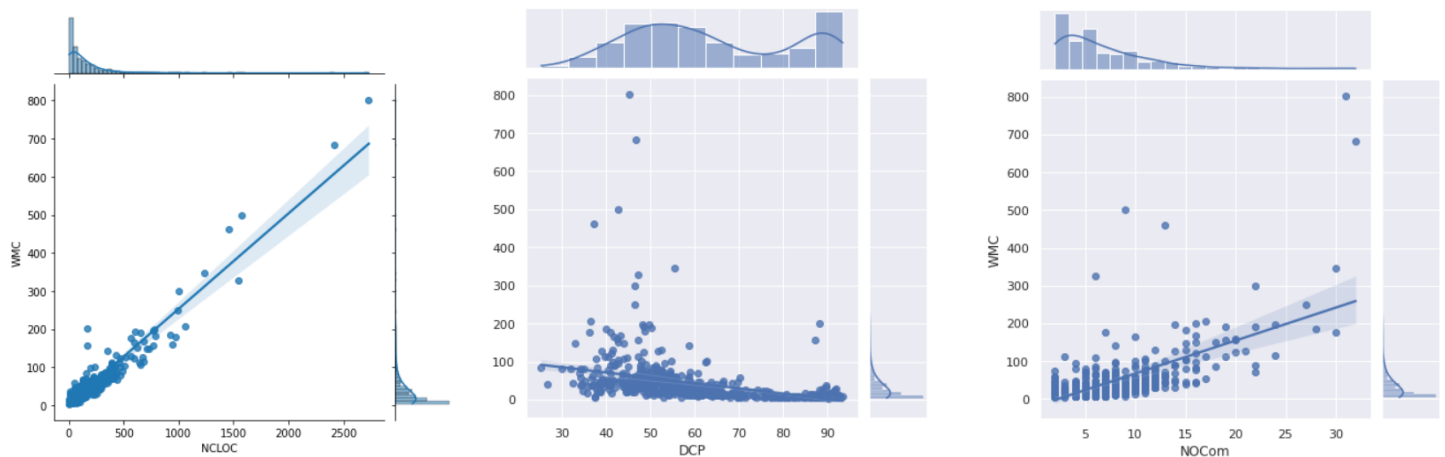
## T3 : Étudier les corrélations

Nous avons tout d'abord exploré toutes les associations de deux métriques (voir T3 `sns.pairplot(dfT3)` dans le J.N.).



**Figure 6:** Ensemble de relation possible avec le métriques fournit dans le document *jfreechart-stats*.

Suite à cette première investigation, nous avons identifié trois associations qui présentent potentiellement une corrélation : WMC vs NCLOC, WMC vs DCP et WMC vs NOCom. Tous les autres, ne présentent pas d'attribut qui pourrait nous faire penser qu'il existe une corrélation. En outre les nuages de point ne semblent pas du tout linéaires.



**Figure 6:** Les trois relations soit WMC en fonction du NCLOC, du DCP et du NOCom, où les métriques sont exprimé comme suit : NCLOC est le nombre de ligne qui ne sont pas des commentaires, DCP la densité de commentaire (CLOC/LOC) et NOCom le nombre de commits.

Dans les trois graphiques de la figure 7, on peut potentiellement voir les corrélations. Il est intéressant de noter que les trois sont le WMC en fonction de quelque chose.

## T4 : Évaluer les hypothèses

### Choix de l'étude

Pour cette étude, nous avons effectué une étude de cas dans laquelle, on s'intéresse à la relation qu'il pourrait exister entre la complexité pondérée des méthodes par classe (WMC) et différentes variables. En raison d'exploration préliminaire (T3), nous faisons l'hypothèse qu'il pourrait exister une relation linéaire entre WMC et NCLOC, WMC et DCP ainsi que WMC et NOCom.

## Énoncé des hypothèses

Alors nous allons explorer les hypothèses suivante :

$H_0$  : Il n'existe pas de relation linéaire entre WMC et X.

$H_A$  : Il existe une relation linéaire entre WMC et X.

Où X est l'une des métriques suivantes : NCLOC, DCP et NOCom.

Pour réaliser cette étude, nous allons utiliser comme données les scripts présentés dans le projet jfrechar-stats. Il s'agit donc d'une étude de cas afin d'évaluer si dans ce projet les hypothèses sont validées.

Chacune des hypothèses consiste à vérifier s'il y a bien une corrélation linéaire entre WMC et une autre variable. Ainsi, la même démarche pourra être appliquée dans les trois cas.

Nous allons d'abord évaluer la distribution des variables pour s'assurer de la normalité des données. Dans le cas où la relation distribution est normale, nous allons évaluer la force de la relation en utilisant le test de corrélation de *pearson*. Sinon, nous utiliserons le rang des données avec le test de corrélation de *spearman*.

Finalement, dans le cas d'une relation forte, nous exprimons cette relation avec une régression linéaire.

## Interprétation et généralisation des résultats:

En visualisant les données graphiquement (J.N.), nous remarquons qu'il est peu probable que les relations soient normales. Le résultat du test de Pearson pour chacune des relations confirme que chacune de celles-ci ne sont pas normales ( $p = 1.97e-37$  pour le NCLOC,  $p = 1.87e-17$  pour le DCP et  $p = 2.04e-25$  pour le NOCom ).

**Tableau 1** : Tableaux des degrés de corrélation entre le WMC et soit : NCLOC, DCP ou NOCom, suite à un test de Spearson "two sided"

Relation	Force de la relation ( $R^2$ )
(WMC, NCLOC)	0.918
(WMC, DCP)	-0.771
(WMC, NOCom)	0.676



Alors, afin de tester la corrélation nous avons effectué des test de spearman pour chacune des relations. Les p-values était de 2.11e-255 pour le NCLOC, 1.16e-125 pour le DCP et 1.98e-85, alors pour chacune on peut rejeter l'hypothèse null et admettre qu'il y a un relation monotone entre le WMC et les métriques étudié. Pour chacun des tests nous avons aussi les degrés de corrélations (voir Tableau 1).

On coefficient de corrélation  $\pm 0.6$  peut-être considéré comme forte, nous avons aussi calculer une régression linéaire pour chacune de métrique (Tableau 2).

**Tableau 2 :** Tableaux des équation linéaire entre le WMC et soit : NCLOC, DCP ou NOCom

Relations	Équations
(WMC, NCLOC)	$WMC = 0.25NCLOC + 2.49$
(WMC, DCP)	$WMC = -1.38DCP + 126.48$
(WMC, NOCom)	$WMC = 8.72NOCom - 20.01$

### Conclusion et menace à la validité

Nous avons testé s'il existait une fonction linéaire entre le WMC et le NCLOC, DCP ou NOCom. Suite à notre analyse, il y aurait une corrélation entre les trois métriques et le WMC. Nous avons pu déterminer à l'aide de la régression linéaire des équations pour chacune des relations. Ces résultats sont très préliminaire, malgré que nous avons un bon nombre de classes pour la puissance statistique, il ne faut pas oublier qu'on travaille avec un seul projet et qu'on ne peut pas vraiment généraliser ces formules. Des facteurs comme l'expérience de l'équipe ou les normes dans une entreprise pourraient peut-être influencer les relations. C'est une situation qui pour l'instant est spécifique à la situation de l'étude. Il serait intéressant de tester sur plusieurs projets différents.

## **T5 : Conclusion**

Nous avons effectué deux études sur le projet jfreechart afin de d'évaluer des relations entre des métriques de qualité du logiciel. Nous avons d'abord testé les classes avec plus 10 modifications était mieux documenté. Dans le cas spécifique de ce projet, nous avons déterminé que les classes avec plus de 10 modifications étaient mieux documentées. Dans une seconde étude, nous avons testé s'il y avait une relation linéaire entre le WMC et NCLOC, DCP ou le NOCom. Il semble, dans ce projet, qu'il existe une relation linéaire entre le WMC et tous les autres métriques. Finalement, ces études nous ont permis de réaliser qu'il existe potentiellement des relations entre

les différentes métriques liées à la qualité du logiciel. Il faut cependant, émettre un bémol, les deux études ont été effectuées sur le même projet. Ce qui implique peut-être les mêmes développeurs, la présence ou l'absence de modèle travail ce qui peut menacer la validité externe. Il serait intéressant de répéter les mêmes études sur différents projets.