

## CHAPITRE 5

### PRÉSENTATION ET INTERPRÉTATION DES RÉSULTATS

Avant le début de ces travaux, avec l'utilisation de quatre photos par échantillon, l'erreur moyenne mesurée entre un compteur et la moyenne de cinq compteurs était de 25%. De plus, chaque échantillon était compté par un chercheur différent, ce qui engendre des erreurs supplémentaires.

La démonstration fut faite à l'interne, qu'à ce niveau d'incertitude, l'utilisation du taux de croissance brut est impossible. En réponse à cette observation, une deuxième recherche à l'interne fut réalisée pour minimiser l'erreur engendrée par tout ce qui est externe au recensement visuel, c'est-à-dire, le traitement enzymatique, la dilution avant le dépôt sur la plaque ainsi que le nombre de photos prises. Ce nombre est passé de quatre à dix. Ces recherches ont diminué l'incertitude moyenne des chercheurs à 10%, dans des conditions optimales. Ces conditions se détériorent rapidement avec le temps. Durant le compte, un humain ne peut conserver ses critères de sélection et ses estimations constantes peu importe sa condition. De plus, les chercheurs doivent se soumettre à une vérification régulière pour réajuster leur « facteur de correction », soit un décalage linéaire, pour se rapprocher de la moyenne des cinq comptes qui sert d'objectif. Ce facteur diminue l'écart moyen.

Rappelons que l'objectif de cette recherche n'est pas de diminuer cette incertitude mais bien de s'en approcher. Cette recherche vise surtout à uniformiser la technique de compte, de libérer du temps pour les chercheurs et d'offrir une occasion de rendre l'utilisation du taux de croissance plus fiable. Même si la précision est différente, il n'y a aucun problème tant qu'elle est constante. Puisque le but est d'utiliser la dérivée de la population, soit le taux de croissance, un décalage linéaire est transparent par rapport au résultat nécessaire.

### **5.1 Critères de sélection**

L'utilisation de critères de sélection fixes pour tous les comptes permet de minimiser les variations des comptes causées par le changement de compteur et permet d'harmoniser les critères de sélection. L'utilisation du logiciel permet de varier, de manière déterminée, chaque seuil de chaque critère en plus de donner accès à de l'information incalculable à la main, tels la taille moyenne des cellules libres, du taux de regroupement, de la forme, etc.

Chacun des objets conservés par la segmentation sera sélectionné pour le traitement si

- La surface est supérieure à 100 pixels
- La circularité est supérieure à 0.5 et la surface est inférieure à 200 pixels
- L'elongation est inférieure à 0.85
- L'écart type d'intensité est supérieur à 10 sur 255

Un exemple en couleur est présenté à la Figure 17.

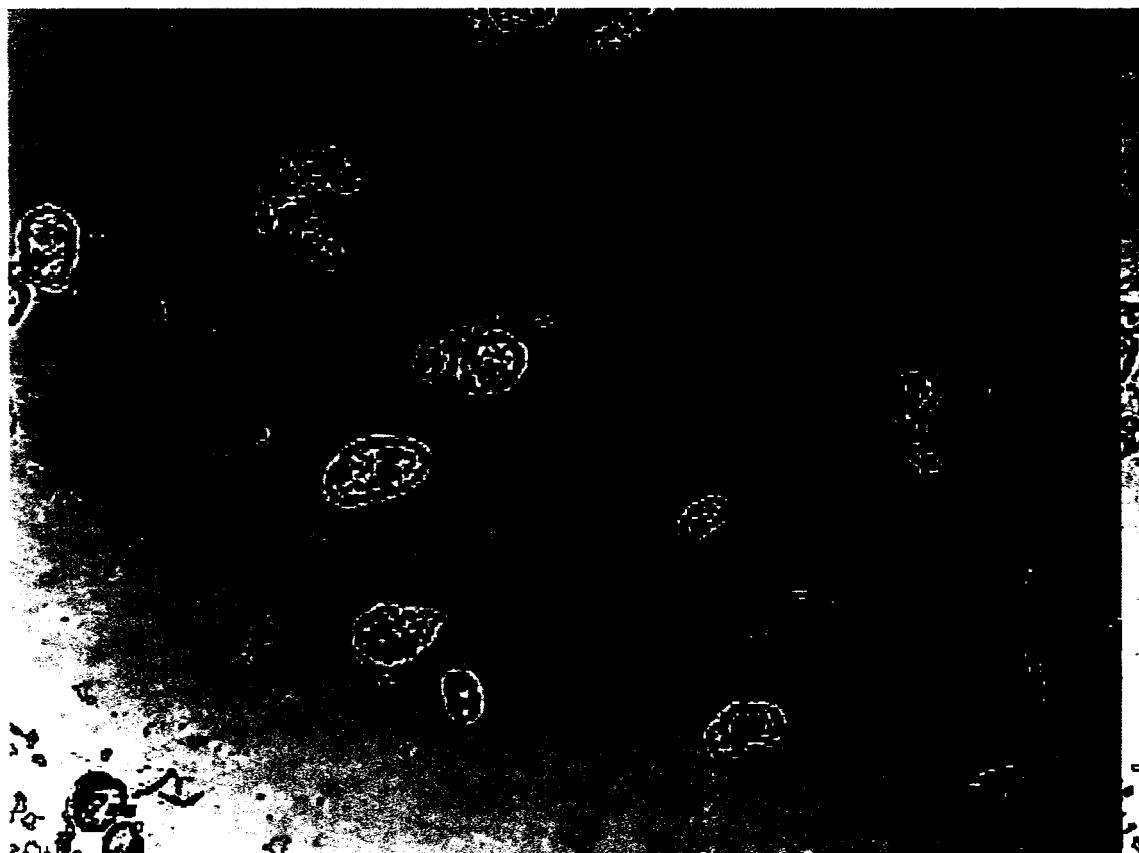


Figure 17 Résultat de la sélection

### 5.1.1 Choix du seuil de la surface minimale

Pour choisir un seuil minimal adéquat, la référence à la littérature est nécessaire. Une cellule végétale type a de 10 à 100  $\mu\text{m}$  de diamètre, soit une surface de 79 à 7854  $\mu\text{m}^2$ . Si un pixel couvre 4.39873  $\mu\text{m}^2$  avec une résolution de 100X, alors une cellule végétale type couvrirait de 18 à 1785 pixels.

Plusieurs débris se retrouvent dans la photo. À la résolution utilisée, ils ressemblent à un volume fermé telle une très petite cellule. Malheureusement, à cette taille, il peut s'agir

soit d'une très petite cellule, soit d'une organelle libérée d'une cellule éclatée ou soit d'un débris quelconque.

Donc, une organelle d'une grosse cellule peut être plus grosse qu'une petite cellule. Il est très difficile de trouver une limite théorique à la taille minimale d'une cellule. La technique la plus sûre est l'observation.

Après classement manuel, les cellules de l'espèce observée ont un diamètre moyen de 55  $\mu\text{m}$ , donc une surface de 540 pixels. Le seuil minimum décidé expérimentalement est de 5.65  $\mu\text{m}$  soit 100 pixels. Ce choix empirique fut basé sur une vingtaine d'images de deux bioréacteurs différents à des temps différents.

Ce seuil influence la taille des cellules libres retenues et mesurées qui servent de références pour le compte.

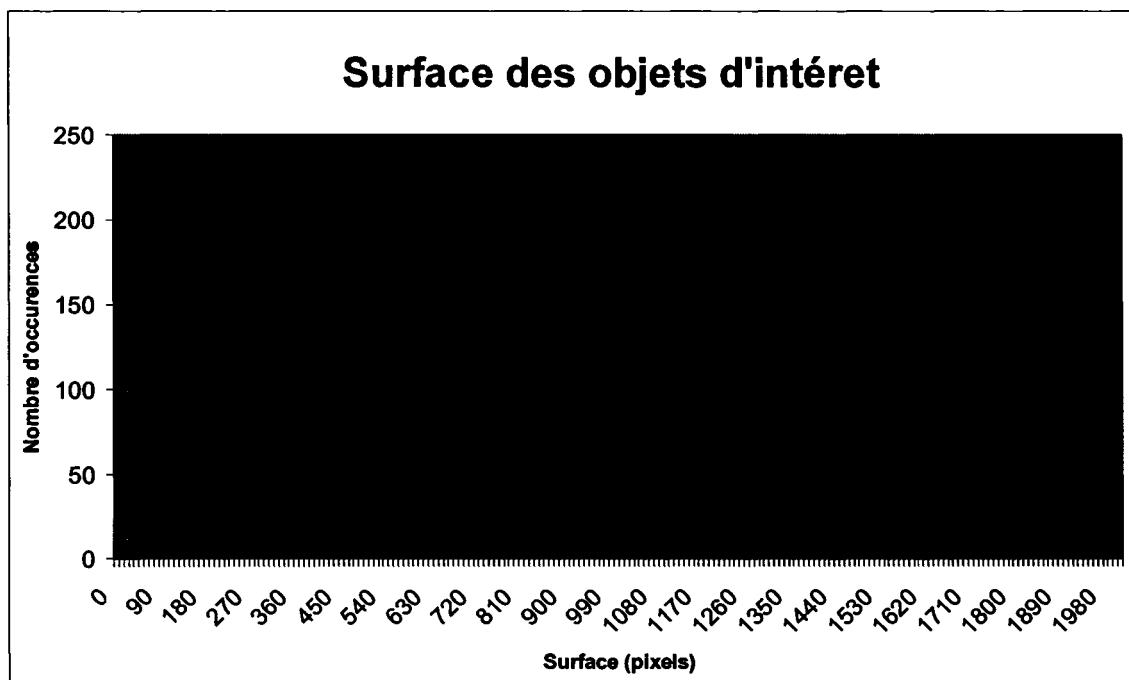


Figure 18 Graphique de la répartition des surfaces (Pixels)

Tel que démontré sur le graphique de la Figure 18, plus les cellules sont petites, plus elles sont nombreuses et se confondent en taille aux débris.

### 5.1.2 Choix du seuil de circularité

Théoriquement, une cellule devrait être relativement sphérique en suspension. Question d'économie naturelle, la cellule possède un maximum de volume pour un minimum de surface. En analysant tous les objets recensés, on trouve une répartition de type moyenne (Figure 19).

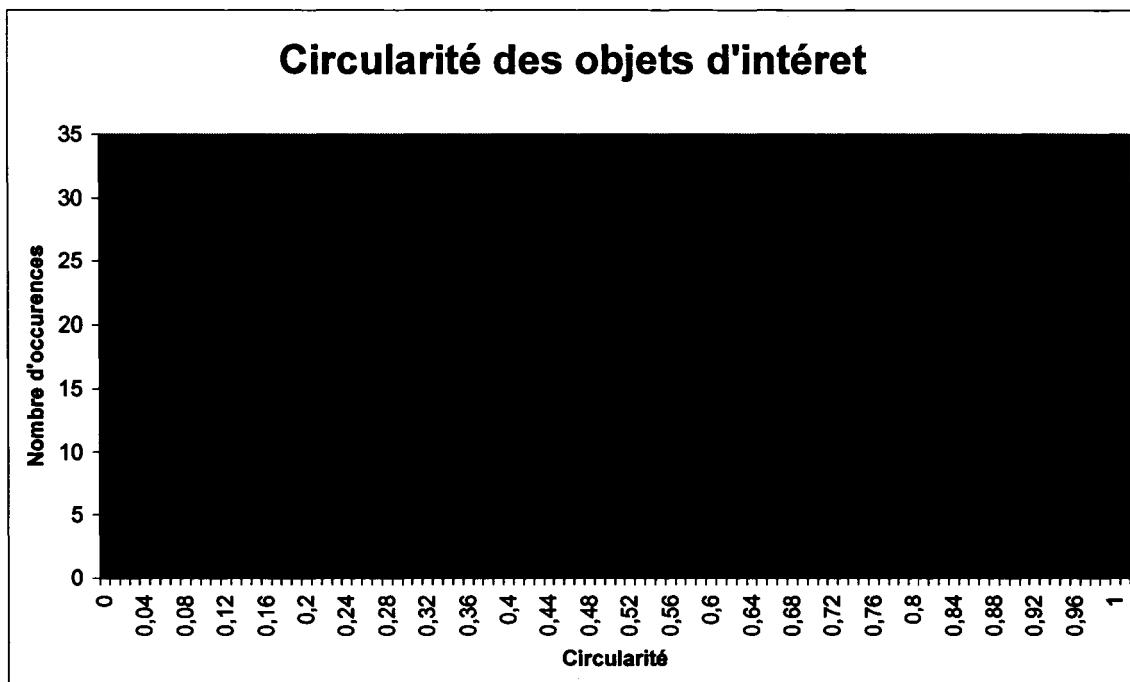


Figure 19 Graphique de la répartition de la circularité

Il est donc difficile d'extraire à nouveau un seuil entre les classes. Suite à l'observation d'une vingtaine d'images, un seuil de 0,5 fut choisi si les cellules ont moins de 200 pixels de surface. La circularité correspond à la formule suivante :

$$\text{Circularité} = 4 \pi * (\text{aire}/\text{périmètre}^2)$$

Un cercle parfait égale 1.0 et tend vers zéro pour une forme plus complexe. Mais au-delà de 200 pixels de surface, la limite absolue de circularité dans les images observées est de 0.15. Cette valeur correspond généralement aux débris possédant des contours flous.

### 5.1.3 Choix du seuil d'elongation

Encore une fois, il n'est pas possible de distinguer une information claire de la répartition des objets en fonction de leur élongation. Après l'observation d'une vingtaine d'images, le seuil fut fixé à 0.85. Cette valeur correspond principalement à l'élongation des lignes du fond de la chambre de compte qui sont sélectionnées. Il est possible de voir une grande quantité de ces objets (près de 120) possédant une valeur de 1 (Figure 20). Un étant une ligne parfaite et zéro étant un cercle, la formule est la suivante:

$$\text{Élongation} = (\text{Axe majeur} - \text{axe mineur}) / (\text{axe majeur} + \text{axe mineur})$$

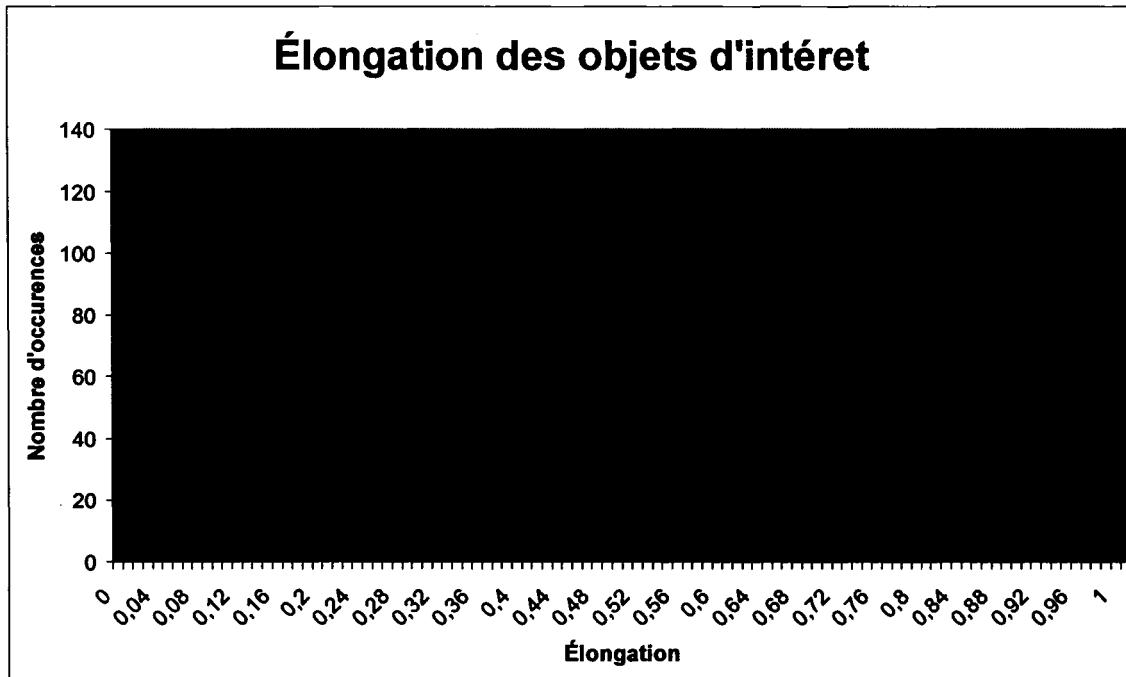


Figure 20 Graphique de la répartition de l'élongation

#### 5.1.4 Choix du seuil de l'écart type d'intensité

Le graphique de la Figure 21 nous permet de constater que plus l'écart type augmente, plus le nombre d'objets augmente. De nouveau, après l'observation d'une vingtaine d'images, sur une échelle de 0 à 255, le seuil fut fixé à 10. En dessous de cette limite, les objets sélectionnés ne sont que des variations d'éclairages ponctuels et autres variations non désirées.

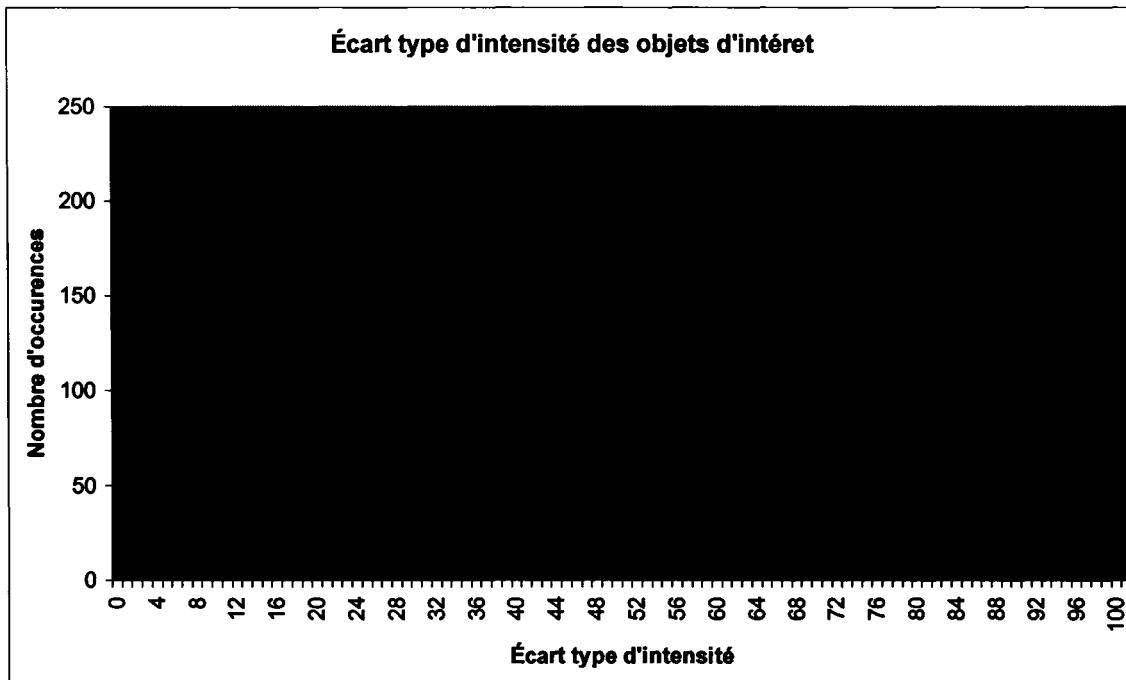


Figure 21 Graphique de la répartition de l'écart type de l'intensité

### 5.1.5 Difficultés reliées aux critères de sélection

Comme expliqué précédemment, le fait de ne pas pouvoir utiliser des éléments caractéristiques fiables empêche d'avoir des taux de reconnaissance supérieurs à l'être humain. L'utilisation de plusieurs critères simples à seuil large permet de s'en rapprocher.

Il demeure qu'il est difficile de reconnaître la taille des cellules au sein d'un amas car souvent, lorsque les cellules sont regroupées en amas, leur taille est plus petite. Il est impossible, avec la méthode actuelle, de compter adéquatement les cellules. Elles se superposent, elles sont floues parce qu'elles sont hors champ et la cellulose reste opaque.

Ceci est le résultat des précédentes discussions; le manque de caractéristiques distinctives rend la reconnaissance d'une cellule seule ou en amas très complexe.

### **5.1.6 Assistance aux comptes par vision artificielle**

Tel que discuté dans la présentation de l'interface, un opérateur peut modifier, valider ou recompter une image. Une étude pourrait être conduite pour savoir l'impact des comptes sur les comptes. Mais le résultat est prévisible et inutile à l'heure actuelle puisque ultimement les résultats des comptes par vision artificielle assisté en totalité par un être humain seraient identiques aux comptes humains.

## **5.2 Projection de croissance, Bioréacteur 1**

Les comptes affichés sont le résultat des comptes des échantillons. Les comptes manuels en production représentent la moyenne de cinq compteurs, ce qui nous permet d'effectuer la comparaison avec les comptes automatiques par vision artificielle sans assistance.

Les résultats numériques sont présentés au Tableau II. La colonne « Date » représente les dates de prise d'échantillons, la colonne « Manuel » représente la moyenne de cinq compteurs expérimentés, la colonne « Automatique » représente le compte généré par vision artificielle sans assistance et la colonne « Écart » représente l'écart en pourcentage entre les comptes automatiques et manuels. La colonne « Facteur de dilution » représente le facteur de dilution puisque tous les ajouts de nutriment, glucose, etc. sont ajoutés en solution dans le bioréacteur. Il faut donc tenir compte du facteur de dilution pour bien estimer la croissance cellulaire.

Tableau II

## Différence en comptes automatiques et comptes manuels

2004-08-31	<b>687 335</b>	<b>684 173</b>	0,4%	1,00
2004-09-01	<b>633 231</b>	<b>673 716</b>	6,3%	1,00
2004-09-02	<b>797 767</b>	<b>649 815</b>	18,5%	1,03
2004-09-03	<b>1 003 315</b>	<b>1 068 385</b>	6,4%	1,07
2004-09-04	<b>1 238 372</b>	<b>971 585</b>	21,5%	1,11
2004-09-05	<b>1 598 975</b>	<b>1 109 615</b>	30,6%	1,14
2004-09-06	<b>1 385 266</b>	<b>1 106 926</b>	20,0%	1,17
2004-09-07	<b>2 184 836</b>	<b>2 083 889</b>	4,6%	1,27
2004-09-08	<b>2 043 079</b>	<b>2 039 074</b>	0,2%	1,34
2004-09-09	<b>1 800 982</b>	<b>1 420 630</b>	21,1%	1,42
2004-09-10	<b>2 783 485</b>	<b>1 940 481</b>	30,2%	1,48
2004-09-11	<b>2 233 668</b>	<b>1 817 689</b>	18,6%	1,70
2004-09-12	<b>2 138 102</b>	<b>2 312 444</b>	8,1%	1,73
2004-09-13	<b>1 917 552</b>	<b>2 237 156</b>	16,6%	1,77
2004-09-14	<b>4 190 734</b>	<b>4 447 422</b>	6,1%	1,82
2004-09-15	<b>2 255 410</b>	<b>2 398 489</b>	6,3%	1,87

Dans le Tableau III, l'erreur moyenne représente la moyenne des erreurs entre tous les comptes par vision artificielle et humaine de chaque échantillon. L'écart type permet d'estimer la répartition de cette erreur. Un facteur est utilisé pour un décalage linéaire qui minimise l'erreur moyenne. Ce facteur, différent à chaque fois, est utilisé sur le résultat journalier de chaque bioréacteur. Ce facteur est calibré régulièrement. Le décalage linéaire est autorisé puisqu'il n'influence pas le taux de croissance. Voir le graphique de la Figure 22 pour un profil plus visuel de ces comptes.

Tableau III

Résultat moyen entre le compte automatique  
et le compte manuel du bioréacteur no1

13,52%
9,87%
1,21

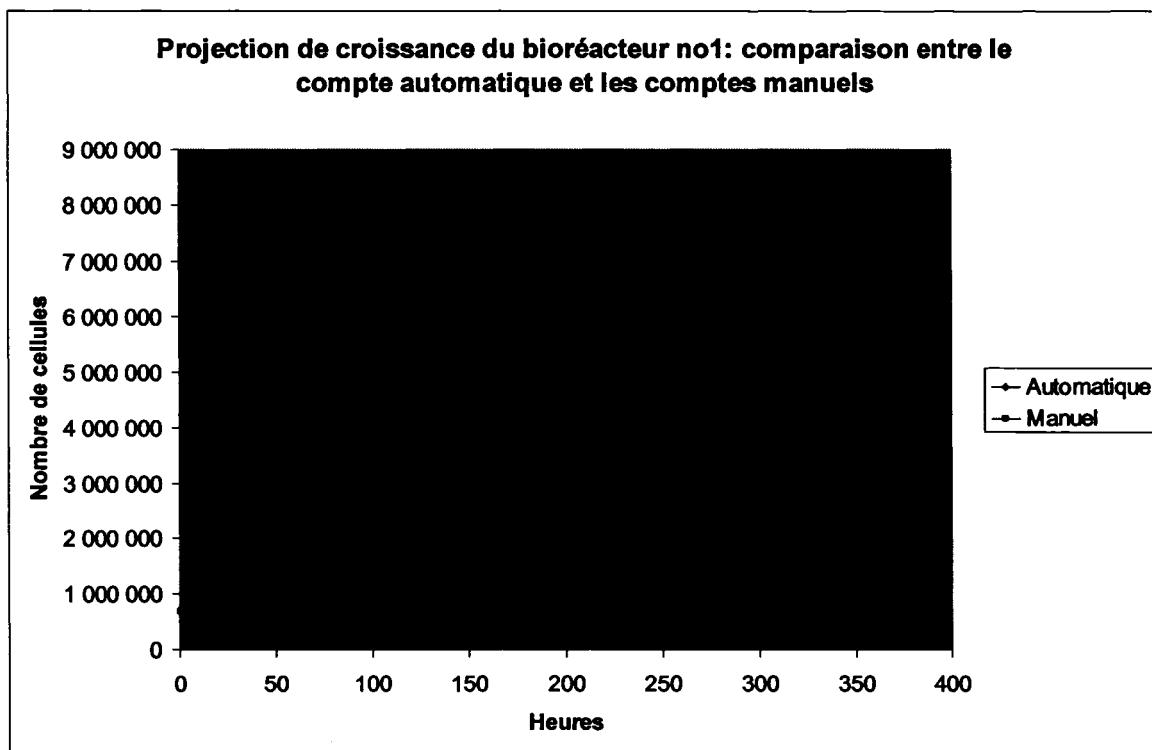


Figure 22 Différence entre le compte automatique et le compte en production

Premièrement, il est important de noter que le compte en manuel est une moyenne de cinq compteurs pour avoir un comparatif optimal, mais que le compte manuel en production est effectué quotidiennement par diverses personnes et sa constance diminue. Cependant, il est à noter que le but n'est pas de comparer les compteurs mais le compte automatique.

En se basant sur ce graphique, il est possible de constater que le système automatique sans assistance est utilisable avec une erreur moyenne comparable à celle entre humain. Seule ombre au tableau, le compte se rapprochant de deux cent quarante heures. Après observation des images individuelles, il est apparu que si l'image ne contient que des petits amas, le compte automatique va alors sous-évaluer. La cause est simple; la séparation entre la cellule simple et l'amas est issue d'une comparaison de la taille moyenne. Donc, si tous les objets sélectionnés sont multiples, les amas sont sous-évalués puisqu'ils sont considérés comme étant des cellules simples, distinction que fait aisément un compteur humain. De plus, la cellule moyenne seule sert de référence, alors si les amas sont composés de plus petites cellules, le nombre de cellules dans l'amas peut être sous-évalué. Mais, puisque dix images sont accessibles une règle pourrait être ajoutée pour extraire la moyenne sur toutes les images pour enrayer ce problème.

Il est possible de voir à la Figure 23 que le compte de gauche démontre une sous-évaluation ou l'erreur moyenne est de 30,29% et le compte de droite, un compte plus exact. La nuance est importante car au premier coup d'œil, les sélections semblent bonnes, Du à la sélection fixe de 1,5 fois la moyenne, de nombreux petits amas sont classés comme cellule libre, ce qui sous-évalue les images.

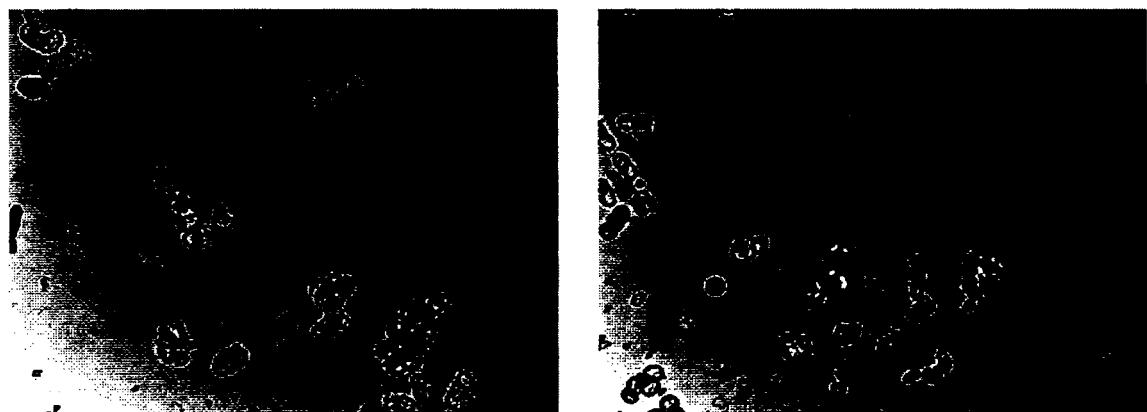


Figure 23 Exemple de sous-évaluation (gauche) et de compte plus exact (droite)

Les écarts entre les comptes sont tous liés directement à la précision de chaque critère de sélection. Pour y parvenir, il faut reconnaître les cellules. Des tests ont été faits, telle la transformé circulaire de Hough. L'approche est bonne puisque le but était de retrouver des cellules dans les amas pour augmenter le nombre de références. Mais les tentatives furent vaines parce que, même en utilisant des cellules rondes et même en ne cherchant que des arcs de cercles, la complexité d'un amas rend l'opération impossible, d'autant plus que telles que présentées précédemment, les cellules n'ont pas de caractéristiques distinctives.

En finalité, la précision est intéressante mais est loin d'être parfaite. Les résultats sont acceptables selon l'hypothèse de travail.

### 5.2.1 Taux de croissance, bioréacteur 1

Le résultat le plus important est la courbe du taux de croissance. Cette courbe est la dérivée de la population totale, les unités résultantes sont en heures<sup>-1</sup>. Le graphique de la Figure 24 montre la différence entre les taux de croissance du compte par vision artificielle et la moyenne des celui des humains.

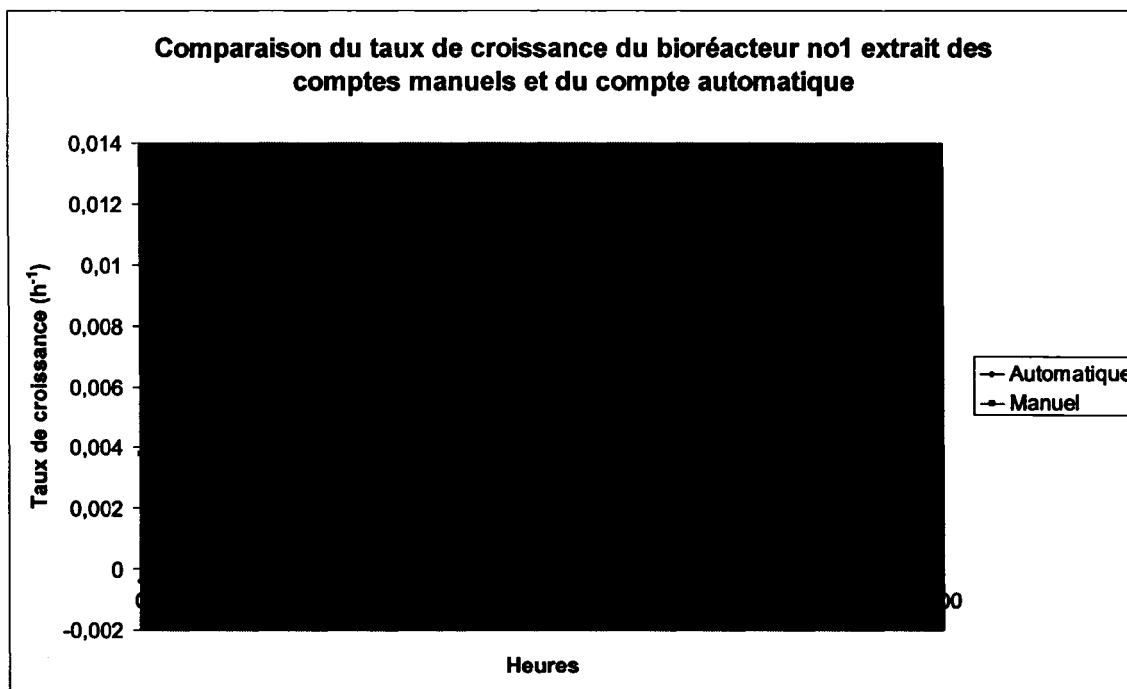


Figure 24 Taux de croissance du bioréacteur 1

Cette courbe est déterminante pour le projet. Les variations de taux de croissance sont primordiales pour les prises de décisions. Si l'on observe les tendances, il est possible de voir deux augmentations et réductions du taux de croissance. Bien que la population décroît dans la courbe de projection de croissance, cette croissance négative n'est pas reflétée dans la Figure 24, puisque le taux de croissance utilisé est une dérivé sur cinq points. Un autre constat peut être fait, les maximum et minimum de croissance sont à peu près identiques, malgré le fait que ce soit une dérivée. De plus, il est normal d'observer une légère décroissance au début de la culture, décroissance montrée seulement par le compte automatique. Mais un décalage dans le temps est observable. Après manipulation des données, il apparaît simplement que c'est l'imprécision du compte automatique qui crée ce déphasage, spécifiquement pour cette courbe.

Cette courbe ne nous permet pas, dans ces conditions, d'obtenir le temps idéal pour prendre une décision telle la récole ou l'inoculation à cause du déphasage. Par contre, un bioréacteur ne devrait pas avoir plusieurs cycles courts comme observés ici.

### 5.2.2 Projection de croissance, bioréacteur 2

Un deuxième bioréacteur fut suivi un peu moins longtemps puisqu'il fut utilisé pour remplacer d'autres réacteurs contaminés en production. Il est présenté ici puisque les comptes manuels nous démontrent une courbe de croissance idéale. Ceci prouve que l'amélioration de la technique de comptes manuels a porté fruit et prouve aussi que la moyenne des comptes peut être fiable. Les comptes quotidiens sont présentés numériquement au Tableau IV, graphiquement à la

Figure 25 et les résultats synthétisés au Tableau V.

Tableau IV

Différence entre les projections de croissance automatique et celles en production du deuxième bioréacteur

2004-08-31	<b>799 878</b>	<b>805 852</b>	0,7%	1,00
2004-09-01	<b>723 707</b>	<b>821 185</b>	13,5%	1,00
2004-09-02	<b>698 483</b>	<b>698 519</b>	0,0%	1,03
2004-09-03	<b>1 097 043</b>	<b>1 324 800</b>	20,8%	1,07
2004-09-04	<b>1 527 128</b>	<b>1 410 667</b>	7,6%	1,12
2004-09-05	<b>1 250 590</b>	<b>899 556</b>	28,1%	1,15
2004-09-06	<b>2 383 076</b>	<b>1 625 333</b>	31,8%	1,18
2004-09-07	<b>2 287 168</b>	<b>2 179 889</b>	4,7%	1,30
2004-09-08	<b>1 719 345</b>	<b>2 208 000</b>	28,4%	1,37
2004-09-09	<b>2 794 285</b>	<b>2 340 889</b>	16,2%	1,44
2004-09-10	<b>2 306 536</b>	<b>2 100 667</b>	8,9%	1,50

Tableau V

Résultat moyen entre le compte automatique et le compte en production du bioréacteur no2

	<b>15,02%</b>
	<b>12,47%</b>
	<b>0,93</b>

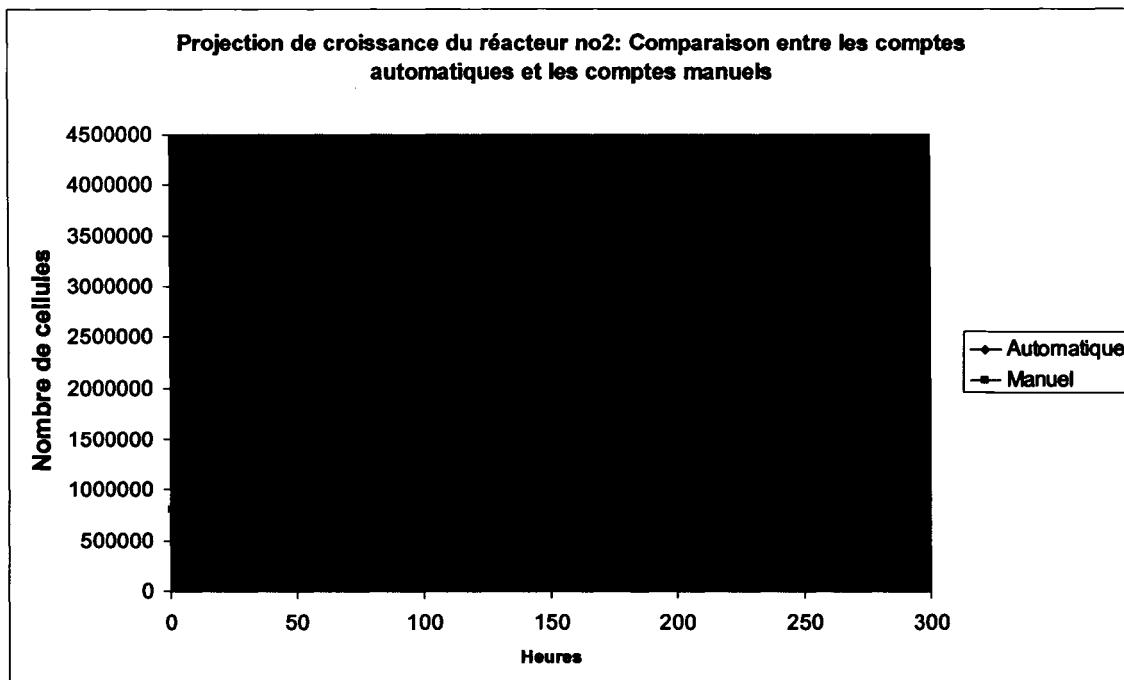


Figure 25 Projection de croissance du deuxième bioréacteur

### 5.2.3 Taux de croissance, bioréacteur 2

La Figure 26 montre que la courbe du taux de croissance extraite des comptes humains est similaire à une courbe théorique. Cette forme, montée rapide et lente décroissance, est la forme attendue d'une culture normale. Cette courbe fut la première courbe de cette qualité au sein de l'entreprise dans la cadre d'un bioréacteur de production. La différence : le compte journalier. Dans le but de sauver du temps et sans preuve du besoin, les comptes s'effectuaient aux deux à trois jours. Le compte journalier fut appliqué à tous les bioréacteurs à partir de ce moment, augmentant ainsi le besoin d'automatiser les comptes. Ces résultats ont permis de confirmer que la moyenne des comptes des chercheurs était une référence fiable.

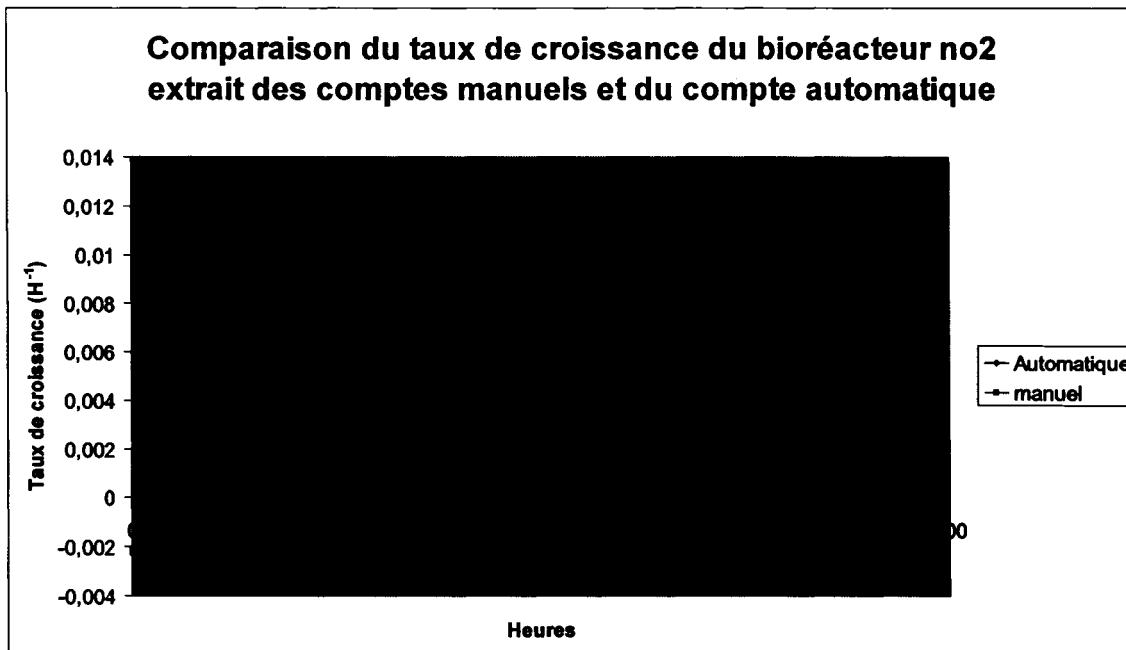


Figure 26 Taux de croissance du deuxième bioréacteur

Aussi, il est possible de constater que la courbe extraite du compte par vision artificielle est très près de celle extraite des comptes manuels. Les deux courbes débutent au même moment, ont la même pente initiale, ont le presque même taux maximum et terminent au même taux.

Une fois de plus, le même décalage est observé. Les erreurs en cause sont exactement les mêmes que celles observées lors de la comparaison des courbes de croissance du bioréacteur no1.

Cette erreur est due principalement au phénomène de sous-évaluation tel que décris précédemment. Ce problème peut être éliminé en améliorant le classificateur qui sépare les amas des cellules libres.

### **5.3 Compte par vision artificielle comparé à un compteur unique**

Bien que la moyenne de cinq compteurs puisse refléter la réalité, chaque compteur contribue à une erreur relative et l'erreur totale est réduite par la moyenne des comptes. Il est intéressant de pourvoir comparer le compte automatique et le compte d'une seule personne pour déterminer si le compte est comparable.

Un seul compteur sera comparé pour l'exemple suivant.

#### **5.3.1 Différence entre le compte par vision artificielle et le compteur unique**

Le Tableau VI démontre les différences entre les comptes du compteur unique et le compte par vision artificielle sans assistance. Un décalage linéaire est effectué pour minimiser l'erreur moyenne. Le compteur 1 fut choisi pour démontrer à quel point le compte par vision artificielle peut être similaire à un être humain. Les résultats sont synthétisés dans le Tableau VII.

### 5.3.1.1 Résultats :

Tableau VI

Différence entre un compte par vision artificielle et un  
compteur no 1

2004-08-31	<b>714 815</b>	<b>718099</b>	0,46%
2004-09-01	<b>632 099</b>	<b>707123</b>	11,87%
2004-09-02	<b>865 432</b>	<b>682037</b>	21,19%
2004-09-03	<b>1 096 296</b>	<b>1121363</b>	2,29%
2004-09-04	<b>1 309 630</b>	<b>1019763</b>	22,13%
2004-09-05	<b>1 428 148</b>	<b>1164637</b>	18,45%
2004-09-06	<b>1 164 815</b>	<b>1161815</b>	0,26%
2004-09-07	<b>2 146 091</b>	<b>2187222</b>	1,92%
2004-09-08	<b>2 074 074</b>	<b>2140185</b>	3,19%
2004-09-09	<b>2 003 704</b>	<b>1491074</b>	25,58%
2004-09-10	<b>2 666 667</b>	<b>2036704</b>	23,62%
2004-09-11	<b>2 586 667</b>	<b>1907822</b>	26,24%
2004-09-12	<b>2 862 222</b>	<b>2427111</b>	15,20%
2004-09-13	<b>2 093 333</b>	<b>2348089</b>	12,17%
2004-09-14	<b>4 471 111</b>	<b>4667956</b>	4,40%
2004-09-15	<b>2 586 667</b>	<b>2517422</b>	2,68%

Tableau VII

Résultat moyen entre le compte automatique et le compteur 1

11,98%
9,84%
1,27

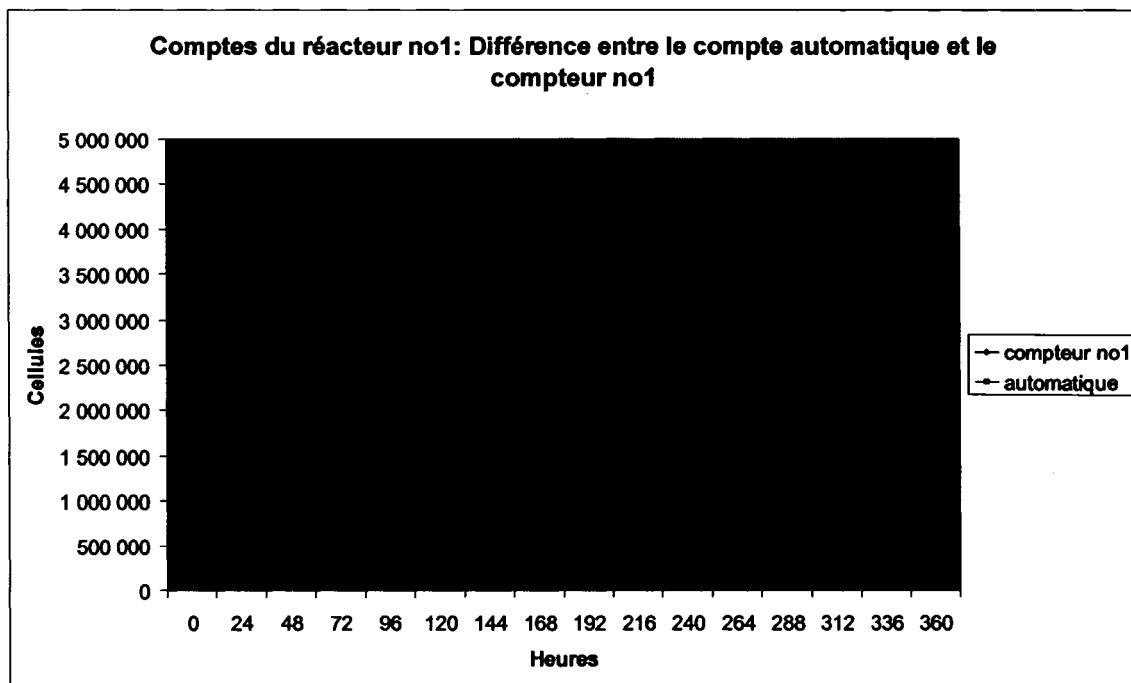


Figure 27 Différence entre le compte automatique et le compteur no1

Le compteur 1 donne un compte très comparable au compte automatique (Figure 27). Il y a une légère surévaluation aux environs de 250 heures qui est causé par la présence de beaucoup de débris. En fait, ce n'est pas une surévaluation locale mais bien une sous-évaluation globale transformée en surévaluation par la multiplication d'un facteur linéaire supérieur à 1.

L'écart moyen de ce compteur par rapport au compte automatique est faible. En comparaison, la différence entre ce compte et d'autres compteurs peut être aussi élevée que 20.4% avec un écart type de 11.0%.

### **5.3.2 Temps utilisé**

Un compteur, au début d'une journée où ses critères de sélection sont les plus sévères, peut prendre en moyenne 60 secondes par image pour compter. À dix images par seconde, un compteur prend 10 minutes pour calculer la concentration cellulaire et ce sans compter la manipulation de données.

Sur un PIV 3.06GHz HT, l'interface créée dans le cadre de ce projet en Visual Basic prend en moyenne 2 minutes pour compter les 10 images calculées.

Mais le temps total n'est pas un facteur important. Une partie de l'échantillon est envoyée à un HPLC (High Performance Liquid Chromatography), un chromatographe liquide de haute performance pour mesurer les concentrations de multiples autres composés chimiques présents dans le milieu du bioréacteur. Avec le traitement de données, ce processus peut facilement prendre près de 30 minutes par bioréacteur. En tenant compte du fait que ce traitement étant plus long et que les comptes doivent se faire en dedans de 24 H, le temps ne fut pas un facteur prédominant dans la recherche de solution. Par contre, plus il est possible de traiter des photos sans assistance, plus l'augmentation du nombre de photos pourrait augmenter la précision.

### **5.4 Choix des compteurs**

Pendant la récolte des données, des écarts majeurs entre un compteur et la moyenne des autres futs observés (de 169% à 226%). Ce compteur a expliqué que puisqu'un facteur linéaire est appliqué aux comptes, il préférait tout compter dans l'image pour être plus

représentatif. Le décalage entraînerait une correction qui superposerait ses comptes aux autres. Ce qui est faux, car ce serait laisser le facteur linéaire devenir un critère de sélection pour les objets dans l'image.

Pour valider cette théorie, l'effet sur la moyenne fut vérifié. Après l'addition de ce seul compteur, la moyenne des compteurs fut changé de 2% à 29% et ce, malgré l'application du facteur linéaire de 0.64.

Après observations, ses comptes ne furent pas retenus pour le calcul de moyenne qui sert de référence tout au long de ce projet. Une discrimination optimale des cellules est nécessaire pour représenter le plus possible la réalité.