



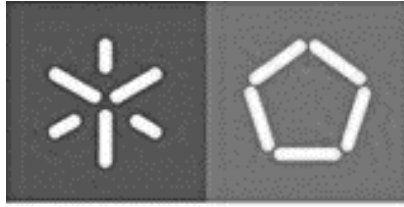
**Universidade do Minho**  
Escola de Engenharia

André Paço da Fonseca

A digital image based method to quantify intracellular polyphosphates in microbial aggregates

Novembro de 2018





**Universidade do Minho**  
Escola de Engenharia

André Paço da Fonseca

## A digital image based method to quantify intracellular polyphosphates in microbial aggregates

Dissertação de Mestrado

Mestrado em Bioinformática

Trabalho realizado sob a orientação de

Investigadora Auxiliar Daniela Mesquita

Professor Doutor Eugénio Ferreira

Novembro de 2018

## Declaração

Nome: André Paço da Fonseca

Endereço eletrónico: [andrepfons@hotmail.com](mailto:andrepfons@hotmail.com)

Telefone: 926579490

Cartão de Cidadão: 14920072

Título da dissertação: A digital image based method to quantify intracellular polyphosphates in microbial aggregates

Orientadores:

Investigadora Auxiliar Daniela Mesquita

Professor Doutor Eugénio Ferreira

Ano de conclusão. 2018

Mestrado em Bioinformática

É AUTORIZADA A REPRODUÇÃO INTEGRAL DESTA DISSERTAÇÃO APENAS PARA EFEITOS DE INVESTIGUAÇÃO, MEDIANTE DECLARAÇÃO ESCRITA DO INTERESSADO, QUE A TAL SE COMPROMETE.

Universidade do Minho, 30/11/2018

Assinatura: André Paço da Fonseca

## Acknowledgments/Agradecimentos

Em primeiro lugar, gostaria de agradecer à minha orientadora, Daniela Mesquita, por todo o tempo que dispensou, todos os conselhos dados e pela boa disposição com que sempre se apresentou ao longo deste trabalho, o qual sem ela não seria possível de realizar. Gostaria também de agradecer ao Professor Eugénio Ferreira pela oportunidade de realizar esta tese sob o seu teto.

Gostaria de agradecer também aos colegas do LBBS por toda a simpatia e boa disposição demonstrados no local de trabalho, principalmente ao Miguel Carvalho por me ter acompanhado durante todo o processo e me ter ajudado sempre que possível.

Aos meus amigos e colegas do mestrado e licenciatura, queria agradecer pela companhia durante todos estes anos, pelos risos e pelas bebedeiras, pelos jogos de futebol e kings no PDF, pelas noitadas de poker ou de conversa até de madrugada, por invadirem a minha casa mesmo sem eu saber, por me acordarem às 6 da manhã ao chegarem do enterro, por... bem, é melhor parar aqui antes que piore. Por isto tudo e muito mais, o meu mais sincero obrigado a todos vocês.

Aos colegas do 3B, que me acompanharam dia e noite durante mais de 5 anos, além de tudo o que já escrevi em cima apenas quero que saibam que vocês transformaram o 3B num verdadeiro lar e vos considero como família. Mas não, continuam a não poder fumar na sala!

À “pinguim dos montes”, obrigado por estes anos todos. Obrigado por estares comigo tanto nos bons como nos maus momentos. Evoluí imenso como pessoa, amigo, estudante e amante nestes últimos anos e muito deveu-se a ti. Obrigado!!!

Aos meus pais, agradeço por todo o esforço que fazem e sempre fizeram para me proporcionar tudo o que possa precisar, e por vezes o que não preciso também. Ao meu irmão, obrigado simplesmente por seres o meu mano mais velho, não poderia escolher melhor nem que quisesse. Ao resto da minha família, agradeço o apoio e amor incondicional que me dão.

A todos vocês, muito obrigado!



## Resumo

A eutrofização é um grave problema de ecossistemas aquáticos que pode ameaçar a biodiversidade e a saúde humana. É causada por uma alta concentração de nutrientes, tal como fósforo, levando a um crescimento excessivo de plantas aquáticas e, conseqüentemente, à estagnação e diminuição dos níveis de oxigênio nos sistemas de água.

Para evitar este problema, as águas residuais das zonas rurais, industriais e urbanas são tratadas com recurso a remoção biológica de fósforo melhorada. Esta técnica utiliza microrganismos já presentes em águas residuais que incorporam o fósforo dentro das suas células na forma de polifosfato, removendo assim o nutriente da água. Uma parte muito importante desta técnica consiste em avaliar os níveis de polifosfato presentes nos microrganismos, o que é geralmente efetuado por métodos analíticos caros e demorados.

Para melhorar a etapa de medição de polifosfatos, o objetivo desta tese foi criar um programa de análise quantitativa de imagem que pudesse correlacionar dados obtidos de imagens com dados analíticos, criando modelos de regressão linear úteis para a quantificação de polifosfatos. Este programa utilizou segmentação de cores entre regiões de polifosfato e biomassa, com base nos espaços de cores RGB, HSV e LAB, e foi aplicado em imagens de amostras tingidas com diferentes colorações (azul de metileno e azul de toluidino) e com diferentes fatores de diluição (sem diluição e diluição 10x). Estas imagens foram ajustadas nos seus valores de gamma e contraste, para avaliar diferentes configurações de imagem em segmentação de cor. Os modelos para amostras de azul de metileno e amostras diluídas apresentaram os maiores coeficientes de regressão entre os resultados dos dados de imagem e os dados analíticos (0,920 e 0,992, respetivamente), existindo também algumas diferenças entre os modelos de imagens com diferentes contrastes e valores de gamma.

O espaço de cores escolhido para a análise de imagens parece ser muito importante, pois o RGB e o LAB apresentam resultados mais satisfatórios do que o HSV. Mais trabalho é necessário antes que este método possa ser usado, no entanto os resultados vistos nesta tese são promissores.

Palavras-chave: fósforo; polifosfato; remoção biológica de fósforo melhorada; análise quantitativa de imagem; azul de metileno; azul de toluidino; RGB; HSV; LAB.





## Abstract

Eutrophication is a serious problem of aquatic ecosystems that can threaten biodiversity and human health. It is caused by a high concentration of nutrients, such as phosphorus, leading to excessive growth of aquatic plants and, consequently, to stagnation and decreased levels of oxygen in water systems.

To prevent this problem, wastewater from rural, industrial and urban backgrounds are treated using enhanced biological phosphorous removal. This technique uses microorganisms already present in wastewater that incorporate phosphorous into their cells in the form of polyphosphate, removing the nutrient from the water. A very important part of this technique consists of evaluating the polyphosphate levels present in the microorganisms, which is usually done through expensive or time consuming analytical methods.

To improve the polyphosphate measurement step, the objective of this thesis was to create a quantitative image analysis program that can correlate image obtained data with analytical data, creating linear regression models useful for polyphosphate quantification. This program used color segmentation between polyphosphate and biomass regions, based on RGB, HSV and LAB color spaces, and was applied on images of samples dyed with different stains (methylene blue and toluidine blue) and with different dilution factors (no dilution and 10x dilution). These images experienced gamma values and contrast adjustments, to evaluate different image settings on color segmentation. Models for methylene blue stained and diluted samples presented the highest regression coefficients between the image data results and analytical data (0.920 and 0.992, respectively), also existing some differences between models of images with different contrast and gamma values.

The color space chosen for image analysis seems to be very important with RGB and LAB presenting more satisfying results than HSV. More work is needed before this method can be used, however the results seen on this thesis are promising.

Key words: phosphorous; polyphosphate; enhanced biological phosphorous removal; quantitative image analysis; methylene blue; toluidine blue; RGB; HSV; LAB.



# Index

|  |      |
|--|------|
| Acknowledgments/Agradecimientos .....              | iii  |
| Resumo.....  | v    |
| Abstract.....                                      | vii  |
| List of figures .....                              | xi   |
| List of Tables.....                                | xv   |
| List of Abbreviations and Acronyms .....           | xvii |
| 1. Introduction.....                               | 1    |
| 1.1 Motivation .....                               | 1    |
| 1.2 Objectives .....                               | 2    |
| 2. State of the art.....                           | 3    |
| 2.1 History of wastewater treatment .....          | 3    |
| 2.2 Enhanced biological phosphorus removal.....    | 5    |
| 2.2.1 EBPR process – poly-P accumulation.....      | 5    |
| 2.2.2 Anaerobic phase.....                         | 5    |
| 2.2.3 Aerobic phase .....                          | 6    |
| 2.2.4 Phosphorus accumulating organisms (PAO)..... | 7    |
| 2.3 Poly-P .....                                   | 7    |
| 2.3.1 Characteristics .....                        | 7    |
| 2.3.2 Functions .....                              | 8    |
| 2.4 Poly-P analysis .....                          | 8    |
| 2.4.1 Chemical analysis.....                       | 8    |
| 2.4.2 Biological analysis.....                     | 9    |
| 2.4.3 Molecular analysis .....                     | 9    |
| 2.4.4 Microscopy and staining analysis .....       | 10   |
| 2.5 Quantitative Image Analysis (QIA) .....        | 11   |
| 2.5.1 Image .....                                  | 11   |
| 2.5.2 Color spaces.....                            | 12   |
| 2.5.3 Gamma.....                                   | 14   |
| 2.5.4 Image analysis .....                         | 15   |
| 3. Methods .....                                   | 17   |
| 3.1 Experimental set-up .....                      | 17   |
| 3.2 Analytical procedures.....                     | 19   |
| 3.3 Staining and image acquisition .....           | 20   |

|                                     |    |
|-------------------------------------|----|
| 3.4 Image analysis.....             | 20 |
| 3.4.1 Early script development..... | 21 |
| 3.4.2 Late script development.....  | 21 |
| 3.5 Linear regression analysis..... | 23 |
| 4. Results and Discussion .....     | 25 |
| 4.1 Samples and images .....        | 25 |
| 4.2 Analytical results .....        | 25 |
| 4.3 QIA results .....               | 27 |
| 4.3.1 Early script development..... | 27 |
| 4.3.2 Late script development.....  | 30 |
| 4.3.3 Obtained parameters.....      | 35 |
| 4.4 Linear regression models.....   | 42 |
| 4.4.1 Gamma and contrast.....       | 42 |
| 4.4.2 Dye used.....                 | 46 |
| 4.4.3 Dilution factor.....          | 48 |
| 5. Conclusion and future work ..... | 53 |
| 6. Literature .....                 | 55 |

## List of figures

|  |    |
|--|----|
| <b>Figure 1</b> Representation of the anaerobic phase biochemical reactions in a PAO. Image adapted from Tarayre [9].<br>.....   | 6  |
| <b>Figure 2</b> Representation of the aerobic phase biochemical reactions in a PAO. Image adapted from Tarayre [9].<br>.....   | 6  |
| <b>Figure 3</b> Example of poly-P linear structure. ....   | 7  |
| <b>Figure 4</b> Representation of the RGB color space in a three-dimensional shape – RGB cube. The x, y and z are represented with the colors red, green and blue respectively, with each coordinate representing a different color. Image adapted from Kuehni [62]. ....  | 12 |
| <b>Figure 5</b> Three-dimensional representation of HSV color space. HSV model is similar to a cylinder with hue (H) being arranged as a radial slice around a central axis of neutral color, saturation (S) as the horizontal depth of the structure, counting from the central axis, and value (V) as the height of the structure. This image was obtained from <a href="http://lib.povray.org">http://lib.povray.org</a> and is a creative property of Michael Horvath. ....  | 13 |
| <b>Figure 6</b> Representation of CIELAB color space as a sphere. This image was obtained from <a href="http://sheriffblathur.blogspot.com/2013/07/cie-lab-color-space.html">http://sheriffblathur.blogspot.com/2013/07/cie-lab-color-space.html</a> .....   | 14 |
| <b>Figure 7</b> Example of data obtained on an image with QIA. Image adapted from Meijering and Cappellen [60].<br>.....   | 15 |
| <b>Figure 8</b> Representation of the lab sized bioreactor used. There were 3 tubes permanently connected to the bioreactor, namely tubes 1, 2 and 3. Tube 1 had holes at the end of it and was connected to a source of compressed air, being responsible for aeration. Tube 2 was responsible for feed, being connected to a jerrycan filled with nutrients, along with the carbon and phosphate source. Tube 3 was responsible for the discharge of the effluent. Opening 4 was usually closed, only used when a sample was taken. ....   | 17 |
| <b>Figure 9</b> Filtration system used during VSS and TSS analysis. ....   | 19 |
| <b>Figure 10</b> Representation of the RGB_mask_segmentation.m script's workflow. ....   | 21 |
| <b>Figure 11</b> Representation of the Poly_P_analysis_all.m script's workflow. The step number 1 happens before step number 2 in order to select the sample regions which will be compared to the remaining images...   | 23 |
| <b>Figure 12</b> TSS and VSS content over the experimental period. ....  | 26 |
| <b>Figure 13</b> Intracellular poly-P values obtained with Hach LCK 350 Phosphate Kit. ....  | 27 |
| <b>Figure 14</b> Intracellular poly-P values per TSS content. ....   | 27 |
| <b>Figure 15</b> Images obtained through microscopy (a, c and e) and their respective segmented poly-P regions (b, d and f) according to the <b>RGB_mask_segmentation.m</b> script using 50-180, 10-90 and 110-190 threshold values on red, green and blue bands respectively. Image a was taken on day 1, c on day 13, and e on day 29. All images were obtained from samples with no dilution, dyed with MB. Gamma and contrast settings were not altered for these images.....  | 29 |
| <b>Figure 16</b> Images obtained through microscopy (a and c) and their respective segmented poly-P regions (b and d) according to the <b>RGB_mask_segmentation.m</b> script using 50-180, 10-90 and 110-190 threshold values on red, green and blue bands respectively. Image a was taken on day 1 and b on day 13. both images were obtained from samples with no dilution, dyed with TB. Gamma and contrast settings were not altered for these images.<br>.....  | 30 |
| <b>Figure 17</b> Segmentation of image a into poly-P and biomass regions (b and c, respectively), when using areas 1, 2 and 3 of image a as an example of said regions plus background. The remaining pixels were assigned to each region by Euclidean distance. Images b and c were also segmented into poly-P and biomass regions (images e and f, and images h and i, respectively) with the areas assigned in image a. RGB color space was used for these images. These images were taken on day 8, MB_sd. Gamma and contrast settings were not altered for these images. .... | 32 |
| <b>Figure 18</b> Segmentation of image a into poly-P and biomass regions (b and c, respectively), when using areas 1, 2 and 3 of image a as an example of said regions plus background. The remaining pixels were assigned to each   |    |

region by Euclidean distance. Images b and c were also segmented into poly-P and biomass regions (images e and f, and images h and i, respectively) with the areas assigned in image a. HSV color space was used for these images. These images were taken on day 8, MB\_sd. Gamma and contrast settings were not altered for these images. .... 33

**Figure 19** Segmentation of image a into poly-P and biomass regions (b and c, respectively), when using areas 1, 2 and 3 of image a as an example of said regions plus background. The remaining pixels were assigned to each region by Euclidean distance. Images b and c were also segmented into poly-P and biomass regions (images e and f, and images h and i, respectively) with the areas assigned in image a. LAB color space was used for these images. These images were taken on day 8, MB\_sd. Gamma and contrast settings were not altered for these images. .... 34

**Figure 20** Grayscale intensity per poly-P area in RGB color space during the experimental period. Plots a-b and c-d show data from images with 0.8 and 1.0 gamma values, respectively. Plots a-c and b-d show data from images without and with contrast, respectively. .... 36

**Figure 21** Grayscale intensity per poly-P area in HSV color space during the duration of the experiment. Plots a-b and c-d show data from images with 0.8 and 1.0 gamma values, respectively. Plots a-c and b-d show data from images without and with contrast, respectively. .... 37

**Figure 22** Grayscale intensity per poly-P area in LAB color space during the duration of the experiment. Plots a-b and c-d show data from images with 0.8 and 1.0 gamma values, respectively. Plots a-c and b-d show data from images without and with contrast, respectively. .... 38

**Figure 23** Average of poly-P area per biomass area for each day in the different gamma and contrast settings for RGB segmented images. Plots a-b and c-d show data from images with 0.8 and 1.0 gamma values, respectively. Plots a-c and b-d show data from images without and with contrast, respectively. The analytically measured intracellular poly-P concentration and intracellular poly-P per TSS are present as a red and green lines for better comparison between parameters. .... 39

**Figure 24** Average of poly-P area per biomass area for each day in the different gamma and contrast settings for HSV segmented images. Plots a-b and c-d show data from images with 0.8 and 1.0 gamma values, respectively. Plots a-c and b-d show data from images without and with contrast, respectively. The analytically measured intracellular poly-P concentration and intracellular poly-P per TSS are present as a red and green lines for better comparison between parameters. .... 40

**Figure 25** Average of poly-P area per biomass area for each day in the different gamma and contrast settings for LAB segmented images. Plots a-b and c-d show data from images with 0.8 and 1.0 gamma values, respectively. Plots a-c and b-d show data from images without and with contrast, respectively. The analytically measured intracellular poly-P concentration and intracellular poly-P per TSS are present as a red and green lines for better comparison between parameters. .... 41

**Figure 26** Linear regression models of RGB treated images using poly-P area per biomass area as predictor variables and analytical intracellular poly-P concentration per day (left side) or per TSS (right side) as response variables. The models were created using different data based on the image settings: models **A** and **B** – G\_2\_C\_1; models **C** and **D** – G\_2\_C\_2; models **E** and **F** – G\_4\_C\_1; models **G** and **H** – G\_4\_C\_2. All image sets corresponding to their respective settings were used in the models (MB\_dil, MB\_sd, TB\_dil and TB\_sd). Data is represented as blue stars, the fitted regression line is represented in red and the model's 95% confidence bounds are represented as light red dotted lines. The models' regression coefficients are the following: **A** = 0.440, **B** = 0.519, **C** = 0.764, **D** = 0.655, **E** = 0.442, **F** = 0.519, **G** = 0.274 and **H** = 0.132. .... 43

**Figure 27** Linear regression models of HSV treated images using poly-P area per biomass area as predictor variables and analytical intracellular poly-P concentration per day (left side) or per TSS (right side) as response variables. The models were created using different data based on the image settings: models **A** and **B** – G\_2\_C\_1; models **C** and **D** – G\_2\_C\_2; models **E** and **F** – G\_4\_C\_1; models **G** and **H** – G\_4\_C\_2. All image sets corresponding to their respective settings were used in the models (MB\_dil, MB\_sd, TB\_dil and TB\_sd). Data is represented as blue stars, the fitted regression line is represented in red and the model's 95% confidence

bounds are represented as light red dotted lines. The models' regression coefficients are the following: **A** = 0.203, **B** = 0.202, **C** = 0.302, **D** = 0.291, **E** = 0.358, **F** = 0.450, **G** = 0.194 and **H** = 0.087. ....44

**Figure 28** Linear regression models of LAB treated images using poly-P area per biomass area as predictor variables and analytical intracellular poly-P concentration per day (left side) or per TSS (right side) as response variables. The models were created using different data based on the image settings: models **A** and **B** – G\_2\_C\_1; models **C** and **D** – G\_2\_C\_2; models **E** and **F** – G\_4\_C\_1; models **G** and **H** – G\_4\_C\_2. All image sets corresponding to their respective settings were used in the models (MB\_dil, MB\_sd, TB\_dil and TB\_sd). Data is represented as blue stars, the fitted regression line is represented in red and the model's 95% confidence bounds are represented as light red dotted lines. The models' regression coefficients are the following: **A** = 0.337, **B** = 0.250, **C** = 0.459, **D** = 0.283, **E** = 0.417, **F** = 0.344, **G** = 0.290 and **H** = 0.200. ....45

**Figure 29** Linear regression models of RGB treated images using poly-P area per biomass area as predictor variables and analytical intracellular poly-P concentration per day (left side) or per TSS (right side) as response variables. The models were created using different data based on the dye used: models **A** and **B** – MB; models **C** and **D** – TB. Data is represented as blue stars, the fitted regression line is represented in red and the model's 95% confidence bounds are represented as light red dotted lines. The models' regression coefficients are the following: **A** = 0.891, **B** = 0.824, **C** = 0.640 and **D** = 0.787. ....46

**Figure 30** Linear regression models of HSV treated images using poly-P area per biomass area as predictor variables and analytical intracellular poly-P concentration per day (left side) or per TSS (right side) as response variables. The models were created using different data based on the dye used: models **A** and **B** – MB; models **C** and **D** – TB. Data is represented as blue stars, the fitted regression line is represented in red and the model's 95% confidence bounds are represented as light red dotted lines. The models' regression coefficients are the following: **A** = 0.403, **B** = 0.431, **C** = 0.883 and **D** = 0.651. ....47

**Figure 31** Linear regression models of LAB treated images using poly-P area per biomass area as predictor variables and analytical intracellular poly-P concentration per day (left side) or per TSS (right side) as response variables. The models were created using different data based on the dye used: models **A** and **B** – MB; models **C** and **D** – TB. Data is represented as blue stars, the fitted regression line is represented in red and the model's 95% confidence bounds are represented as light red dotted lines. The models' regression coefficients are the following: **A** = 0.920, **B** = 0.919, **C** = 0.690 and **D** = 0.654. ....48

**Figure 32** Linear regression models of RGB treated images using poly-P area per biomass area as predictor variables and analytical intracellular poly-P concentration per day (left side) or per TSS (right side) as response variables. The models were created using different data based on the dilution factor: models **A** and **B** – diluted samples; models **C** and **D** – non-diluted samples. Data is represented as blue stars, the fitted regression line is represented in red and the model's 95% confidence bounds are represented as light red dotted lines. The models' regression coefficients are the following: **A** = 0.869, **B** = 0.978, **C** = 0.888 and **D** = 0.869. ....49

**Figure 33** Linear regression models of HSV treated images using poly-P area per biomass area as predictor variables and analytical intracellular poly-P concentration per day (left side) or per TSS (right side) as response variables. The models were created using different data based on the dilution factor: models **A** and **B** – diluted samples; models **C** and **D** – non-diluted samples. Data is represented as blue stars, the fitted regression line is represented in red and the model's 95% confidence bounds are represented as light red dotted lines. The models' regression coefficients are the following: **A** = 0.218, **B** = 0.429, **C** = 0.706 and **D** = 0.611. ....50

**Figure 34** Linear regression models of LAB treated images using poly-P area per biomass area as predictor variables and analytical intracellular poly-P concentration per day (left side) or per TSS (right side) as response variables. The models were created using different data based on the dilution factor: models **A** and **B** – diluted samples; models **C** and **D** – non-diluted samples. Data is represented as blue stars, the fitted regression line is represented in red and the model's 95% confidence bounds are represented as light red dotted lines. The models' regression coefficients are the following: **A** = 0.989, **B** = 0.992, **C** = 0.510 and **D** = 0.359. ....51





## List of Tables

|   |    |
|---|----|
| <b>Table 1</b> Composition and concentration of nutrients that compose the synthetic feed used in the batch reactor. Trace metals composition can be seen in Table 2. This feed was based on [8], [75]..... | 18 |
| <b>Table 2</b> Composition and concentration of metals that compose the trace metals solution used in the synthetic medium. This solution is based on [8], [75].....  | 18 |



## List of Abbreviations and Acronyms

- ADP – Adenosine diphosphate
- AS – Activated sludge
- ATP – Adenosine triphosphate
- C\_1 – No contrast
- C\_2 – Higher contrast
- $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$  – Calcium chloride dihydrate
- CIE – *Commission International d'Eclairage*
- $\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$  – Cobalt(II) chloride hexahydrate
- COD – Chemical oxygen demand
- $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$  – Cupric sulfate pentahydrate
- DAPI – 4',6-Diamidino-2-phenylindole dihydrochloride
- EBPR – Enhanced biological phosphorous removal
- EDTA – Ethylenediaminetetraacetic acid
- $\text{FeCl}_3 \cdot 6\text{H}_2\text{O}$  – Ferric chloride hexahydrate
- G\_2 – Gamma value of 1
- G\_4 – Gamma value of 0.8
- $\text{H}_3\text{BO}_3$  – Boric acid
- HSV – Hue, saturation and value
- Int. poly-P – Intracellular polyphosphate
- KI – Potassium iodide
- LAB – Lightness,  $a^*$  and  $b^*$  spaces
- MB – Methylene blue
- MB\_dil – Methylene blue diluted
- MB\_sd – Methylene blue non-diluted
- $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$  – Magnesium sulfate heptahydrate
- $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$  – Manganese(II) chloride tetrahydrate
- $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$  – Sodium Molybdate Dihydrate
- $\text{NH}_4\text{Cl}$  – Ammonium chloride
- NMR – Nuclear magnetic resonance
- OLS – Ordinary least squares

P – Phosphorous  
PAO – Polyphosphate accumulating organisms  
PHA – Polyhydroxyalkanoates  
Poly-P – Polyphosphate  
PPK – Poly-P kinases  
 $P-PO_4^{1-}$  – Orthophosphate as phosphorus  
PPX – Exopolyphosphatase  
QIA – Quantitative image analysis  
RGB – Red, green and blue  
SBR – Sequencing batch reactor  
TB – Toluidine blue  
TB\_dil – Toluidine blue diluted  
TB\_sd – Toluidine blue non-diluted  
TCA – Tricarboxylic acid cycle  
TSS – Total suspended solids  
VSS – Volatile suspended solids  
 $ZnSO_4 \cdot 7H_2O$  – Zinc sulfate heptahydrate

# 1. Introduction

## 1.1 Motivation

Water is the most important resource for human survival and life in general. There is around  $1,4 \times 10^9$  km<sup>3</sup> of water on Earth, however only 2,5% of the total water volume is drinkable with only a very small percentage of this fresh water being accessible to man [1], [2]. According to the AQUASTAT database of the Food and Agriculture Organization of the United Nations, there is an estimated use of 3,9 km<sup>3</sup> of fresh water per year, with the greatest consumption being agriculture. Out of this amount 56% (2,2 km<sup>3</sup>) is released into the environment as wastewater in the form of industrial or municipal effluent or agricultural drainage water. At the same time and due to increasing population, changing consumption habits, greater urbanization and industrialization, and others, water demand is predicted to increase significantly on the coming decades [3]–[5].

All these factors lead to an immense pollution of fresh water sources which can cause severe health and environmental issues, especially in poorer countries and regions. On average, high-income countries treat about 70% of wastewater, while middle-income countries treat less than 50%. Low-income countries are even worse as they treat less than 10% of their wastewater further escalating the situation for the poor who are sometimes in direct contact with polluted wastewater. It is estimated that, globally, over 80% of wastewater is released to the environment without proper treatment [3]. Although there has been an increase in wastewater treatment levels over the years, further investment and study must be made to achieve even better outcomes [6].

The lack of treatment can also directly affect ecosystems and the services they provide. Both agriculture drainage water and municipal effluent have a high concentration of nutrients such as phosphorous (P) or nitrogen that increase eutrophication of freshwater ecosystems [3]. Eutrophication is the excessive growth of aquatic plants and algae which leads to an increase in water turbidity and stagnation and to a decrease in oxygen levels, ultimately affecting wildlife in those waters [4], [7]. The removal of these nutrients from wastewater is therefore of utmost importance and it is why there are more and more wastewater treatment plants with the purpose to reduce the nutrients levels. The most common biological process for P removal, known as Enhanced Biological Phosphorous Removal (EBPR), uses microorganisms present in the wastewater to incorporate the abnormal levels of P on their cell bodies in the form of inorganic polyphosphate (poly-P). Over time the wastewater will have its phosphate concentration diminished for discharge into natural water sources [7]–[9].

For a proper EBPR system, fast and efficient analytical methods must be a requirement, however current poly-P monitoring techniques mostly consist of time consuming chemical methods which require prior extraction of cell's content or other more complicated, or even more expensive methods [10]. Thus, there is an obvious need for new and faster techniques in this analytical process [8]. This thesis proposes a different approach to this problem by use of an algorithm that identifies and quantifies intracellular poly-P (int. poly-P) using images obtained with staining analysis and brightfield microscopy.

## 1.2 Objectives

This thesis main objective is the development of a quantitative image analysis (QIA) program that allows for the identification and quantification of poly-P inclusions in samples from an EBPR system. In detail, the technological objectives are:

- 1) To review relevant bibliography about poly-P and state of the art techniques and existing tools for poly-P quantification;
- 2) Studying and testing available relevant software tools;
- 3) To develop an algorithm that identifies and quantifies poly-P inclusions;
- 4) Compare the developed algorithm with a traditional poly-P quantification method;
- 5) Validation of the tool.

## 2. State of the art

### 2.1 History of wastewater treatment

Wastewater management has been in existence since early in the known history of mankind. In the Mesopotamian empire (3500 to 2500BC) there were drainage systems relying on storm water designed to carry away wastes into cesspools [11]. The ancient Greeks (300 BC to 500 AD) had a different approach to the problem, they used public latrines with drainage systems that collected sewage and storm water in a collection basin located outside the city which connected to nearby agricultural fields, using the wastewater for irrigation and fertilization[11]. The Romans (800 BC to 450 AD) had a strong sanitation ethic, creating a central sewer system that led wastewater into the Tiber along with aqueducts which provided clean water not only for personal use, but also for cleaning public baths, latrines and even the streets, which were used as open sewers by common people [11], [12].

With the Roman empire's collapse so too did their sanitary ethics. Their far-reaching aqueducts were neglected and ruined over time. These sanitary "dark ages" (450 to 1750 AD) saw some attempts to properly dispose of waste and wastewater with the building of cesspools, sewers and the hiring of sanitation workers to clean the streets, however the monarchy wasn't too concerned with the situation as long as they weren't affected by the smell, which slowed down further development [11], [12].

In the second half of the 18th century came the industrial revolution and with it an increase in science and technology, however it was also the cause for an exponential increase in population, which worsened the already difficult situation of waste management. Due to the increasing population densities the early part of the 19th century was marked by outbreaks of diseases the likes of cholera and typhus, which is now known to be a consequence of water and waste-borne disease [11], [13]. At the time, these diseases were thought to be provoked by miasma, a type of poisoned or foul air caused by rotting waste, leading to the government placing emphasis on sanitation to prevent the spread of the disease. In London, measures such as construction of sewers and water closets helped the situation with wastewater going directly into the sewers instead of cesspools. By doing this, however, the government was unknowingly turning the river Thames into a virtual cesspool since all the sewers were connected into it [11], [13]. In 1854, Dr John Snow proved the link between water and cholera, by showing that water containing contaminated vomit or fecal matter was the cause of the disease[14]. After this the government implemented a system that collected the discharge wastewater before releasing it downriver from the city [13].

Only after this discovery and in the later parts of this century did people realize that pollutants needed to be removed from wastewater, which led to the first instances of wastewater treatment [11]. There were already some instances of chemical treatment in 1740, using lime as a precipitant [15], but the movement only gained favor in the mid to later stages of the 19th century. Despite its advantages, chemical treatment wasn't completely efficient at pollutant removal and created a lot of sludge, which was difficult to dispose of [11]. The search of better treatment methods led to the creation of septic tanks and biological filters. Septic tanks consisted of large cisterns with inlets and outlets located below the surface of water, since wastewater divided itself into 3 components based on density: oils went to the top, water in the mid and settled sludge on the bottom, allowing cleaner water to be removed. It was observed that the number of solids decreased in these tanks with the passage of time, this being credited to anaerobic organisms [12], [16], [17]. Biological filters consisted of a medium used to filtrate wastewater, with porous soil being used. Wastewater that passed through sandy, gravelly soil had a decrease in pollutants, but the method later evolved into synthetic filters [16], [18]. From both methods came the idea that microorganisms could be used for wastewater treatment, which later led to extensive biological treatment research and development in the 20th century [11]. Chemical treatment was casted to the background, even though it is still used for purposes such as disinfection, chemical precipitation, neutralization and others [19].

The main biological treatment process, called activated sludge (AS), was only discovered in 1913. It was previously known that aeration of sewage allowed for oxidation of organic matter, thus diminishing the pollutants in wastewater, but it was common practice to dispose of the sludge that resulted from settling [11], [20]. In 1913, William Lockett and his team decided to aerate new portions of sewage along with previous obtained sludge. They discovered that with each aeration the amount of sludge increased and the period needed for oxidation of organic matter in the wastewater was reduced [20]–[22]. In a mere 25 years, AS was present in hundreds of full-scale wastewater treatment facilities all around the world [23]. Research into this biological process was halted until after the Second World War. From that point on the main, emphasis has been on the removal of suspended solids, organic material and nutrients, disinfection and also overall performance of the method, along with research of AS's many variants and also anaerobic processes [11].



## 2.2 Enhanced biological phosphorus removal

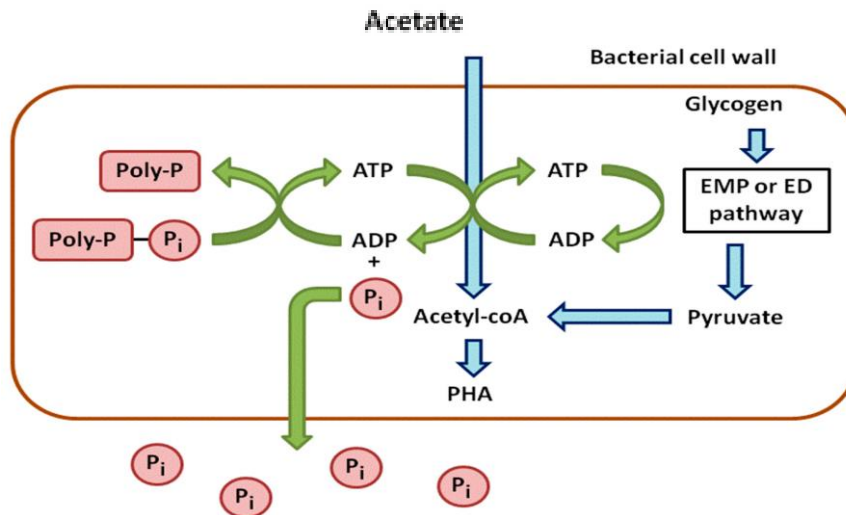
EBPR was discovered in 1975 in South Africa, where its semi-arid environment, along with population growth and high chemical prices, made reuse of water essential, leading to the search of cheaper alternatives to chemical precipitants, which were, at the time, the conventional method for phosphorus and nitrogen removal [24]. EBPR is a variant of the AS process that achieves P removal by recirculating sludge through aerobic and anaerobic conditions [24]. The organisms capable of P removal are generally known as polyphosphate accumulating organisms (PAO). These incorporate P in the form of int. poly-P granules (inclusions), which leads to P removal from the liquid phase by means of cell removal in the waste AS [25].

### 2.2.1 EBPR process – poly-P accumulation

Poly-P accumulation doesn't occur in anaerobic conditions; however, a single aerobic phase is not enough for efficient removal of P from wastewater. Therefore, EBPR is characterized by exposure of AS to both aerobic and anaerobic phases in cycles. Keeping in mind that oxygen can be replaced by nitrite or nitrate, the aerobic phase can be replaced with an anoxic phase or even a combination of both. This will not alter the general process of EBPR.

### 2.2.2 Anaerobic phase

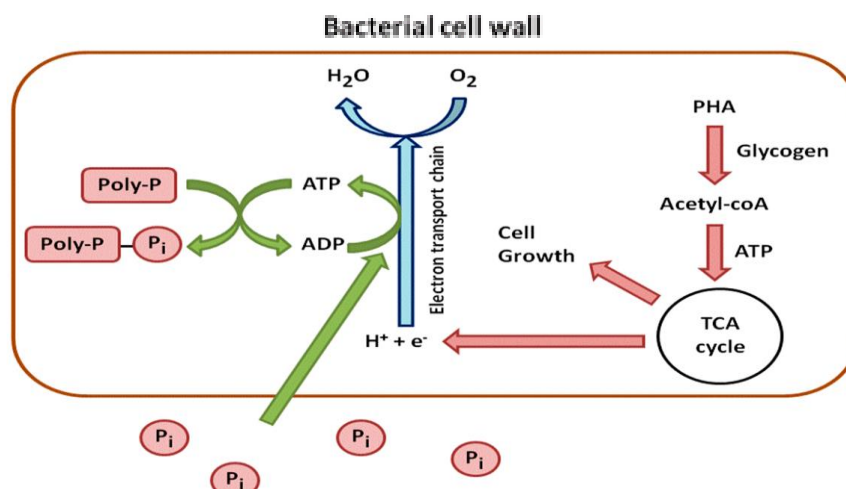
The anaerobic phases of the EBPR process (**Figure 1**) are stressful for PAOs leading them to take up and accumulate carbon sources to prepare for the possibility of long-term oxygen absence. Acetate and other molecules that serve as carbon sources, such as sucrose or propionate, are thus transformed into acetyl-CoA, a reaction that consumes energy leading into the release of P, due to the hydrolysis of ATP into ADP. In this step poly-P reserves in the cell are consumed to maintain ATP levels in the cell. The accumulated acetyl-CoA is then converted and stored as polyhydroxyalkanoates (PHA), a carbon storage compound. The anaerobic phase may seem counterproductive since P is released into the medium, however, the aerobic phase intakes more than enough P to compensate [9], [26].



**Figure 1** Representation of the anaerobic phase biochemical reactions in a PAO. Image adapted from Tarayre [9].

### 2.2.3 Aerobic phase

In the aerobic phase (**Figure 2**), the previously stored carbon compounds, PHA, are used as fuel for cell growth and replenishment of poly-P reserves. The compounds are degraded into acetyl-CoA and enter the tricarboxylic acid cycle (TCA) cycle, which will produce the necessary energy and carbon for new cell growth. Some of the produced energy (ATP) will be used to incorporate P from the surrounding environment in the form of intracellular poly-P granules. In the aerobic stage, the microorganisms will take up almost all the surrounding P, including the one removed during the anaerobic phase. Under optimal conditions, EBPR will consume around 85% of wastewater phosphate, a great improvement to the conventional method's 30% [9], [26].



**Figure 2** Representation of the aerobic phase biochemical reactions in a PAO. Image adapted from Tarayre [9].

### 2.2.4 Phosphorus accumulating organisms (PAO)

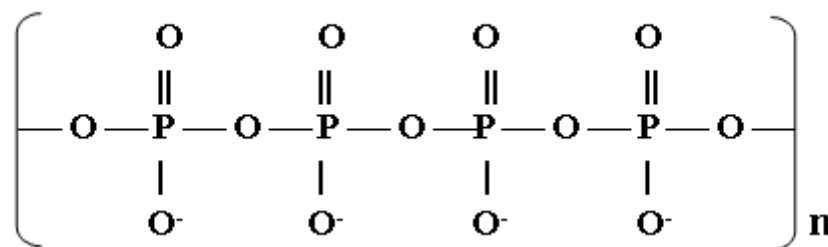
A great number of organisms can be used in EBPR as accumulators of P, from bacteria to algae and even fungi [9], however, bacteria plays the most significant role in this process. For EBPR to be efficient as a P removal method, not all PAOs can be used since the process of incorporating P intracellularly will rely mainly on the ability of the organisms present in the sludge, usually bacteria. Thus, a proper knowledge regarding the molecular mechanisms of poly-P accumulation on bacteria, and perhaps on other types of PAO, could become of help in the biological P removal from wastewater [27].

In the early beginnings of EBPR it was thought that PAOs could only accumulate P under aerobic conditions. However, based on a biological point of view, oxygen is not the only viable electron acceptor with both nitrate and nitrite being possible candidates. It is known that most bacteria possess the ability to denitrify and accumulate poly-P [28] and in fact, literature states that there is P removal in the presence of nitrate in AS systems [29], [30]. This means that there are at least two groups of PAOs, aerobic and denitrifying [28], with the latter able to remove nitrogen alongside P. Many bacteria have been reported as PAO in AS, such as: *Acinetobacter*, *Aeromonas*, *Pseudomonas*, *Paracoccus*, *Bacillus*, and more.[28] *Acinetobacter* spp. are the most common isolates found in the EBPR process, however they are not predominant, in fact several bacteria belonging to different genus are constantly being identified in EBPR processes, with both Gram-negative and Gram-positive bacteria being found.

## 2.3 Poly-P

### 2.3.1 Characteristics

Poly-P is a linear polymer of orthophosphate residues linked by phosphoanhydride bonds, the same found in ATP. This polymer can have lengths from tens to hundreds of residues, depending on its location and metabolic state [8], [10], [31]. Due to its composition (**Figure 3**), it has a highly negative charge. Organic poly-P is usually found in large granules inside the cells, especially large if the organism is a PAO [25], [32].



**Figure 3** Example of poly-P linear structure.

### 2.3.2 Functions

Poly-P is a molecule with a wide range of biological functions, depending on the subcellular compartment where it is located and when it is needed. These functions include substitution for ATP in kinase reactions, regulation of enzymatic activities, capsule of bacteria, regulating some genes expressions, the physiologic adjustments to growth, development, stress and deficiency, and, with special interest to this thesis, disposal of pollutant phosphate in wastewater [33]. The biological removal of P from wastewater through poly-P production has long since been documented [4], [7], [25], [34], with poly-P serving as a high energy storage molecule that, upon hydrolysis, can supply large amounts of energy for biochemical reactions within the cell [8], however, this is not its only function in wastewater treatment since poly-P's characteristics make it useful in other areas. Its highly negative charge, for instance, makes it helpful in preventing microbial adhesion to filter surfaces, which can prove useful in recovering microflora from drinking water [35]. Poly-P can also keep minerals in suspension during industrial processing and decrease bacterial adhesion to soil [36]. Poly-P's reusability in wastewater treatment is yet another reason to why EBPR is the best current choice in P removal.

## 2.4 Poly-P analysis

Quantification and characterization of different forms of P are not easy to accomplish due to the vast uses of P in biology and its presence in several forms. Most quantification methods require a primary step to separate and concentrate the selected P fractions of interest from the rest. Poly-P can associate with other cellular components and depending on its size poly-P solubility will vary, making it a hard target for extraction [10]. Also, extraction has been reported to lead to an underestimation of poly-P values in samples owing to poly-P loss or composition change in the extraction step [37]–[39]. This implies a drop of efficiency in analytical methods with a prior extraction step, making other methods more appealing to the general scientific community.

There are several methods that enable P detection and quantification in biological sources, but not all are suitable for poly-P, with some detecting organic P or ortho-P. These methods can be divided into chemical, biological, molecular, and microscopy/staining analysis.

### 2.4.1 Chemical analysis

Traditional chemical poly-P quantification methods involve a digestion step to convert poly-P to ortho-P. Organic P is also converted therefore increasingly rigorous digestion methods are used to separate both types of P. The ortho-P corresponding to poly-P is then analyzed by ion chromatography or inductively

induced plasma atomic emission spectrometry. Since the analysis is conducted on ortho-P levels and not directly on poly-P these results may not be completely accurate due to high background ortho-P levels and the method's own susceptibility to other P compounds [40]–[42].

High performance liquid chromatography is a very precise chemical quantification method, however it can only identify poly-P with small chain lengths, less than 35 P units, making it not compatible with the quantification of EBPR poly-P [43].

More advanced methods, such as electron ionization mass spectrometry, allow for poly-P quantification without the need of pre-extraction steps, preventing any underestimation [44].

These methods are widely used due to their cost-effectiveness and no need of advanced instruments. Still, they require work and time and are not techniques used for fast quantification of poly-P.

#### 2.4.2 Biological analysis

Biological analysis can be divided into either enzymatic assays or protein affinity labeling. The second method can locate and visualize poly-P at the ultra-structural level tagging and detecting an epitope, the part of poly-P that is detected by the immune system [45]. This technique is not used for poly-P quantification. The first method, however, can quantify poly-P levels.

Enzymatic assays are based not on poly-P itself but on the enzymes responsible for poly-P metabolism. Poly-P is synthesized in bacteria by poly-P kinases (PPK) and degraded by exopolyphosphatase (PPX). One method involves measuring P concentration after poly-P is degraded by PPX. Another measures ATP concentration, since poly-P is converted to ATP by PPK in the presence of an excess of ADP. The reaction products can then be analyzed by chromatography, electrophoresis or other methods [37], [46], [47]. The PPK method is considered more effective for longer poly-P chains, more than 20 P units, while the PPX method seems more suited for the opposite. These techniques are highly dependent on the extraction step beforehand since they require a high poly-P purity [39].

#### 2.4.3 Molecular analysis

Methods that pertain to molecular analysis are great non-invasive techniques that can differentiate between different species of P, such as poly-P with possible quantification. X-ray spectrometry, nuclear magnetic resonance (NMR) and RAMAN microscopy are possible molecular analysis methods for poly-P in wastewater.

When X-ray spectrometers are associated with scanning or transmission electron microscopy it is possible to directly measure the levels of elemental contents of the cells and its inclusions, such as

poly-P granules [48], [49]. However, quantification of environmental samples using electron microscopy is difficult, if not impossible to accomplish. There is also the possibility of fluorescence X-ray microscopy, a technique that shows the spatial concentration of elements and can be applied to poly-P. This method can only be applied to cells larger than 3  $\mu\text{m}$ , making it somewhat undesirable [50], [51].

The NMR technique requires labeled substrates to operate ( $^{31}\text{P}$ ) which are included in poly-P granules through normal cell metabolism [52]. This method allows for a good characterization of P-containing molecules, allowing for the visualization of the spectra of all P species, including poly-P. A drawback from this method is that it recognizes molecules based on their bond, meaning that other molecules that possess phosphoanhydride bonds, such as nucleotides, may interfere with poly-P detection. Also, this technique allows for a relative quantification at best, not being suitable for quantifying total poly-P levels on samples [53].

RAMAN microscopy requires a minimal sample preparation and can be used with complementary molecular techniques for better results [10]. This method has been recently used for the identification and quantification of poly-P in EBPR systems [52], however there is still no literature on this method's ability to differentiate between poly-P's with different chain lengths.

#### 2.4.4 Microscopy and staining analysis

Microscopy and staining analysis involves some of the most common methods used when compound visualization and quantification is involved. These range from fluorometric to simple colorimetric techniques and provide a fast, low-cost and easy set of operations [10].

Fluorometric techniques involve staining of poly-P granules with specific fluorescent dyes. The most common fluorescent staining method for poly-P detection uses 4',6-Diamidino-2-phenylindole dihydrochloride (DAPI) [54], [55]. This dye is not specific to poly-P, staining DNA as well as other compounds, but this behavior can be circumvented with high enough concentrations of DAPI, with DNA-DAPI turning blue while poly-P-DAPI yellow [39], [54], [56]. Methods for quantification of poly-P using DAPI have already been created with promising results [8], [55].

Intracellular poly-P granules are not visible through an optical microscope, unless stained which led to colorimetric methods in poly-P visualization and quantification. These methods were the first to be implemented and are still widely used due to their simplicity. They consist of staining biomass with chemicals (dyes) and observing them through a brightfield microscope [9]. There are two major stains commonly used in EBPR studies, Neisser and Loeffler staining, and both use Methylene Blue (MB) as

their active component. This dye can be replaced with Toluidine Blue (TB), which shares similar properties, and can also be used for poly-P staining. MB is a metachromatic dye with a high positive charge that has a high affinity for poly-P granules, due to their anionic properties. When this dye binds to poly-P it produces a color different from the rest of the cell, allowing for the differentiation between them through microscopy [57], [58]. These dyes alone are already enough for poly-P differentiation, with Loeffler and Neisser methods further increasing the visual difference between cell and poly-P granules.

Loeffler's method has been reported to have a higher efficiency towards larger poly-P granules, with good results in EBPR systems with large poly-P quantities [57]. Through this method, pink-violet poly-P granules are seen amidst a blue background corresponding to the cells [59].

When using Neisser staining method, MB is dissolved in an acidic solution that helps better differentiate poly-P granules and the rest of the cell since MB only binds to molecules with lower pH than the solution, which is the case of poly-P. Afterwards a counterstain solution is added. This method makes Neisser staining more suitable for poly-P staining than Loeffler's since it provides a better contrast between poly-P and the cell. Through this method, purple-black poly-P granules are seen amidst a yellow-brown background corresponding to the cells [57].

In EBPR systems, cell staining after the aerobic phase with any of these methods, should show the entire cell as stained instead of individual granules, due to large amounts of poly-P present in PAO [32].

## 2.5 Quantitative Image Analysis (QIA)

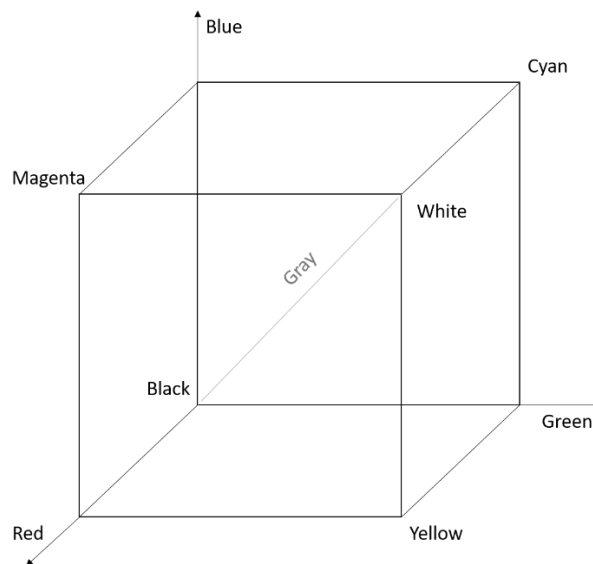
### 2.5.1 Image

An image may be defined as a two-dimensional function,  $f(x, y)$ , where  $x$  and  $y$  are spatial coordinates, and the amplitude in any given coordinate (pixel) is the intensity of the image at that point. When  $f$ ,  $x$  and  $y$  are finite and discrete quantities, the image is a digital image. In monochrome images, the intensity value of pixel is usually known as gray level, since the variation between black and white is what enables us to see a picture. Mathematically speaking, images are  $n$ -dimensional matrices (with  $n$  typically 1-5) with each dimension corresponding to a different parameter or coordinate [60]. For example, in the RGB (red, green and blue) color system, an image consists of three individual monochrome images, each corresponding to their respective color [61]. The junction of these three individual monochromatic images is the original color image.

### 2.5.2 Color spaces

A color can be represented through a series of different mathematical systems. These are called color spaces or color models, and each possess different qualities that enables them to better perform certain tasks when compared to each other.

RGB is the most common color system, but it's not the only one, in fact there are several of these color spaces, such as HSV or CIELAB, which are mathematical representations of sets of color [62]. In RGB color space there are three primary colors: red (R), green (G), and blue (B), that added together can form all other colors including black and white. This model is usually represented as a cube due to the models three main colors (**Figure 4**), which act as the axis for the model. RGB values are conventionally normalized to range from 0 (not saturated) to 1 (fully saturated), however they can also be represented in an 8-bit system, ranging from 0 (not saturated) to 255 (fully saturated) [63].

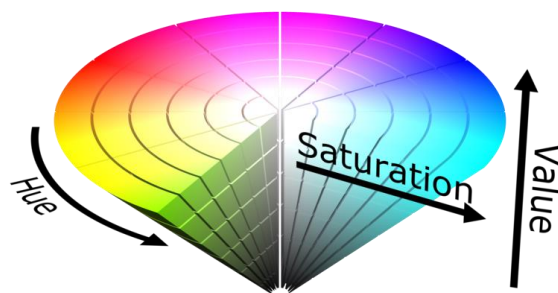


**Figure 4** Representation of the RGB color space in a three-dimensional shape – RGB cube. The x, y and z are represented with the colors red, green and blue respectively, with each coordinate representing a different color. Image adapted from Kuehni [62].

The simplicity and versatility of this color space makes it the most common choice for computer graphics, however RGB is not very efficient when dealing with “real-world” images so other color systems were designed [62], [63]. One color space that takes human perception and interpretation of color into account is HSV (hue, saturation, and value), among others. It was designed to be more intuitive in manipulating color in a time when color had to be specified manually in systems. This color model allowed for a simplification of programming, processing, and end-user manipulation processes, making it suitable for color manipulation such as color identification and segmentation. This color space

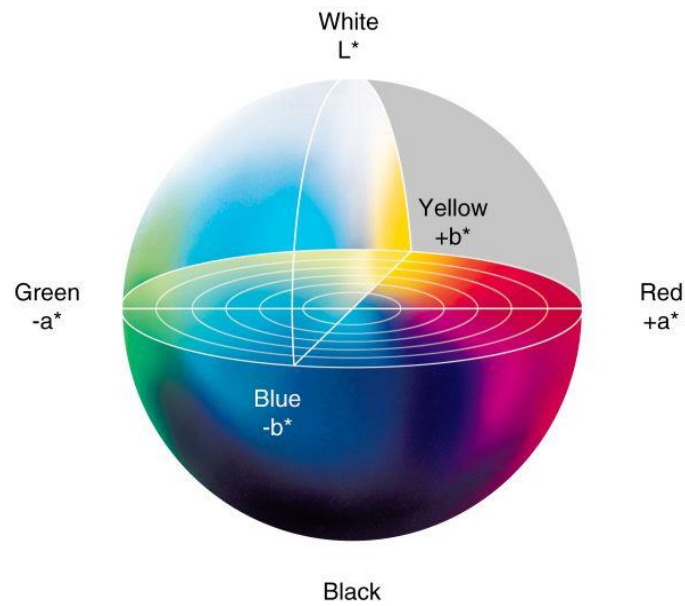


is usually represented as a cylinder or as a cone (Figure 5) [62]. Unlike RGB this color space doesn't have the same range of values in its "dimensions". Hue (H) refers to the basic color and is defined as the angle around the color plane from a reference to the specific color, ranging from 0 (red) to 360 (also red, since it's a circle). Saturation (S) and value (V) are defined as the departure of the color from white and black, respectively, and are defined as two vectors, with saturation being the distance between the central circle point and the specific pure color, and value being the vertical position where the circle plane intersects the grey vertical axis. Both saturation and value conventionally possess normalized values ranging from 0 (white or black) to 1 (specific color), or as a percentage.



**Figure 5** Three-dimensional representation of HSV color space. HSV model is similar to a cylinder with hue (H) being arranged as a radial slice around a central axis of neutral color, saturation (S) as the horizontal depth of the structure, counting from the central axis, and value (V) as the height of the structure. This image was obtained from <http://lib.povray.org> and is a creative property of Michael Horvath.

Another important color space and presently the most important system of color representation is the CIELAB model, also called LAB or  $L^*a^*b^*$ . Created by the *Commission Internationale d'Eclairage* (CIE) in 1976, this model is a result of several years of study and development and is based on the opponent color theory. Designed to be perceptually uniform with respect to human color vision and considered one of the most accurate color systems in the organization of colors, this 3D model is composed of three different axes: a vertical  $L^*$  axis, that relates to the luminance of the color and ranges from black to white and two perpendicular horizontal axis,  $a^*$  and  $b^*$  which range from red to green and yellow to blue, respectively [63]. Conventionally, red and yellow are positive on their axes with values up to +154 while green and blue are negative with values down to -155 (8-bit integer).  $L$  values range from 0 (darker) to 100 (lighter). This relates to the fact that it's considered impossible for any color to be both green and red, or yellow and blue at the same time, according to the opponent color theory [64]. This color space is usually represented as a sphere (Figure 6).



**Figure 6** Representation of CIELAB color space as a sphere. This image was obtained from <http://sheriffblathur.blogspot.com/2013/07/cie-lab-color-space.html>.

All these color spaces are interchangeable, which allows for an easy transformation of a RGB image to HSV, LAB or others, and even vice-versa, through the use of mathematical equations [62], [63].

### 2.5.3 Gamma

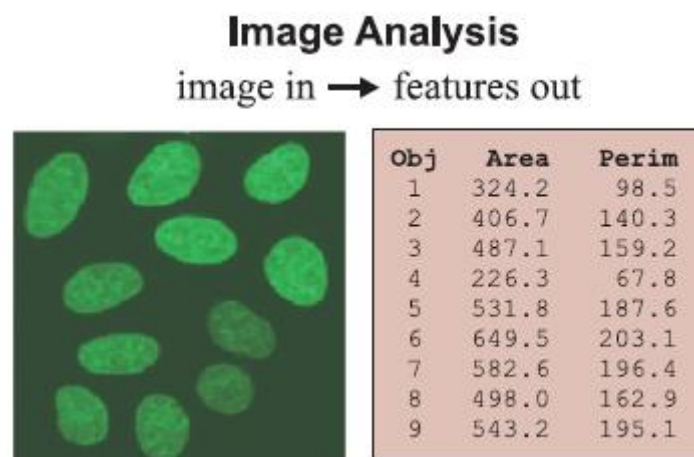
Another important element when dealing with images is gamma correction, or simply gamma. While it is not directly related to color, it can significantly change the perception of colors and brightness of an image.

The intensity of light generated by a physical device is not a linear function of the applied signal, instead being the approximate applied voltage raised to a gamma power. If an image was stored in a linear manner, instead of gamma encoded, it would use an excess number of bits to describe brighter tones due to the higher sensitivity of cameras on this tonal range. Gamma increases the visual quality of an image since humans perceive color and brightness in a nonlinear manner. Gamma correction, which is a simple altering of the gamma value, is therefore a nonlinear operation that optimizes the usage of bits, giving more importance to shadow values that humans are more sensitive to or to highlights that humans cannot differentiate, depending on the purpose of the correction. To put it simply, gamma values define how much dark colors are highlighted in an image. Higher gamma values will darken shadows and darker regions, while lower gamma values will do the opposite, brightening those areas [62], [65].

#### 2.5.4 Image analysis

Image analysis is defined as the act of measuring meaningful object features in an image, with these objects being colors, sections or others (**Figure 7**). The results can be qualitative or quantitative with a tendency for quantification in most fields of research [60].

QIA techniques are becoming increasingly more important in many scientific fields, with a great focus on biology. The increase in both image hardware acquisition efficiency and data storage and processing, along with a need for faster and less subjective approaches has made QIA techniques indispensable and increasingly needed in science [60]. There are numerous QIA methods, adapted for their specific needs, but some steps are common in those methods, such as pre-treatment of images, filtering, and segmentation [66]–[68].



**Figure 7** Example of data obtained on an image with QIA. Image adapted from Meijering and Cappellen [60].

QIA methods have recently been used to obtain valuable information about biological processes, being applied for content and morphology determination in EBPR biomass using bright-field, phase-contrast, and/or fluorescence microscopy [69]–[72].

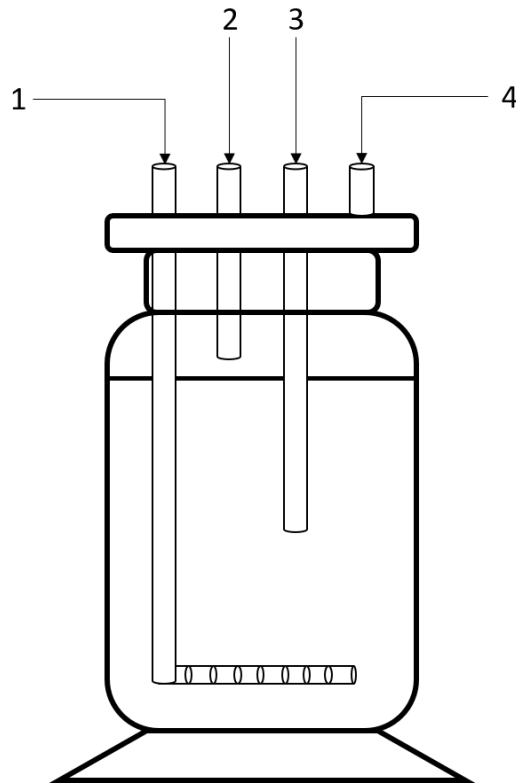
Despite their uses in microorganism or intracellular storage compounds identification (poly-P, PHA, glycogen) [58], [73], these processes are generally not used in routine quantitative analysis, being overshadowed by other more widespread and influential methods, such as chromatography. This implies a need for the development of even more QIA algorithms and techniques in EBPR systems but also on other fields of study.



## 3. Methods

### 3.1 Experimental set-up

Experimental results were obtained in a laboratory-scale sequencing batch reactor (SBR) with a working volume of 4L (**Figure 8**) operated for 29 days. In this experiment, synthetic wastewater was used with acetate as main carbon source and dipotassium phosphate as phosphate source. The chemical oxygen demand (COD)/P ratio was originally planned to slowly decrease from 27 mg COD per mg P- $\text{PO}_4^{3-}$  to 10 mg COD per mg P- $\text{PO}_4^{3-}$  to provide advantages to PAO [8], [74], unfortunately phosphate was not dissolving properly below 20 mg COD per mg P- $\text{PO}_4^{3-}$  so this ratio was kept throughout the duration of the experiment.



**Figure 8** Representation of the lab sized bioreactor used. There were 3 tubes permanently connected to the bioreactor, namely tubes 1, 2 and 3. Tube 1 had holes at the end of it and was connected to a source of compressed air, being responsible for aeration. Tube 2 was responsible for feed, being connected to a jerrycan filled with nutrients, along with the carbon and phosphate source. Tube 3 was responsible for the discharge of the effluent. Opening 4 was usually closed, only used when a sample was taken.

Synthetic feed consisted of the carbon and phosphate source along with other key nutrients according to [8], [75]. Their composition can be found in **Table 1**. Trace metals solution was created separately and then added to the synthetic feed. Their composition is seen in **Table 2**.

**Table 1** Composition and concentration of nutrients that compose the synthetic feed used in the batch reactor. Trace metals composition can be seen in Table 2. This feed was based on [8], [75].

| Synthetic feed (gL <sup>-1</sup> )     |                       |
|--|-----------------------|
| NH <sub>4</sub> Cl                     | 0,59                  |
| MgSO <sub>4</sub> · 7 H <sub>2</sub> O | 0,95                  |
| CaCl <sub>2</sub> · 2 H <sub>2</sub> O | 0,44                  |
| Allyl-N thiourea                       | 0,012                 |
| EDTA                                   | 0,03                  |
| Trace metals                           | 3.16 mL <sup>-1</sup> |

**Table 2** Composition and concentration of metals that compose the trace metals solution used in the synthetic medium. This solution is based on [8], [75].

| Trace metals solution (gL <sup>-1</sup> )             |      |
|---|------|
| FeCl <sub>3</sub> · 6 H <sub>2</sub> O                | 1,5  |
| H <sub>3</sub> BO <sub>3</sub>                        | 0,15 |
| CuSO <sub>4</sub> · 5 H <sub>2</sub> O                | 0,03 |
| KI  | 0,18 |
| MnCl <sub>2</sub> · 4 H <sub>2</sub> O                | 0,12 |
| Na <sub>2</sub> MoO <sub>4</sub> · 2 H <sub>2</sub> O | 0,06 |
| ZnSO <sub>4</sub> · 7 H <sub>2</sub> O                | 0,12 |
| CoCl <sub>2</sub> · 6 H <sub>2</sub> O                | 0,15 |

The system operated with the help of electronic timers, with a cycle of 6h consisting of 120 min anaerobic including 45 min feed, 180 min aerobic, and 60 min settling including 5 min wasting. One on/off control valve was used, bubbling compressed air into the reactor during aerobic periods. The pH was periodically monitored to ensure that it remained around 7.5.

Samples were obtained from the bioreactor immediately after the end of the aerobic period. Samples were also diluted in a 1/10 ratio, with both the diluted and non-diluted samples being stained for posterior image analysis. The remaining samples were stored at -20°C for posterior analytical procedures.

### 3.2 Analytical procedures

Poly-P content, total suspended solids (TSS), and volatile suspended solids (VSS) were always analyzed at the end of the aerobic stage using the same samples used for staining and image acquisition. Poly-P content was measured using a Hach LCK 350 Phosphate Kit. This kit has two different protocols, one for measuring dissolved orthophosphate (a) and the other for measuring the total P (b) content of the sample. The orthophosphate procedure involves adding 0.5 ml of Reagent B to 0.4 ml of sample on a cuvette and closing it with a DosiCap C. The cuvette is then inverted 2 to 3 times so that the cap's frozen content is dissolved. After 10 minutes the sample is ready for analysis. The total P procedure involves a prior hydrolysis step, with DosiCap Zip content and 15 minutes at 170°C using a HT 200 S, before following the same procedure as the orthophosphate analysis. Poly-P concentration was calculated by subtracting the dissolved orthophosphate from the total P content (Poly-P=b-a).

TSS and VSS were measured using the standard methods [76]. Glass fiber filters were placed on top of an aluminum dish and weighted (a). Afterwards, 5 ml of sample (V) was filtered using a standard filtration system (**Figure 9**). The biomass filled filters, along with the aluminum dish, were heated at 105°C for 24h. They were weighted again after cooling down (b). These were then reheated at 550°C during 2h for removal of volatile substances. All that remains in the filters is inorganic matter, usually referred to as fixed solids. The filters were then weighted again after a cooldown period (c). The TSS and VSS were calculated using **Equation 1** and **Equation 2**, respectively, and expressed in gL<sup>-1</sup>.



**Figure 9** Filtration system used during VSS and TSS analysis.

$$\mathbf{TSS = (b - a) \times 1000/V} \qquad \mathbf{Equation 1}$$

$$\mathbf{VSS = (b - c) \times 1000/V} \qquad \mathbf{Equation 2}$$

### 3.3 Staining and image acquisition

Intracellular poly-P inclusions were observed using standard brightfield microscopy at 20 times magnification with MB and TB staining. To keep the process simple, the dyes were used without any additives, as sent by the manufacturer (MB's manufacturer was CONDA Labs while TB's was PanReac). The staining method used was simple: around 0.5 ml of sample was put on top of a slide and spread carefully. The sample was dried at room temperature, after which it was covered with dye for no more than 30 seconds and washed with distilled water immediately after. After drying at room temperature once again, the sample was ready for visualization.

This method was also used for diluted samples with a 1/10 ratio. This dilution was used to prevent biomass overlapping during image visualization. Images were then acquired from all four stained slides thus obtaining 4 sets of around 50 images each: MB diluted (MB\_dil), MB non-diluted (MB\_sd), TB diluted (TB\_dil), and TB non-diluted (TB\_sd). Images were obtained using an Olympus BX51 optical microscope (Olympus, Tokyo, Japan), coupled with an Olympus DP72 camera (Olympus, Tokyo, Japan). Images were acquired at  $1360 \times 1024$  pixels through the commercial software Cell<sup>^</sup>B (Olympus, Tokyo, Japan). Images were then analyzed with the help of a MATLAB™ written image analysis algorithm.

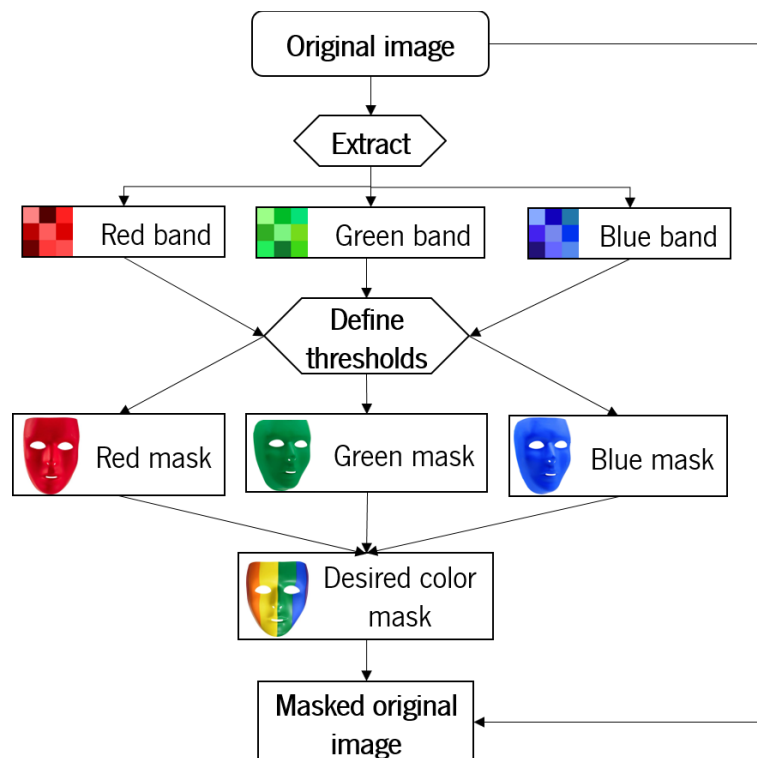
### 3.4 Image analysis

The algorithms developed in this thesis were created using solely MATLAB R2015. The image processing and analysis were based on the identification of poly-P regions, be they intracellular inclusions or other (aggregated, superimposed regions or even ruptured granules), through color segmentation. There are several mathematical representations of color, called color spaces, but for this thesis three were chosen for the color segmentation process: RGB, HSV, and LAB. They were chosen due to their different qualities and their differences in organizing color as to select the most appropriate color space for this work. The mechanism through which color segmentation was performed was basically the same for all color spaces, dividing each image into three different areas representing poly-P, biomass, and background. Two different gamma values and contrast settings were used in an attempt to better visualize the different colors and perform correct segmentation for posterior data analysis. Images obtained by microscopy were organized by day, dye used, and dilution factor, for faster analysis.



### 3.4.1 Early script development

There was an evolution in terms of scripts used during this thesis. A sort of trial and error methodology was applied, with scripts being created and discarded if they did not correspond to expectations or were deemed too difficult to use. Early in this thesis, a script called **RGB\_mask\_segmentation.m**, was created for color segmentation. This script operated on the basis of thresholding RGB values on the obtained microscopy images. The RGB images were divided into their three color bands, onto which a low threshold and a high threshold were designed. The image pixels with values that ranged between both thresholds were used to create color masks while the other pixels were discarded. These masks were then applied in their respective color band creating masked images for each band, or simply put, images corresponding to each color band with their desired values. The combination of all three masked images create a colored mask which, when applied on the original image, led to the final image showing the desired color. The workflow for this script can be seen in **Figure 10**.



**Figure 10** Representation of the **RGB\_mask\_segmentation.m** script's workflow.

### 3.4.2 Late script development

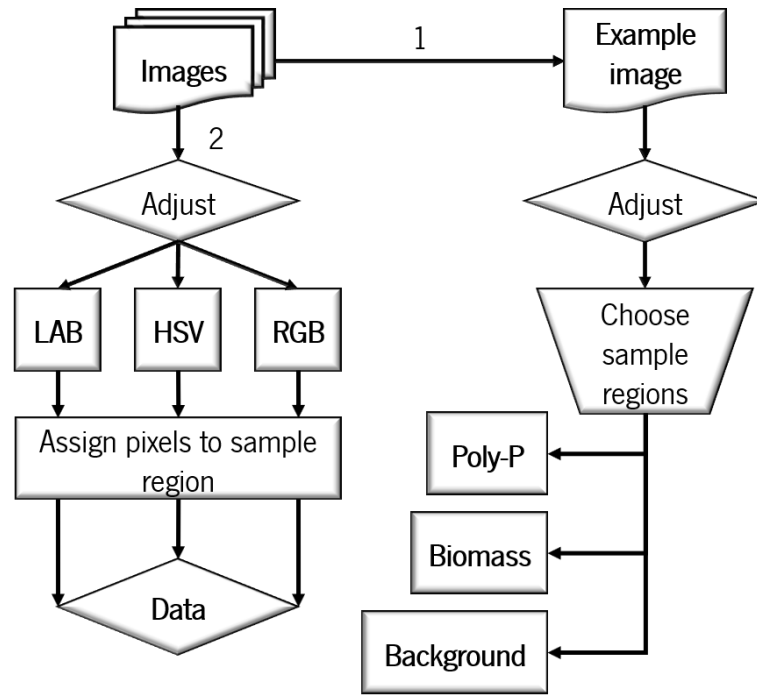
The final two types of scripts used were **Poly\_P\_analysis\_all.m**, a script that utilizes the three color spaces to segment sets of images on the three different regions and saves the data extracted from them on an excel file, and **x\_color\_separation.m**, where x stands for **HSV**, **RGB**, or **LAB**, which are

### 3. Methods

scripts that segment an image based on **Poly\_P\_analysis\_all.m** workflow, and shows them separately on images.

**Poly\_P\_analysis\_all.m** can be further divided into three separate scripts: **Poly\_P\_analysis\_LAB.m**, **Poly\_P\_analysis\_HSV.m** and **Poly\_P\_analysis\_RGB.m**, which process the images in the color spaces associated with their respective names. Their workflow is relatively equal, apart from the original RGB image being transformed into a LAB or HSV image, depending on the script. On any one of these three scripts an image is first selected in order to define the colors belonging to the poly-P, biomass or background region. This enables the script to work with any coloring dye, as long as there's a clear difference between these regions. These regions will then have their pixel intensities averaged on all the color spaces dimensions. This gives the average color for each region to serve as a reference. Other images in the same file folder as the example chosen will have their pixels values compared to the reference colors values through Euclidean distance, which is defined as the shortest distance in a straight line from one point to another. The pixel will be assigned to the region with the shortest distance between their values. The newly assigned regions have their limits smoothed out for a visually better structure. The poly-P and biomass regions will then become RGB images, no matter the previously used color space, and transformed into grayscale and binary images in order to obtain quantitative information. **Figure 11** is a representation of the **Poly\_P\_analysis\_all.m** workflow.

The scripts **x\_color\_separation.m** are similar to the subscripts of **Poly\_P\_analysis\_all.m**, except that it does not store data about all images in a file. A single image is chosen, out of which sample regions will be selected. The image will have its pixels color values compared to the sample regions reference color, such as in **Poly\_P\_analysis\_all.m**, dividing the image in three separate RGB images each with its intended region. The purpose of this script is simply to visualize the separate regions and better identify them.



**Figure 11** Representation of the Poly\_P\_analysis\_all.m script's workflow. The step number 1 happens before step number 2 in order to select the sample regions which will be compared to the remaining images.

### 3.5 Linear regression analysis

In statistics, linear regression is an approach to model the relationship between a dependent variable (or the response variable) and one or several independent variables (as predictor variables)[77]. In this thesis, ordinary least squares multiple linear regression is used to correlate several QIA predictor variables to the analytical response variable. Ordinary least squares (OLS) regression works by minimizing the sum of the squares of the residuals ( $r$ ) on all data points ( $n$ ), a residual being the difference between an observed value ( $y$ ) and the fitted value provided by the model ( $\hat{y}$ ), this way increasing the fit between the data and the model[77], [78]. The sum of the squares of the residuals can be defined **Equation 3**.

$$\sum_{i=1}^n (r_i^2) = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad \text{Equation 3}$$

A linear model is defined as an equation that is linear in its coefficients, meaning that the data points can be modeled by a first-degree polynomial. The linear model can therefore be represented **Equation 4** where the fitted data ( $y$ ) equals the intersection of the y axis with the regression line ( $b$ ) plus the sum of the product of  $n$  number of predictor variables ( $x_n$ ) with their corresponding coefficients ( $\beta_n$ ).

$$y = b + \beta_1 \times x_1 + \beta_2 \times x_2 + \dots + \beta_n \times x_n \quad \text{Equation 4}$$

Since there will be a great number of predictor variables on the models created, a simple response vs predictor variables plot will be incredibly confusing, therefore an added variable plot will be created to represent the model. This type of plot uses the Frisch–Waugh–Lovell theorem [79] to “partial out” variables in the model in order to create a plot that allows visualization of the model according to one predictor, despite having several predictor variables. Basically, this type of plot will visually show the explanatory power of the model as a whole.

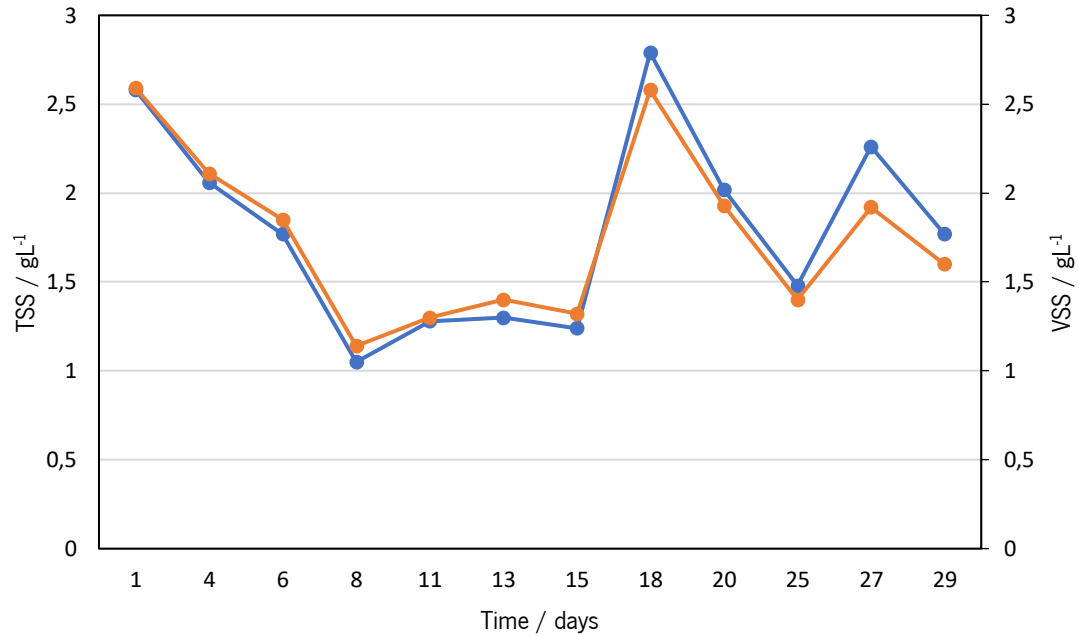
## 4. Results and Discussion

### 4.1 Samples and images

A total of 12 samples were obtained for analytical methods during the 28 days of experiment. These 12 samples were further divided into MB\_dil, MB\_sd, TB\_dil, and TB\_sd sets of images after dilution, staining and image acquisition, which equals to 48 different sets of images, for a total of 2447 images. In this manner around 200 different images were taken for each sample. These images were then pre-processed into four different combinations of two gamma values ( $G_2 = 1$  and  $G_4 = 0.8$ ) and two contrast settings ( $C_1 = \text{no contrast}$  and  $C_2 = \text{higher contrast}$ ) during image analysis, leading to four different datasets of parameters per set of images, each containing poly-P area (pixels), biomass area (pixels), poly-P area per biomass area (ratio), poly-P total grayscale intensity (total grey value), and average grayscale intensity per poly-P pixel (on a scale of 0 to 255).

### 4.2 Analytical results

Standard analytical methods were used to determine the TSS and VSS parameters in the reactor, which were used as both a measurement of biomass growth as well as to relate to QIA information. TSS and VSS values showed high variability during the experimental period, with a decrease from  $2.58 \text{ gL}^{-1}$  and  $2.59 \text{ gL}^{-1}$  on day one to values below  $1.5 \text{ gL}^{-1}$  on days 8 to 15. Some problems with feeding and wasting were observed in the early stages of the experiment, which could have caused the decrease of biomass. Values once again increased above  $1.5 \text{ gL}^{-1}$  after day 15, however both TSS and VSS experienced irregular increases and decreases in their values until the end of the experiment. The highest TSS and VSS values were  $2.79 \text{ gL}^{-1}$  and  $2.59 \text{ gL}^{-1}$  on days 18 and 1, respectively, while the lowest values were  $1.05 \text{ gL}^{-1}$  and  $1.14 \text{ gL}^{-1}$ , respectively, on day 8. TSS and VSS results can be seen in **Figure 12**.



**Figure 12** TSS and VSS content over the experimental period.

Two poly-P parameters were used to relate to QIA information for calibration of the proposed method: intracellular poly-P per volume and per TSS. Only TSS was used since the concentrations were found to be similar to VSS, as can be seen in **Figure 12**. Intracellular poly-P levels increased throughout the experiment, from 25.9 mgL<sup>-1</sup> to 87.1 mgL<sup>-1</sup>, as shown on **Figure 13**. A similar increase can be seen in **Figure 14**, in intracellular poly-P levels per TSS, despite some slight differences due to the abnormal variability in TSS content shown in **Figure 12**. The increase in intracellular poly-P levels was to be expected since the high P concentration in the feed should have led to the selective growth of PAO, who should in turn have incorporated increasing amounts of P as poly-P reserves. Slight traces of microalgae were found in the reactor during the experiment. These organisms competed with PAOs for nutrients and might be the cause for the lower PAO growth, leading to lower intracellular poly-P levels, observed in days 6 and 13 (**Figure 13** and **Figure 14**).

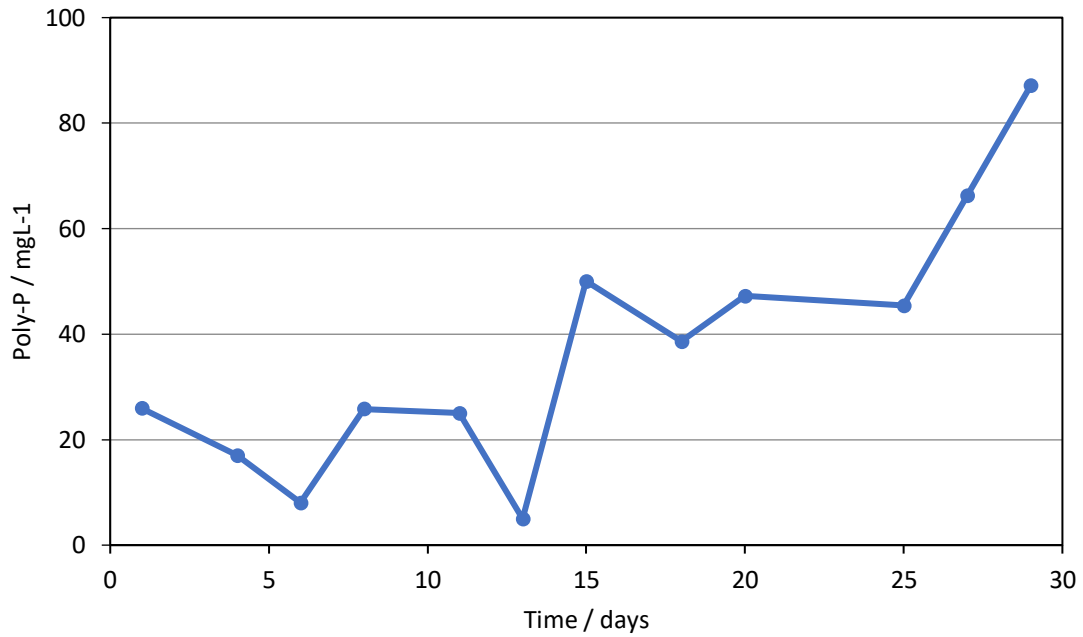


Figure 13 Intracellular poly-P values obtained with Hach LCK 350 Phosphate Kit.

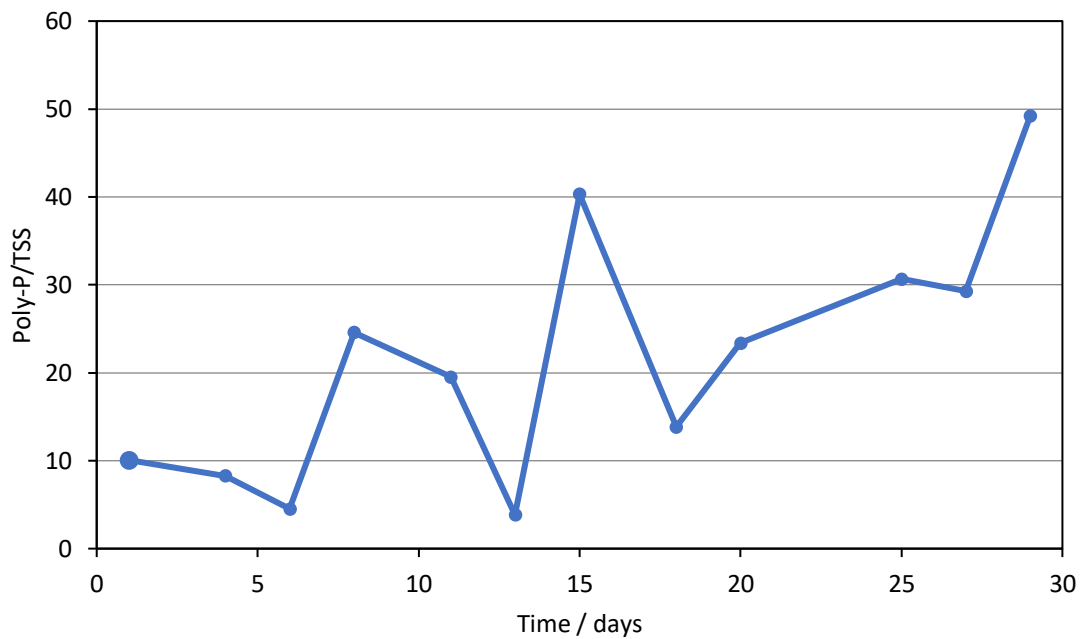


Figure 14 Intracellular poly-P values per TSS content.

### 4.3 QIA results

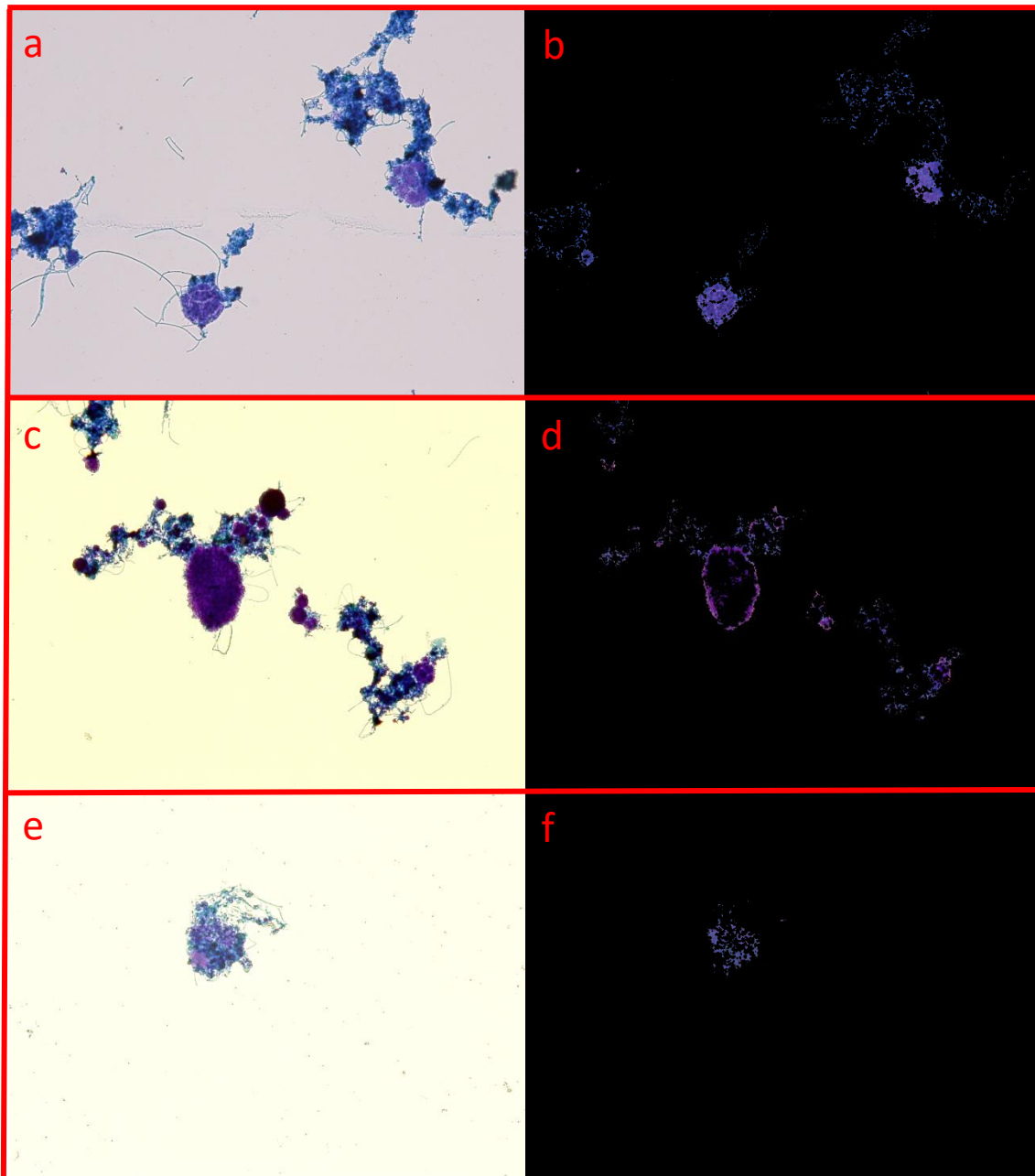
#### 4.3.1 Early script development

The `RGB_mask_segmentation.m` script was able to properly distinguish poly-P regions in an image through color thresholding. The RGB threshold values used for the purple/violet shade of poly-P region

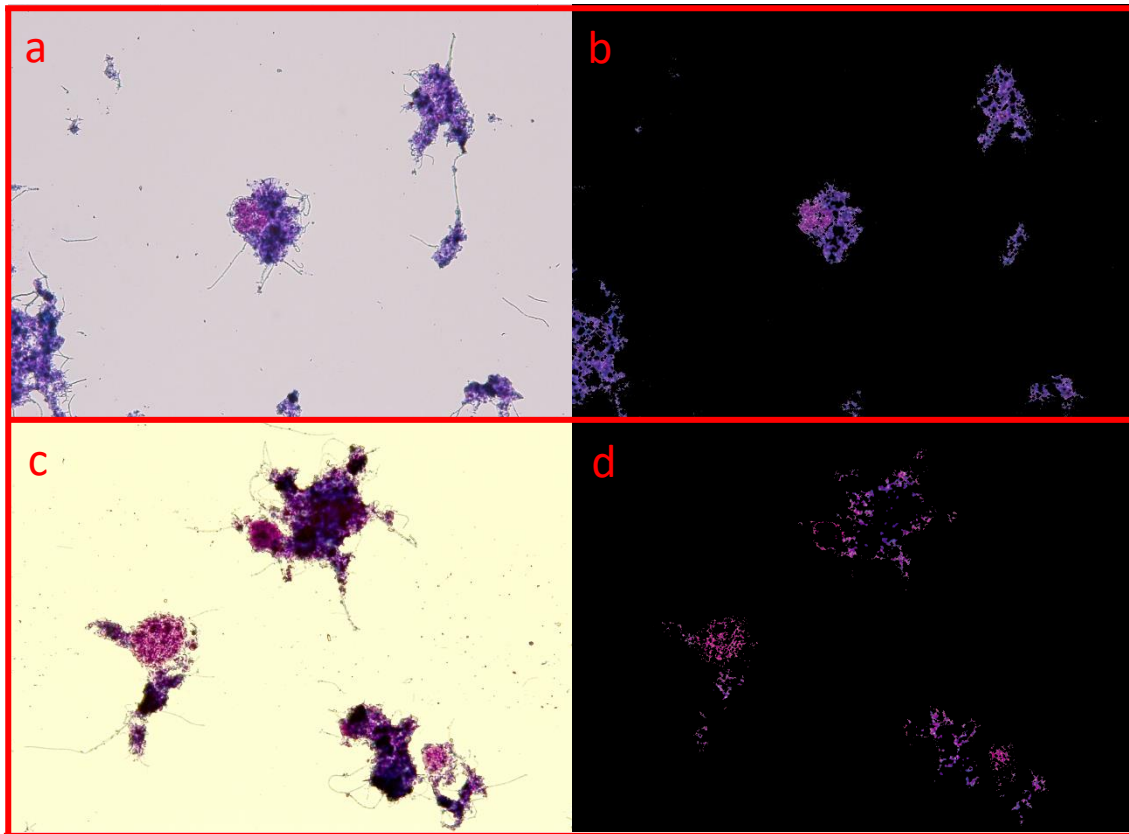
while dyed with MB were 50-180, 10-90 and 110-190, respectively, for the red, green, and blue bands of images. It was noticed that images possessed different gradients of color throughout the experiment, despite similar conditions being applied in all aspects of staining and image acquisition processes. This difference could be seen by comparing the different images of **Figure 15**, which correspond to days 1 (images **a-b**), 13 (images **c-d**), and 29 (images **e-f**) of the experiment. Exposure time and brightness values were deliberately constant over experimental period as to provide similar looking images, however these differences could be due to other imaging software settings that were somehow altered along the experiment. Accidental variations of the staining time (30s by definition) could have also influenced the changes found in poly-P and biomass regions colors (which are lighter in some images and darker on others), as can be seen in **Figure 15**.

Due to these unexpected variations on color, the abovementioned threshold values used managed to properly segment the poly-P region in **Figure 15a**, seen on **Figure 15b**, however they were not the proper values to use on the images obtained from other days of the experiment. It is clearly seen in **Figure 15c-d** and **Figure 15e-f** that the darker or lighter, respectively, shade of purple/violet color did not correspond to the threshold intervals and was therefore not fully identified by the program. This setback causes a need for the discovery of the proper thresholds every day that the samples were collected, and the images obtained. Moreover, two different dyes were used on the process of sample staining, meaning that threshold values also had to differ based on the dye used. **Figure 16** shows examples of the MB thresholds values mentioned above used for segmentation of TB stained images. The results obtained did not conform to the expected poly-P regions, instead representing both poly-P and biomass regions on **Figure 16a-b** and representing some poly-P and biomass regions on **Figure 16c-d**. This implied a proper study of the correct threshold values for each image folder (corresponding to day and dye used) prior to the data analysis itself, which greatly undermined the efficiency of this automated process. For these reasons, this script's development was halted, and another script was implemented in order to avoid this shortcoming and create a faster and easier poly-P quantification method.





**Figure 15** Images obtained through microscopy (a, c and e) and their respective segmented poly-P regions (b, d and f) according to the `RGB_mask_segmentation.m` script using 50-180, 10-90 and 110-190 threshold values on red, green and blue bands respectively. Image a was taken on day 1, c on day 13, and e on day 29. All images were obtained from samples with no dilution, dyed with MB. Gamma and contrast settings were not altered for these images.



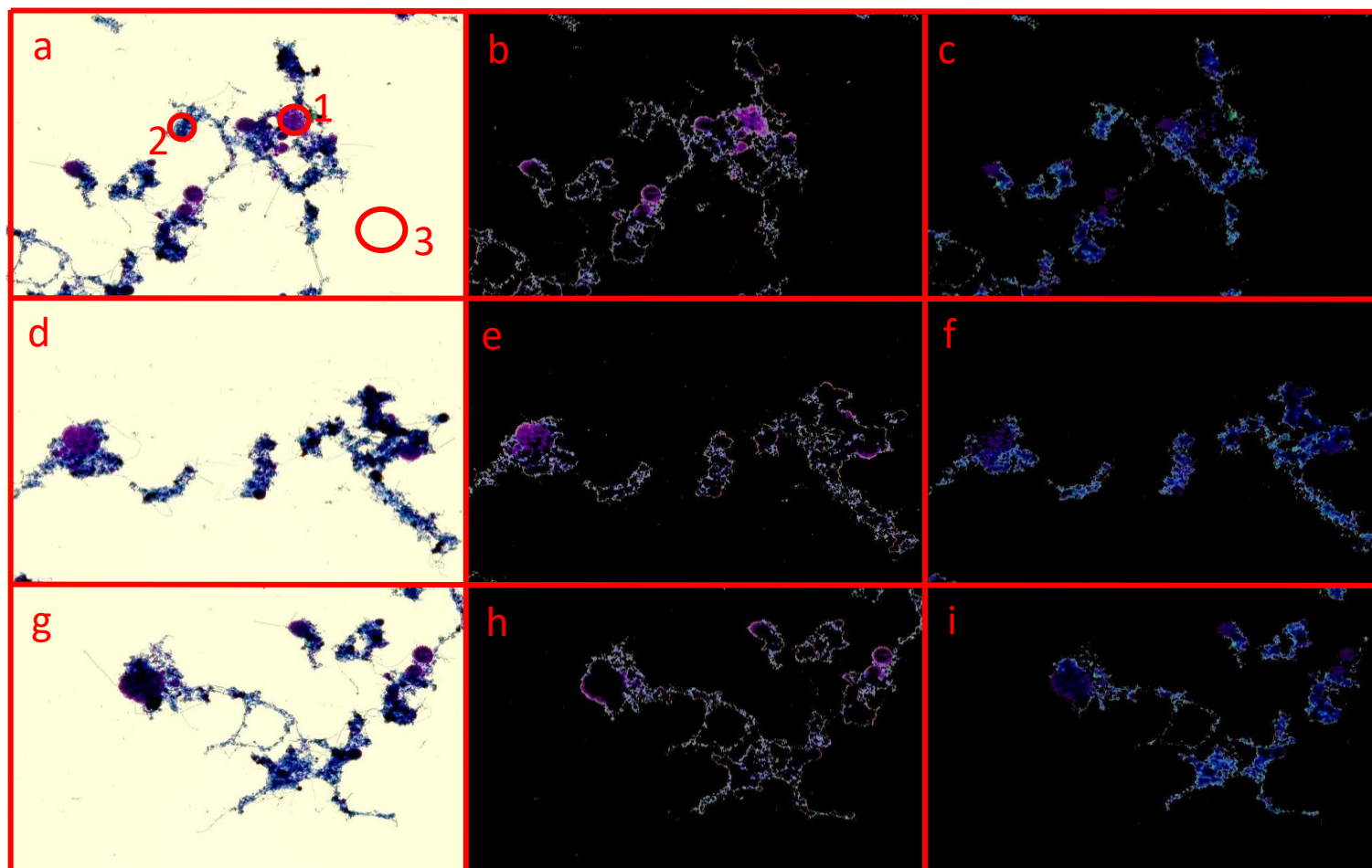
**Figure 16** Images obtained through microscopy (a and c) and their respective segmented poly-P regions (b and d) according to the `RGB_mask_segmentation.m` script using 50-180, 10-90 and 110-190 threshold values on red, green and blue bands respectively. Image a was taken on day 1 and b on day 13. both images were obtained from samples with no dilution, dyed with TB. Gamma and contrast settings were not altered for these images.

#### 4.3.2 Late script development

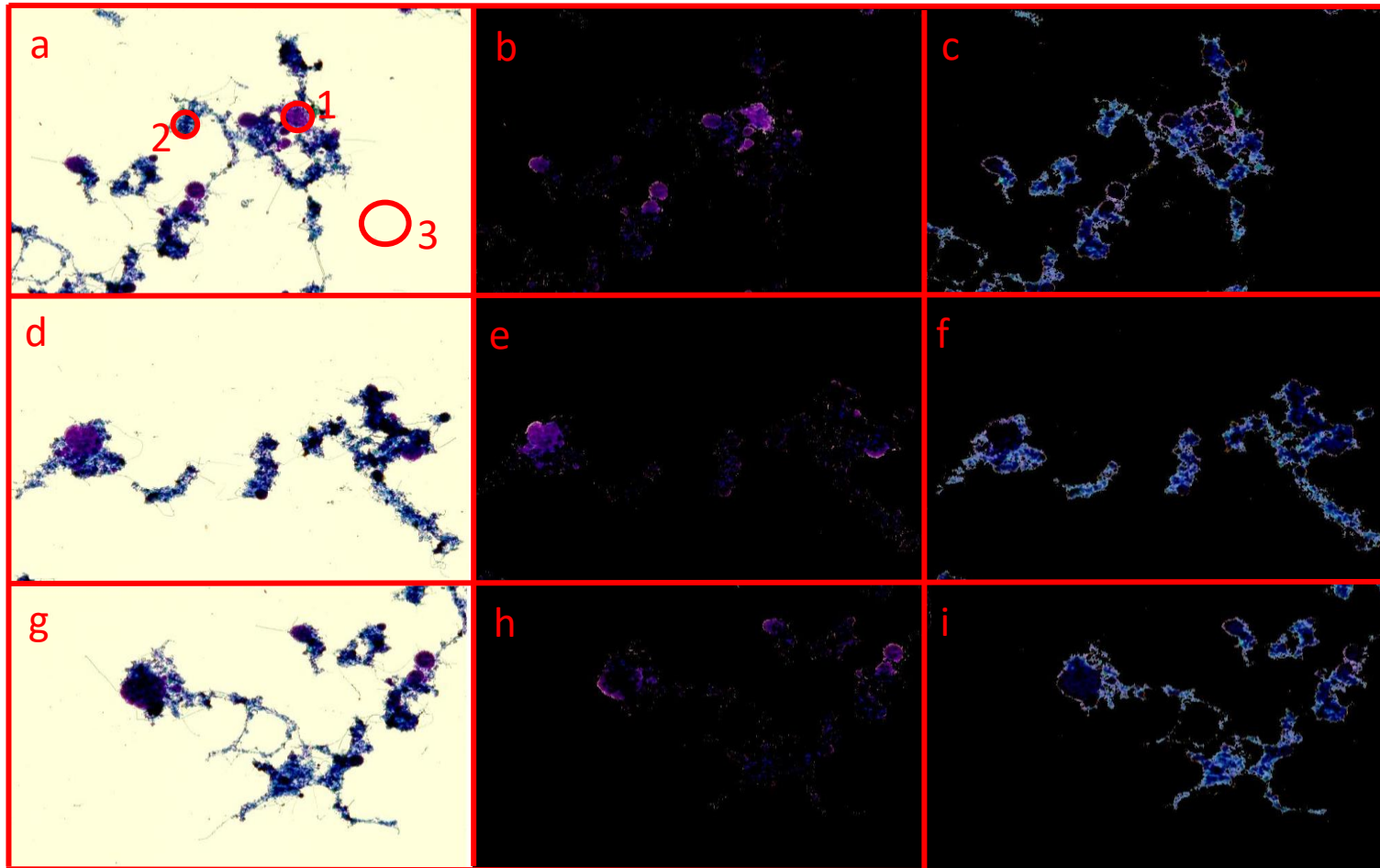
The differences in color and brightness found on the images required the script to reject any manual attempts at defining thresholds, as shown on **4.3.1 Early script development**. The solution found was to define poly-P, biomass and background regions on a selected image, average their pixel values and assign the remaining pixels on the image to a certain region considering the Euclidean distance between their values. This new script, `Poly_P_analysis_all.m` enabled the user to visually select the regions, which excluded the difficulty in finding the proper threshold values. The difference in color between images of different days was also nullified since an image was selected as an example for the region assignment for each different image folder. Although a manual selection by the user is still required, visually choosing regions was faster than finding and changing thresholds for each image folder making this script long term efficient.

A visual representation of the `Poly_P_analysis_all.m` script on MB dyed samples can be seen in **Figure 17**, **Figure 18** and **Figure 19**, with each figure corresponding to the color space used, RGB, HSV, and

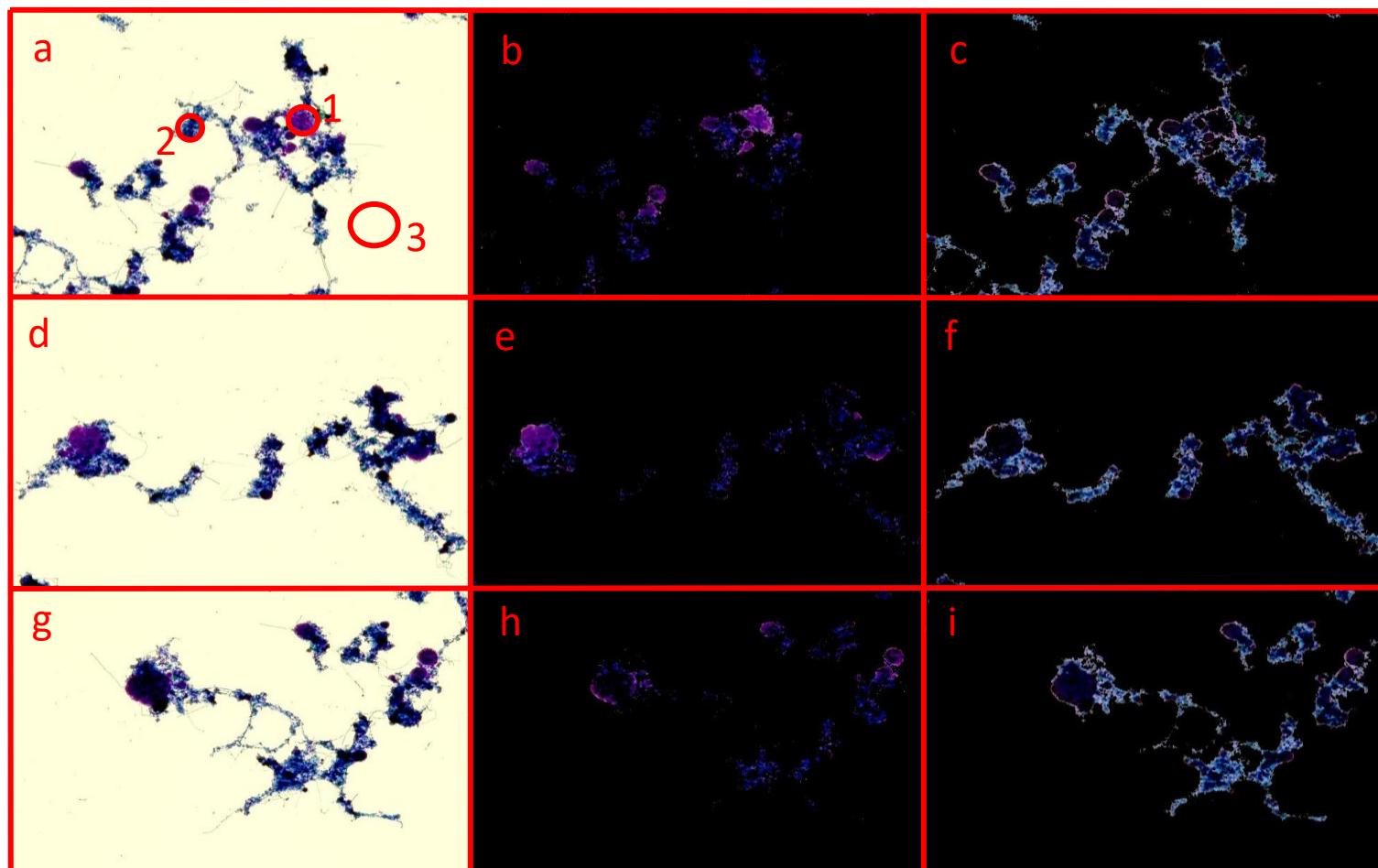
LAB, respectively. The selection of poly-P, biomass and background regions, represented by the numbers 1, 2 and 3, respectively, on image **a** of all three figures provided a good segmentation of these regions on the selected image, seen on the **b** and **c** images, but also on other images in the studied folder, namely images **d** and **g**, with images **e-f** and **h-i** representing the poly-P and biomass regions found, respectively. It could also be seen that despite the selection of the same example areas on the same image, when using different color spaces, a somewhat different result was obtained. It could be noticed that when using the RGB color space, the pixels assigned to the poly-P region were somewhat lighter in color than when using either HSV or LAB color spaces, which seemed somewhat similar. As expected, there was a clear difference between color spaces when segmenting colors, at least visually, since HSV and LAB are designed to be more aligned with human perception and therefore possess more similarities between them than with RGB, that does not have this quality.



**Figure 17** Segmentation of image a into poly-P and biomass regions (b and c, respectively), when using areas 1, 2 and 3 of image a as an example of said regions plus background. The remaining pixels were assigned to each region by Euclidean distance. Images b and c were also segmented into poly-P and biomass regions (images e and f, and images h and i, respectively) with the areas assigned in image a. RGB color space was used for these images. These images were taken on day 8, MB\_sd. Gamma and contrast settings were not altered for these images.



**Figure 18** Segmentation of image a into poly-P and biomass regions (b and c, respectively), when using areas 1, 2 and 3 of image a as an example of said regions plus background. The remaining pixels were assigned to each region by Euclidean distance. Images b and c were also segmented into poly-P and biomass regions (images e and f, and images h and I, respectively) with the areas assigned in image a. HSV color space was used for these images. These images were taken on day 8, MB\_sd. Gamma and contrast settings were not altered for these images.



**Figure 19** Segmentation of image a into poly-P and biomass regions (b and c, respectively), when using areas 1, 2 and 3 of image a as an example of said regions plus background. The remaining pixels were assigned to each region by Euclidean distance. Images b and c were also segmented into poly-P and biomass regions (images e and f, and images h and i, respectively) with the areas assigned in image a. LAB color space was used for these images. These images were taken on day 8, MB\_sd. Gamma and contrast settings were not altered for these images.

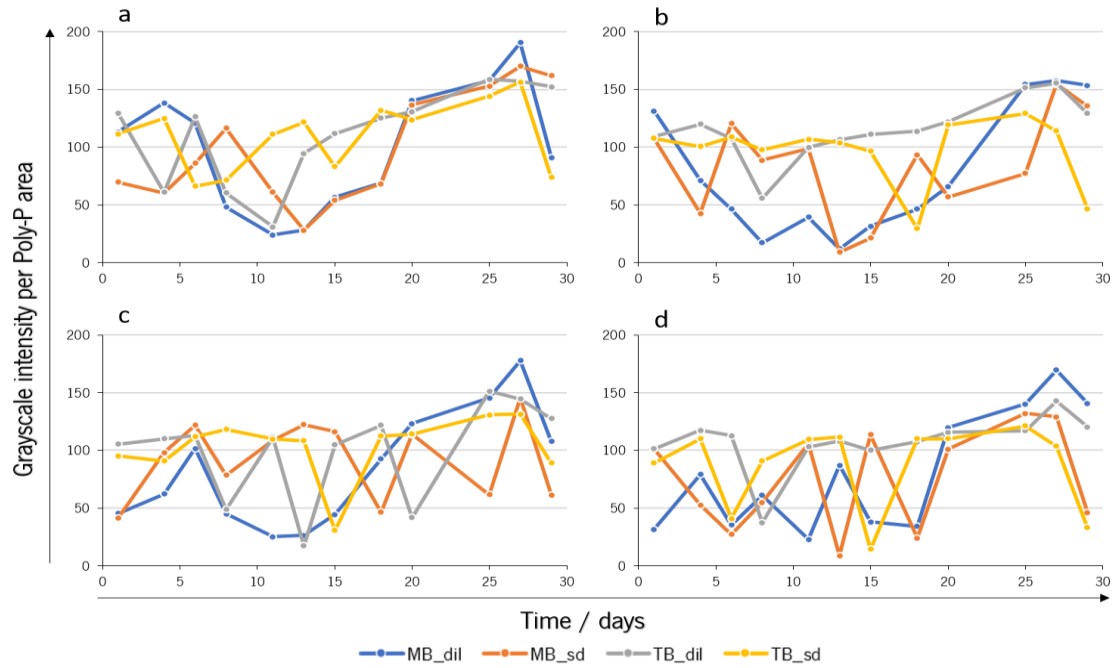
### 4.3.3 Obtained parameters

Five different parameters were obtained by image analysis, using **Poly\_P\_analysis\_all.m**: poly-P area (in pixels), biomass area (in pixels), poly-P area per biomass area, poly-P total grayscale intensity and average grayscale intensity per poly-P pixel.

To understand the importance of intensity in brightfield microscopy, the average grayscale intensity per pixel in the poly-P region was analyzed in all three color spaces and with all gamma and contrast settings. In both RGB (**Figure 20**) and HSV (**Figure 21**) spaces, there were great amounts of variation in grayscale intensity per pixel in a generally random matter, however, in RGB the values differed greatly between gamma and contrast settings, while HSV values were similar no matter the gamma values or contrast chosen. Likewise, LAB color space (**Figure 22**) possessed similar values regardless of gamma or contrast, however, unlike RGB and HSV, its grayscale intensity per pixel values were found to be somewhat constant over time.

In the RGB color space (**Figure 20**), MB\_dil images appeared to have a higher intensity per poly-P pixel in all settings at the final days and early days of the experiment and lower intensity in between, while MB\_sd, TB\_dil and TB\_sd were highly inconstant throughout all stages. **Figure 20a-d** appeared to show a higher intensity between days 20 and 27 in all dyes and dilution factor, with MB\_dil also showing this increase in **Figure 20b-c**. MB\_sd and TB\_dil showed inconstant values on all images, making it difficult for comparison and discussion. TB\_sd showed almost constant values above 100 in intensity, with only some days going below it. The days where this happened changed depending on the gamma and contrast. In this color space, different gamma values and contrast settings seemed to affect grayscale intensity variation, which could be seen by comparing the different values in **Figure 20a-b-c-d**, however these changes appeared to be random. Since these images didn't present an increase similar to **Figure 13** or **Figure 14** it could imply that grayscale intensity per poly-P area does not properly reproduce poly-P concentration values in samples, at least in RGB color space.

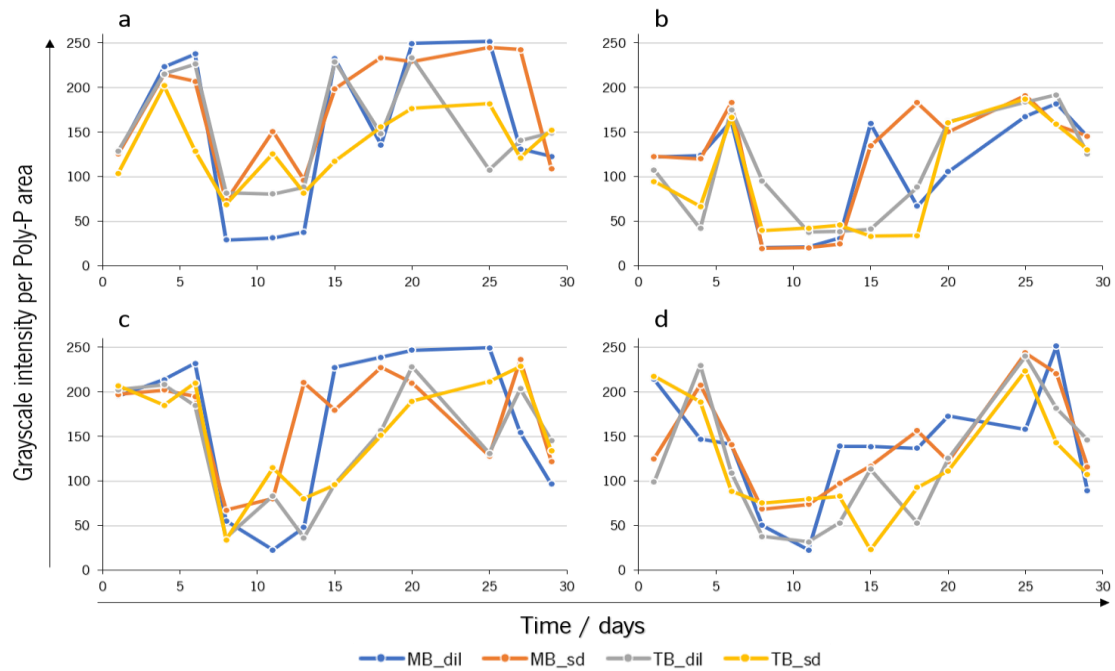
4. Results and Discussion



**Figure 20** Grayscale intensity per poly-P area in RGB color space during the experimental period. Plots a-b and c-d show data from images with 0.8 and 1.0 gamma values, respectively. Plots a-c and b-d show data from images without and with contrast, respectively.

In the HSV color space (**Figure 21**), there was a clear difference between the image analysis results from day 8 to day 13, with a much smaller intensity per poly-P pixel using both dyes and dilution factors, and also in all gamma settings. The results for TB images in **Figure 21b-d** on days 15 and 18 presented intensity values as low as the ones obtained from days 8 to 13. Apart from this difference all graphs showed similar characteristics, with higher values on days 1 and 4, lower values on days 8 to 13 and again higher values on the remaining days. Therefore, in this color space there seemed to be no difference between results obtained with different contrast settings nor with different gamma values. Despite the high grayscale intensity variations during the experiment, the results seemed to indicate that grayscale intensity per poly-P area in HSV may not correlate to poly-P concentration values in the samples.

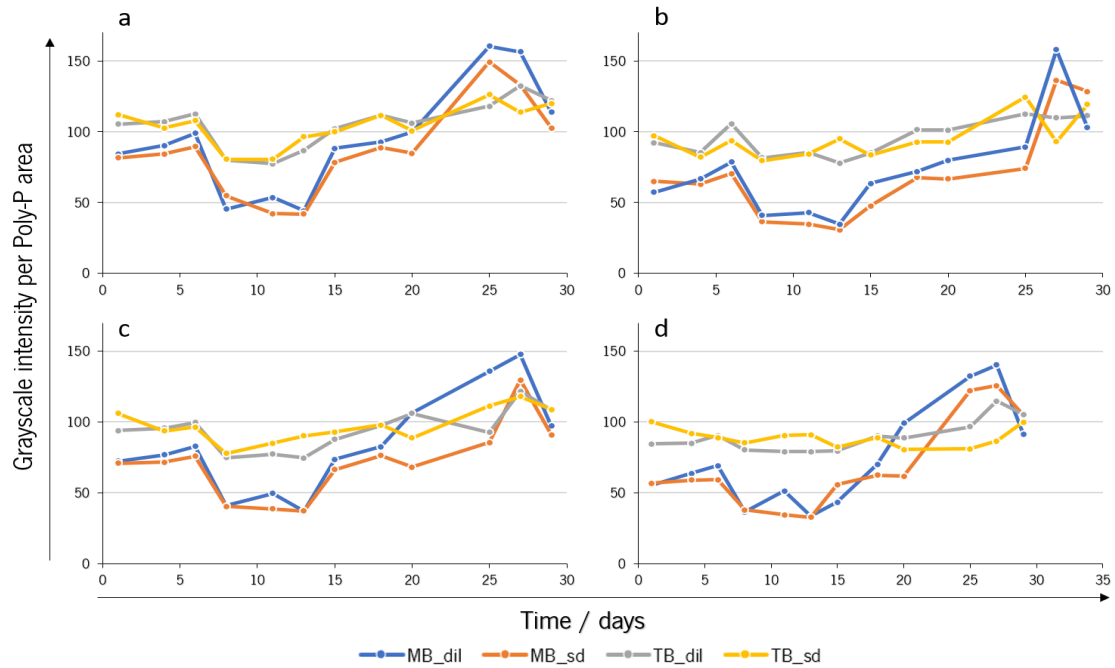




**Figure 21** Grayscale intensity per poly-P area in HSV color space during the duration of the experiment. Plots a-b and c-d show data from images with 0.8 and 1.0 gamma values, respectively. Plots a-c and b-d show data from images without and with contrast, respectively.

The data obtained through images in LAB color space (**Figure 22**) presented a less random group of values than in RGB or HSV with a distinct difference between MB and TB images. TB images had grayscale intensity per poly-P area of around 100 in all settings (**Figure 22a-b-c-d**), while MB images demonstrated a slight decrease from around 60-70 on days 1 to 6, to less than 50 on days 8 to 13 and then a big increase from day 15 and onwards. There seemed to be no difference between results obtained with different gamma values nor with different contrast settings in this color space. This could indicate that MB images in LAB space could correlate to poly-P concentration values. However, taking into account the results obtained in MB images of RGB and HSV spaces, and the results from TB images in the same LAB space, it is highly unlikely that grayscale intensity values per poly-P area in brightfield microscopy obtained images are a determining factor for intracellular poly-P quantification. While grayscale intensity does not seem a proper parameter in this thesis, it must be noted that it shows great potential in fluorescent microscopy[8]. In this technique, and unlike what happens in brightfield microscopy, the concentration of the desired molecule is correlated to the amount of fluorescence in an image, which is quantified by the pixel intensity. It is interesting to note that different QIA approaches require the use of different image parameters.

## 4. Results and Discussion

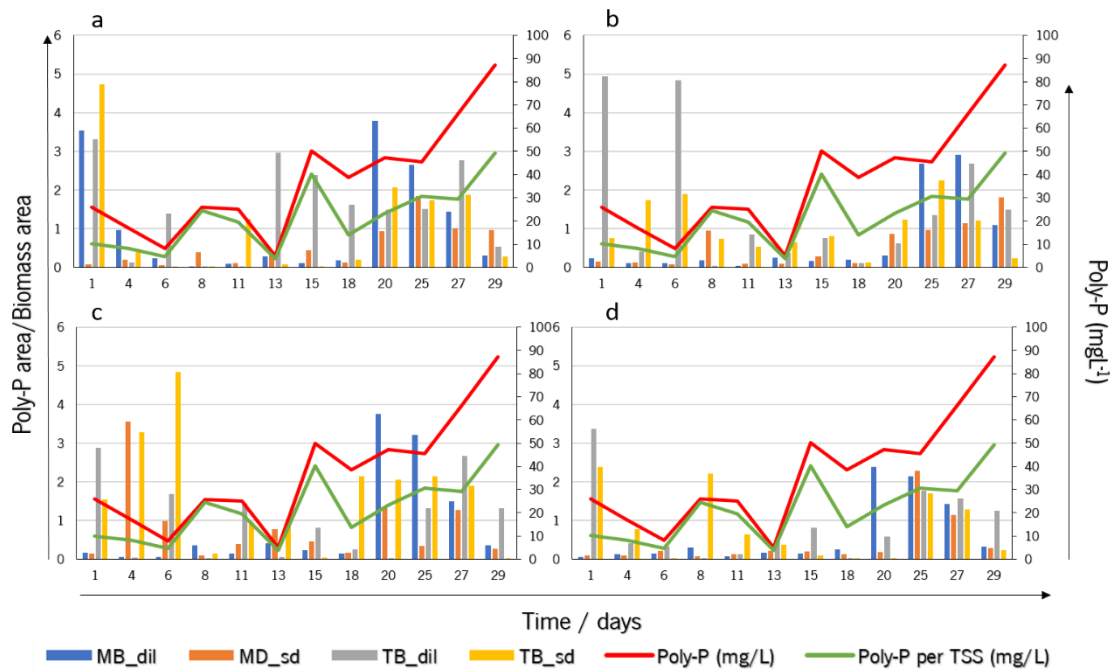


**Figure 22** Grayscale intensity per poly-P area in LAB color space during the duration of the experiment. Plots a-b and c-d show data from images with 0.8 and 1.0 gamma values, respectively. Plots a-c and b-d show data from images without and with contrast, respectively.

Some images had high amounts of biomass and poly-P regions present, almost filling the entire image, while others had smaller amounts that didn't account for even half of the entire image. Therefore, the poly-P area may not properly correlate to the amount of intracellular poly-P in a sample, while the biomass area parameter did not indicate intracellular poly-P concentration at all. However, the amount of poly-P found in biomass was expected to increase during the reactor operation, meaning that the QIA parameter poly-P area per biomass area could be a proper indicator of intracellular poly-P concentration. This ratio was analyzed for all color spaces, gamma values, and contrast settings and compared to the intracellular poly-P concentration quantified by analytical procedures.

In the RGB color space, a high poly-P per biomass ratio was found in TB images (both dil and sd) on both gamma values and contrast settings from days 1 to 8 (**Figure 23a-b-c-d**). **Figure 23a** showed that MB\_dil images provided high poly-P per biomass ratio on days 1 and 4. Regarding **Figure 23c**, MB\_sd showed similar values on days 4 and 6. Days 11 to 18 showed a lower ratio in comparison, with TB\_dil images having higher values in **Figure 23c**. Data from images with higher contrast (**Figure 23b-d**) showed no increase on MB values until day 20. The last days of the experiment, from day 20 onwards, showed a higher ratio value for all gamma values and contrast settings regarding both dyes and dilution factors with the last day (day 29) experiencing a decrease in ratio values in all settings. Overall, there

were signs of an increase on the RGB data, however the initial stages of the experiment and the last day (day 29), were contrary to what was expected when looking at the analytical results shown as red and green lines, especially on TB dyed images. While this did not outright reject RGB as a possible working color space for the desired objective of this thesis, when combined with differences of values between sets of images (dye and dilution factor), it seemed to pose some problems for using this color space alone to correlate to poly-P analytical data.

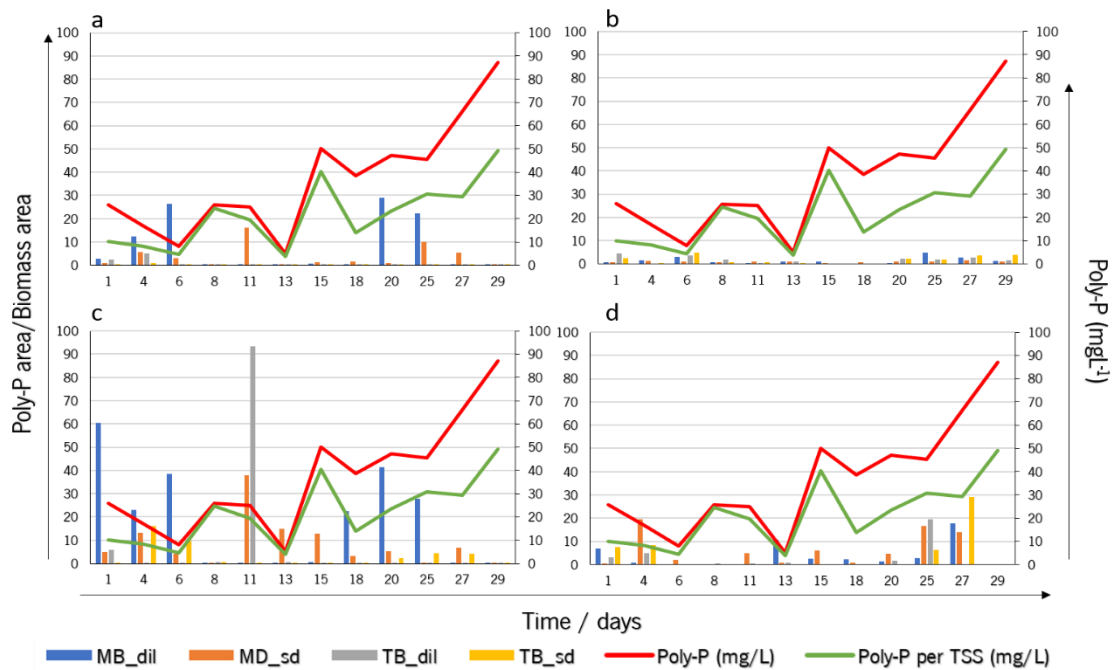


**Figure 23** Average of poly-P area per biomass area for each day in the different gamma and contrast settings for RGB segmented images. Plots a-b and c-d show data from images with 0.8 and 1.0 gamma values, respectively. Plots a-c and b-d show data from images without and with contrast, respectively. The analytically measured intracellular poly-P concentration and intracellular poly-P per TSS are present as a red and green lines for better comparison between parameters.

In the HSV color space (**Figure 24**), results were quite difficult to analyze with several poly-P to biomass ratios being over ten times as large, or more, than their counterparts across the experimental period. MB images with lower gamma values (**Figure 24a-c**) presented several outliers over the experiment, with values reaching 60 in poly-P per biomass area ratio, while normal ratios were generally below 6. On the contrary, Mb images with gamma value of 1 (**Figure 24b-d**) presented “normal” ratio values, with similar to the results obtained in RGB and HSV. TB images showed “normal” results overall, even though TB\_dil images presented the highest ratio value, which was over 90 on G\_4\_C\_1 conditions (**Figure 24c**). Images with higher contrast and lower gamma, settings G\_2\_C\_2 (**Figure 24b**), appeared to show signs of increase in ratio values from days 15 to 27, despite a decrease from days 1 to 15. Images with higher gamma, G\_4\_C\_2, (**Figure 24d**) showed variations in ratio value that was similar to

## 4. Results and Discussion

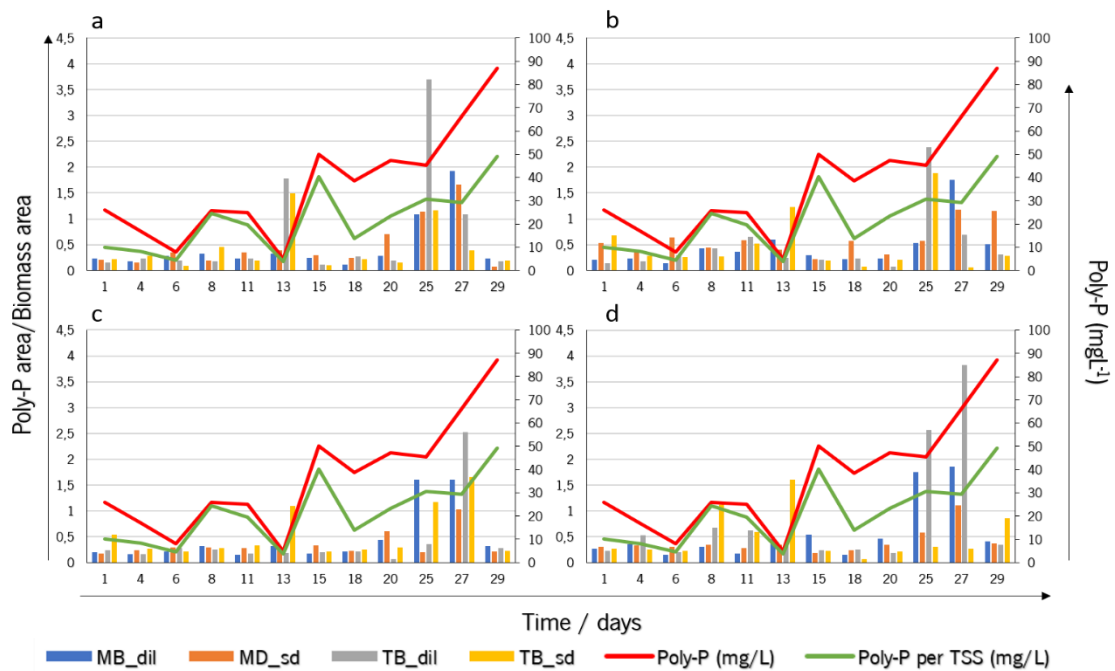
images G\_2\_C\_2, however, the difference between values was enormous with some values reaching 20. Unfortunately, HSV color space results differed greatly making it impossible to properly analyze them. These differences were the highest found between all three color spaces studied. Overall, HSV results seemed to be lacking in all image settings, and while some hope could be seen in **Figure 24b**, it did not seem enough to consider HSV as a good color space for this thesis, and certainly not to use it alone to correlate to poly-P analytical data.



**Figure 24** Average of poly-P area per biomass area for each day in the different gamma and contrast settings for HSV segmented images. Plots a-b and c-d show data from images with 0.8 and 1.0 gamma values, respectively. Plots a-c and b-d show data from images without and with contrast, respectively. The analytically measured intracellular poly-P concentration and intracellular poly-P per TSS are present as a red and green lines for better comparison between parameters.

In the LAB color space (**Figure 25**), the first days of the experiment didn't have higher poly-P to biomass ratio values than the middle or late stages in any image settings or set of images (dye and dilution factor), which, unlike in RGB and HSV, went into accordance with the analytical results obtained for intracellular poly-P concentration per day and per TSS. A small peak was found on day 13 on all image settings for images TB\_sd and on settings G\_2\_C\_1 (**Figure 25a**) for TB\_dil, while analytical results show a decrease in intracellular poly-P on this day. Other than this, the ratio values were relatively uniform throughout the duration of the experiment with increases in days 25 and 27, but not day 29. In **Figure 25a**, an increase was seen on all sets on day 25, all but TB\_sd on day 27 and no increase at all on day 29. Similarly, with a higher contrast (**Figure 25b**), the MB images didn't experience as much of an increase in day 25, showing higher values only on day 27, and MB\_sd in day 29, while TB images

showed higher values on day 27. On settings G\_4\_C\_1 (Figure 25c), day 25 showed an increase in values only for MB\_dil and TB\_sd, day 27 showed an increase for all image sets, and day 29 showed no increase at all, while the remaining days were constant in their values. With increased contrast, G\_4\_C\_2 (Figure 25d), day 25 and 27 showed an increase in ratio values for both diluted image sets (dil), with a small increase for MB\_sd in day 27, while day 29 showed only a small increase for TB\_sd images. TB\_dil images showed higher values on day 13 of all image settings, which surprisingly, went against the analytical results who show a decrease in poly-P concentration. Overall, despite a more uniform set of results than RGB and HSV color spaces, LAB did not show an increase in values from days 15 and onwards, which was expected seeing the analytical results also shown in the image (red and green lines), instead presenting higher values solely on days 25 and 27. These results seem to indicate that LAB alone is not enough to fully represent or correlate to poly-P analytical data.



**Figure 25** Average of poly-P area per biomass area for each day in the different gamma and contrast settings for LAB segmented images. Plots a-b and c-d show data from images with 0.8 and 1.0 gamma values, respectively. Plots a-c and b-d show data from images without and with contrast, respectively. The analytically measured intracellular poly-P concentration and intracellular poly-P per TSS are present as a red and green lines for better comparison between parameters.

Considering the results obtained from all color spaces it was possible to see some similarities, for instance, an increase in values on the later days of the experiment, usually from day 20 and onwards. It could also be seen that, apart from this increase, the analytical results don't quite match any of the individual results from the three color spaces. This lack of similarity between QIA and analytical information could imply some problems with the QIA script itself, or with the staining/image acquisition

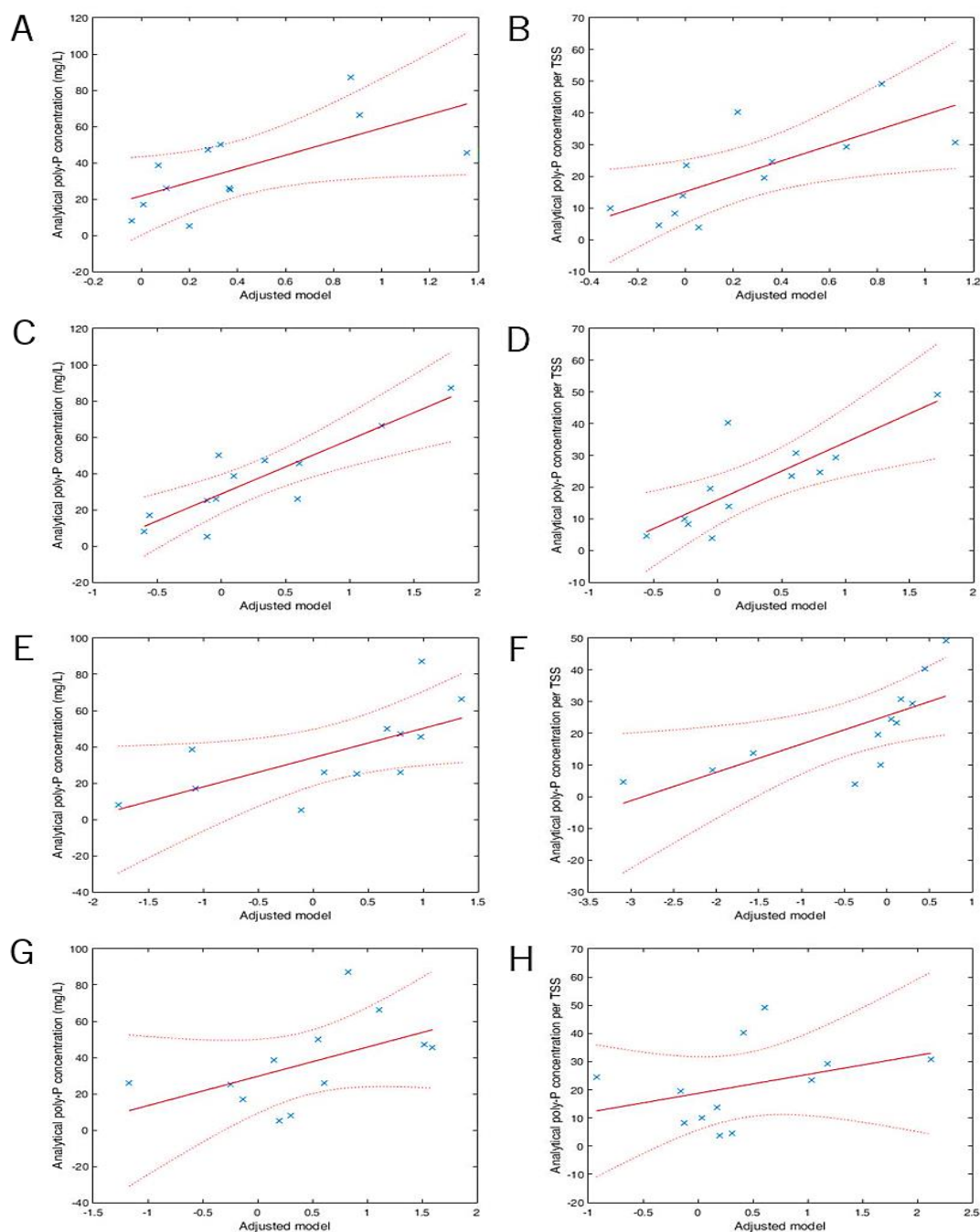
process. The appearance of several superimposed regions on images with both biomass and poly-P on the same space could also be a reason for these differences, since the QIA cannot fully identify these regions.

#### 4.4 Linear regression models

To correlate between QIA information and analytical data, ordinary least squares linear regression was used. Several models were created, with the purpose of finding the one that best fitted the data, and to compare different aspects of the sample images, such as color model used, gamma values and contrast settings, dyes used and dilution factor.

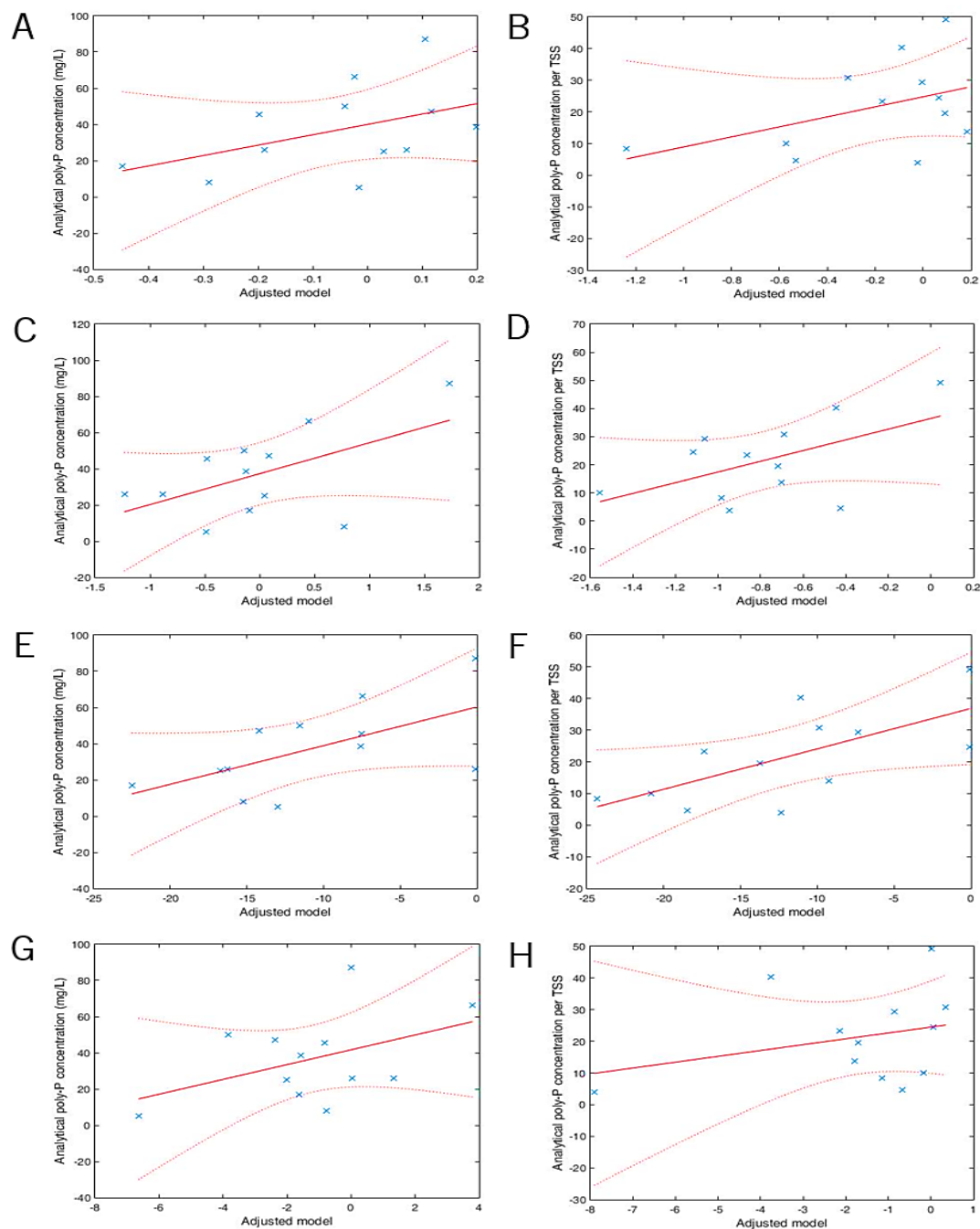
##### 4.4.1 Gamma and contrast

The regression models based on singular gamma and contrast settings showed a really high level of error between the model and the data, seen in **Figure 26**, **Figure 27** and **Figure 28**. It was possible to see several data points far from the fitted regression line and in most cases, there were data points outside of the 95% confidence bounds. The models corresponding to HSV results had the lowest regression coefficient seen, with the G\_4\_C\_2 model having 0.087 when the response variable is intracellular poly-P concentration per TSS (**Figure 27H**). The other HSV models didn't fare much better with their best regression coefficient being 0.450 with the G\_4\_C\_1 model and the same response variable (poly-P per TSS) as its lowest counterpart (**Figure 27F**). LAB models didn't have coefficient values as low as HSV ones, however they also didn't surpass them by much, with the highest coefficient being a disappointing 0.459 on the G\_2\_C\_2 model with intracellular poly-P concentration as the response variable (**Figure 28C**). The worst regression coefficient once again belonged to the G\_4\_C\_2 model, with a measly 0.200 (**Figure 28H**). This model also had the worst regression coefficient value for the RGB models, with a 0.132 (**Figure 26H**). It could be assumed by these values that settings G\_4\_C\_2 were the worst fit for poly-P quantification. The other RGB models presented more acceptable coefficients, with the G\_2\_C\_2 having a promising 0.764 and 0.655 with the response variable intracellular poly-P and intracellular poly-P per TSS, respectively (**Figure 26C-D**). From these values it seemed as if a higher contrast and a different gamma value could be a positive change for QIA of poly-P. Only two gamma values and contrast settings were tested in this thesis, so it is unknown if other combinations of these factors or even other preprocessing image changes could improve the identification and quantification accuracy. Also, RGB models appeared to present relatively better regressions coefficients than the other color spaces, which could imply that RGB was more suitable for this kind of image treatment.



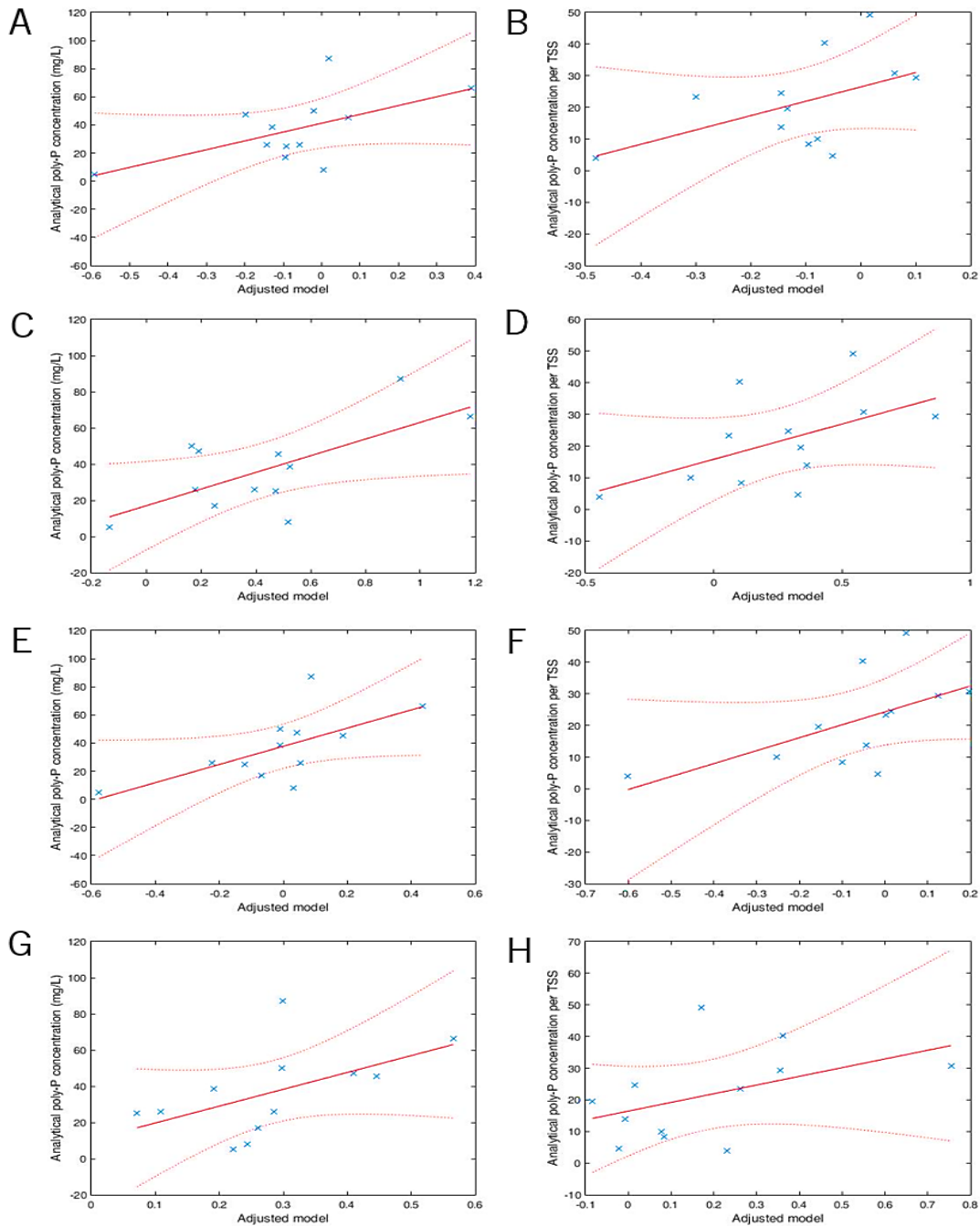
**Figure 26** Linear regression models of RGB treated images using poly-P area per biomass area as predictor variables and analytical intracellular poly-P concentration per day (left side) or per TSS (right side) as response variables. The models were created using different data based on the image settings: models **A** and **B** – G\_2\_C\_1; models **C** and **D** – G\_2\_C\_2; models **E** and **F** – G\_4\_C\_1; models **G** and **H** – G\_4\_C\_2. All image sets corresponding to their respective settings were used in the models (MB\_dil, MB\_sd, TB\_dil and TB\_sd). Data is represented as blue stars, the fitted regression line is represented in red and the model's 95% confidence bounds are represented as light red dotted lines. The models' regression coefficients are the following: **A** = 0.440, **B** = 0.519, **C** = 0.764, **D** = 0.655, **E** = 0.442, **F** = 0.519, **G** = 0.274 and **H** = 0.132. The models' p-values are the following: **A** = 0.334, **B** = 0.218, **C** = 0.023, **D** = 0.079, **E** = 0.331, **F** = 0.218, **G** = 0.639 and **H** = 0.890.

4. Results and Discussion



**Figure 27** Linear regression models of HSV treated images using poly-P area per biomass area as predictor variables and analytical intracellular poly-P concentration per day (left side) or per TSS (right side) as response variables. The models were created using different data based on the image settings: models **A** and **B** – G\_2\_C\_1; models **C** and **D** – G\_2\_C\_2; models **E** and **F** – G\_4\_C\_1; models **G** and **H** – G\_4\_C\_2. All image sets corresponding to their respective settings were used in the models (MB\_dil, MB\_sd, TB\_dil and TB\_sd). Data is represented as blue stars, the fitted regression line is represented in red and the model's 95% confidence bounds are represented as light red dotted lines. The models' regression coefficients are the following: **A** = 0.203, **B** = 0.202, **C** = 0.302, **D** = 0.291, **E** = 0.358, **F** = 0.450, **G** = 0.194 and **H** = 0.087. The models' p-values are the following: **A** = 0.773, **B** = 0.775, **C** = 0.584, **D** = 0.605, **E** = 0.477, **F** = 0.317, **G** = 0.790 and **H** = 0.948.





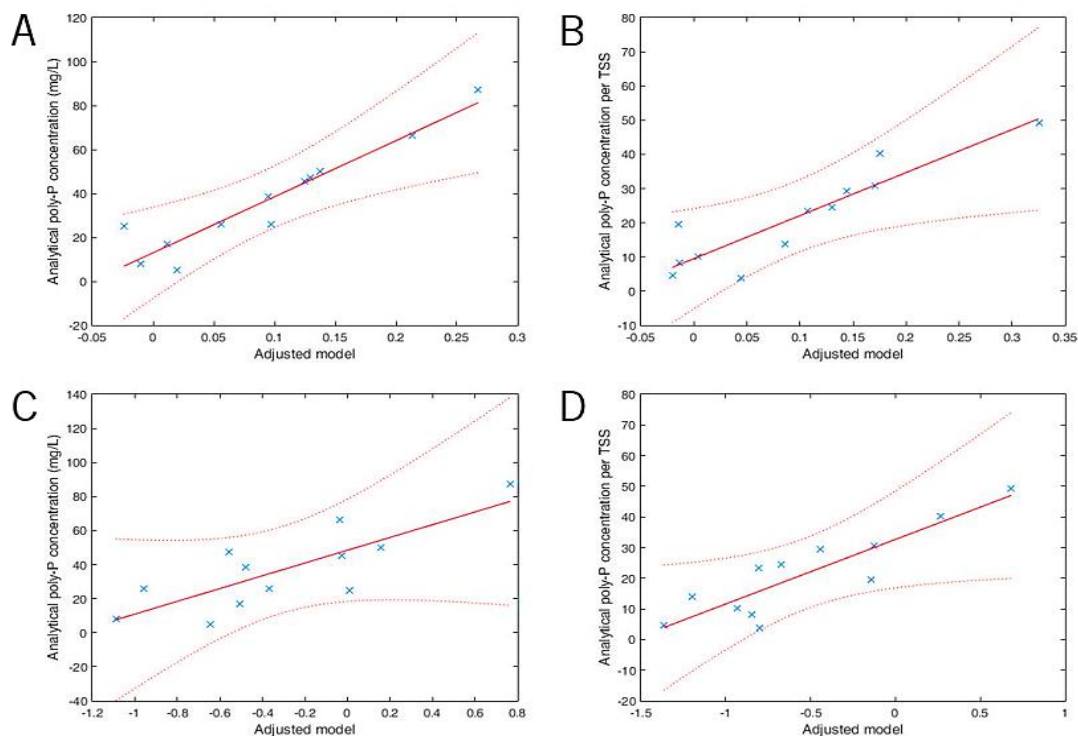
**Figure 28** Linear regression models of LAB treated images using poly-P area per biomass area as predictor variables and analytical intracellular poly-P concentration per day (left side) or per TSS (right side) as response variables. The models were created using different data based on the image settings: models **A** and **B** – G\_2\_C\_1; models **C** and **D** – G\_2\_C\_2; models **E** and **F** – G\_4\_C\_1; models **G** and **H** – G\_4\_C\_2. All image sets corresponding to their respective settings were used in the models (MB\_dil, MB\_sd, TB\_dil and TB\_sd). Data is represented as blue stars, the fitted regression line is represented in red and the model's 95% confidence bounds are represented as light red dotted lines. The models' regression coefficients are the following: **A** = 0.337, **B** = 0.250, **C** = 0.459, **D** = 0.283, **E** = 0.417, **F** = 0.344, **G** = 0.290 and **H** = 0.200. The models' p-values are the following: **A** = 0.517, **B** = 0.686, **C** = 0.304, **D** = 0.621, **E** = 0.372, **F** = 0.503, **G** = 0.608. and **H** = 0.779.

## 4. Results and Discussion

## 4.4.2 Dye used

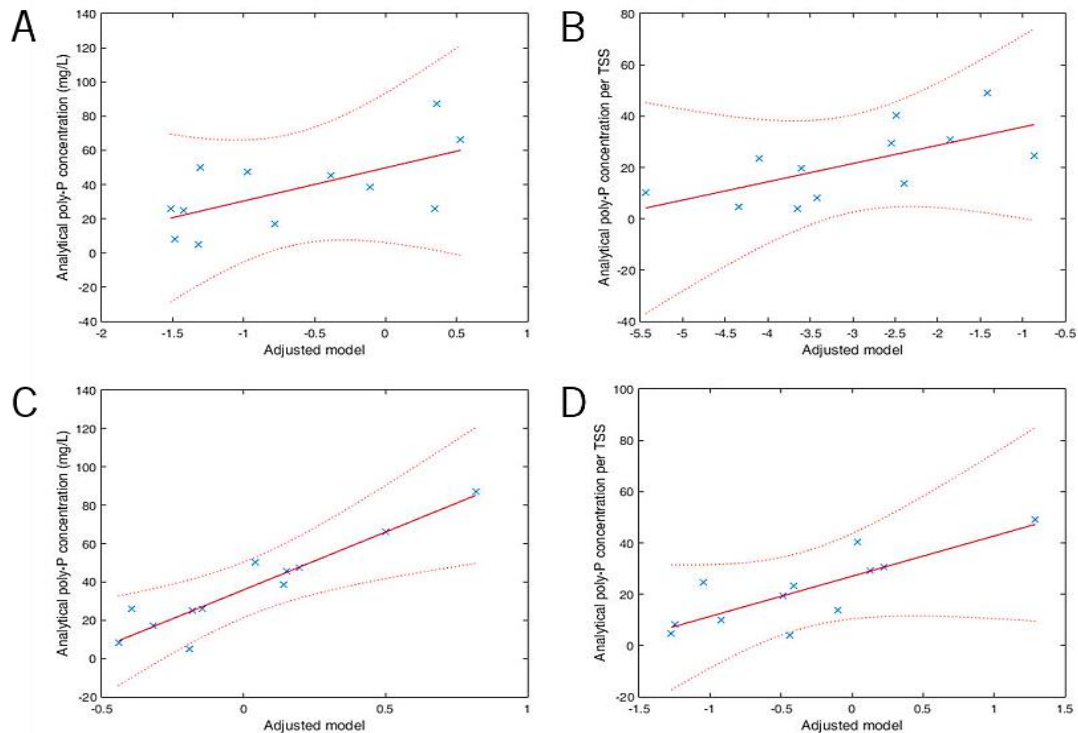
Models based on the dye used (MB or TB) provided a better fit to data than models based on gamma and contrast settings and it was possible to see the data closer to the models regression line without crossing the 95% confidence bounds in **Figure 29**, **Figure 30** and **Figure 31**.

RGB regression models (**Figure 29A\_B**) of MB stained images had better fit to data than their TB counterparts (**Figure 29C\_D**), with satisfactory regression coefficients of 0.891 and 0.824, varying on the response variable, while TB models had lower values of 0.640 and 0.787.



**Figure 29** Linear regression models of RGB treated images using poly-P area per biomass area as predictor variables and analytical intracellular poly-P concentration per day (left side) or per TSS (right side) as response variables. The models were created using different data based on the dye used: models **A** and **B** – MB; models **C** and **D** – TB. Data is represented as blue stars, the fitted regression line is represented in red and the model's 95% confidence bounds are represented as light red dotted lines. The models' regression coefficients are the following: **A** = 0.891, **B** = 0.824, **C** = 0.640 and **D** = 0.787. The models' p-values are the following: **A** = 0.194, **B** = 0.349, **C** = 0.713 and **D** = 0.434.

In HSV models, TB data had higher regression coefficients than MB, contrary to what occurred with the RGB color space. They presented a satisfactory coefficient of 0.883 in the TB model with intracellular poly-P concentration as the response variable (**Figure 30C**), while the other TB model only had 0.651 (**Figure 30D**). The MB models presented even lower regression coefficients with both response variables, having 0.403 and 0.431 (**Figure 30A-B**, respectively).

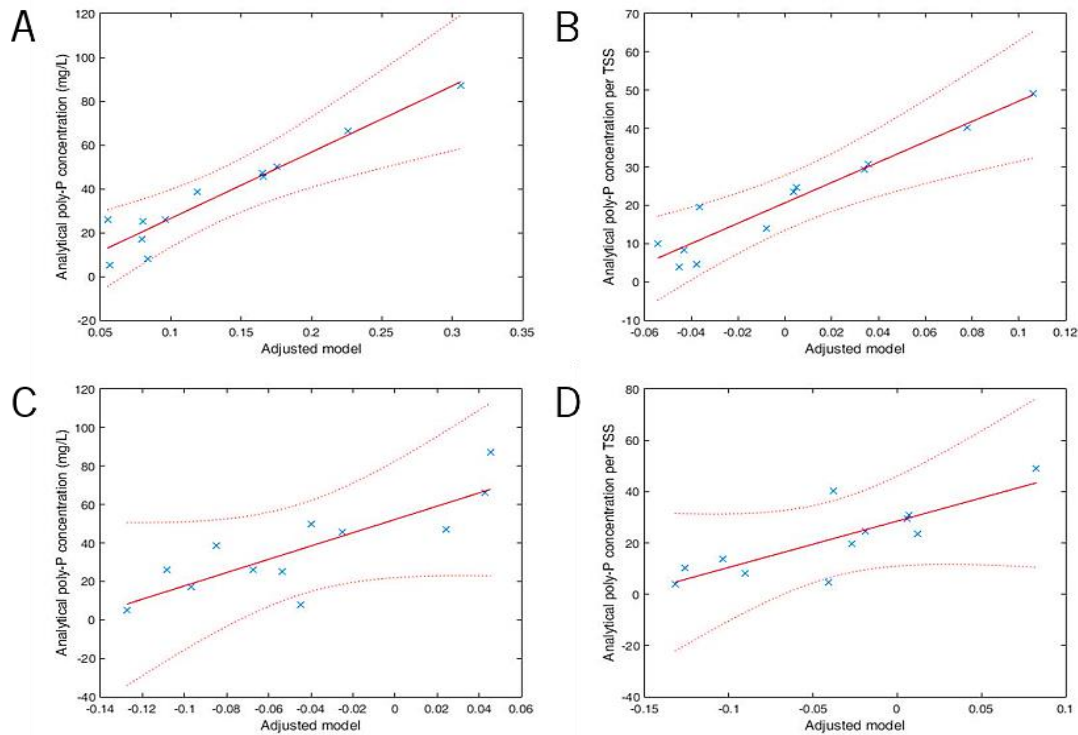


**Figure 30** Linear regression models of HSV treated images using poly-P area per biomass area as predictor variables and analytical intracellular poly-P concentration per day (left side) or per TSS (right side) as response variables. The models were created using different data based on the dye used: models **A** and **B** – MB; models **C** and **D** – TB. Data is represented as blue stars, the fitted regression line is represented in red and the model's 95% confidence bounds are represented as light red dotted lines. The models' regression coefficients are the following: **A** = 0.403, **B** = 0.431, **C** = 0.883 and **D** = 0.651. The models' p-values are the following: **A** = 0.946, **B** = 0.931, **C** = 0.212, and **D** = 0.696.

In LAB models, the MB results (**Figure 31A-B**) once again displayed a better fit to the model than TB ones (**Figure 31C-D**) with satisfactory regression coefficients of 0.920 and 0.919 depending on the response variable. TB models were nowhere near their MB counterparts with disappointing coefficient values below the 0.700 line.

Considering these models, it seems that MB data better described the analytical results, since they display the highest coefficient values in both RGB and LAB color spaces. However, TB models possessed somewhat reasonable coefficient values in both RGB and LAB, with acceptable values in the HSV color space. This could mean that HSV color space could better identify the different tones of color between poly-P and biomass regions with TB than with MB. The opposite could be said for RGB and LAB color spaces. Still, the higher coefficient values indicated better overall results with MB images.

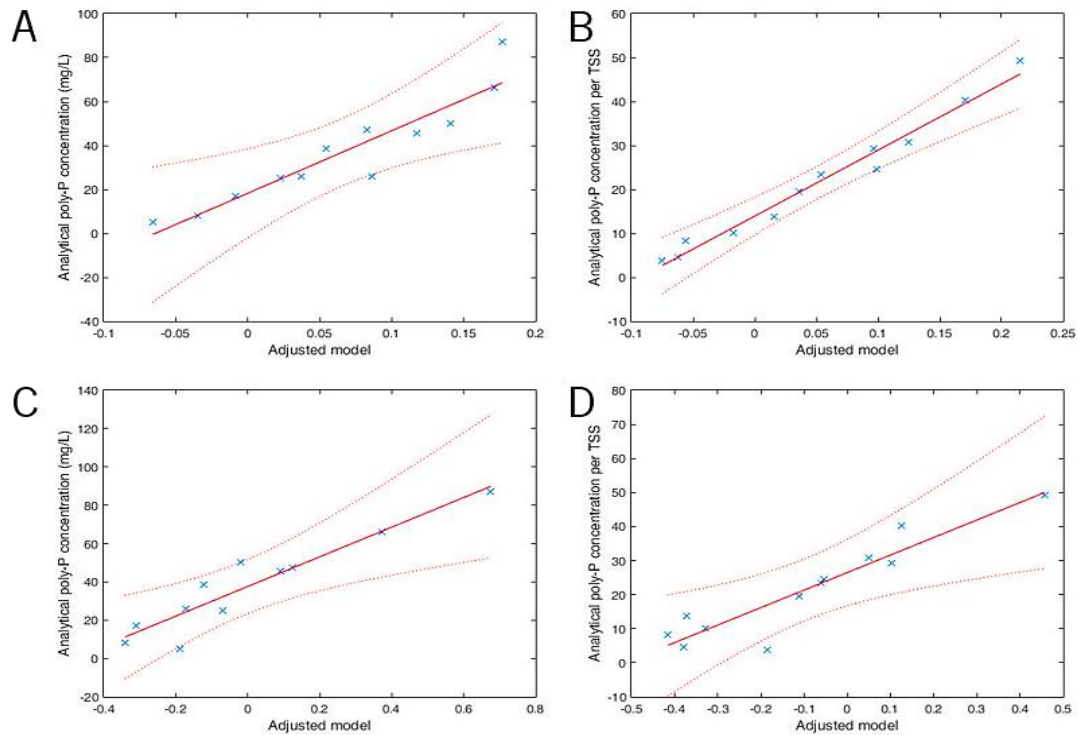
4. Results and Discussion



**Figure 31** Linear regression models of LAB treated images using poly-P area per biomass area as predictor variables and analytical intracellular poly-P concentration per day (left side) or per TSS (right side) as response variables. The models were created using different data based on the dye used: models **A** and **B** – MB; models **C** and **D** – TB. Data is represented as blue stars, the fitted regression line is represented in red and the model's 95% confidence bounds are represented as light red dotted lines. The models' regression coefficients are the following: **A** = 0.920, **B** = 0.919, **C** = 0.690 and **D** = 0.654. The models' p-values are the following: **A** = 0.127, **B** = 0.130, **C** = 0.630 and **D** = 0.692.

4.4.3 Dilution factor

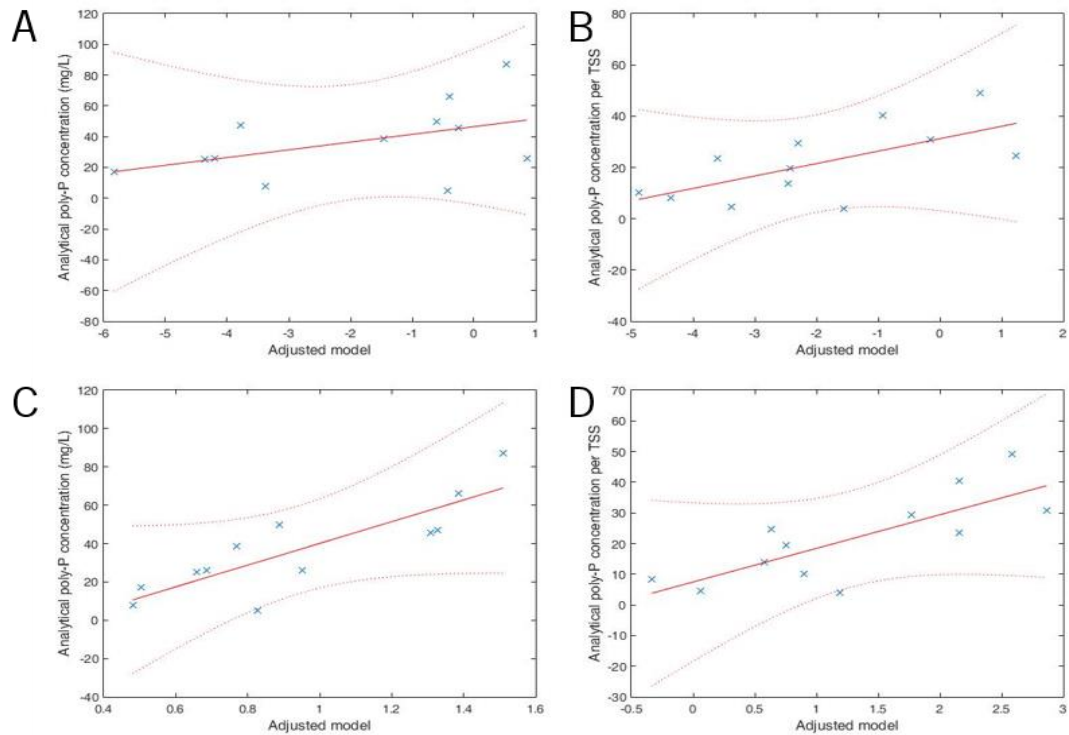
RGB regression models based on the dilution factor samples presented really good regression coefficients on both diluted and non-diluted results with values of 0.869 and 0.978 for diluted samples (**Figure 32A-B**, respectively) and 0.888 and 0.869 for non-diluted ones (**Figure 32C-D**, respectively), depending on the response variable. While the difference was small, diluted samples appeared as more reliable, especially when using int. poly-P per TSS as response variable, which could be seen in **Figure 32**.



**Figure 32** Linear regression models of RGB treated images using poly-P area per biomass area as predictor variables and analytical intracellular poly-P concentration per day (left side) or per TSS (right side) as response variables. The models were created using different data based on the dilution factor: models **A** and **B** – diluted samples; models **C** and **D** – non-diluted samples. Data is represented as blue stars, the fitted regression line is represented in red and the model's 95% confidence bounds are represented as light red dotted lines. The models' regression coefficients are the following: **A** = 0.869, **B** = 0.978, **C** = 0.888 and **D** = 0.869. The models' p-values are the following: **A** = 0.246, **B** = 0.021, **C** = 0.199 and **D** = 0.245.

HSV regressions models based on the dilution factor of samples had high levels of error between model and data. Diluted models using this color space had especially low regression coefficients, with 0.218 and 0.429 depending on the response variable used (**Figure 33A-B**, respectively). A better correlation could be found in the non-diluted data, however even then the regression coefficients of the models were acceptable at best, with values of 0.706 and 0.611, depending on the response variable used (**Figure 33C-D**, respectively). The wide confidence intervals and poor correlation between data and HSV models were evidences to these low regressions.

## 4. Results and Discussion

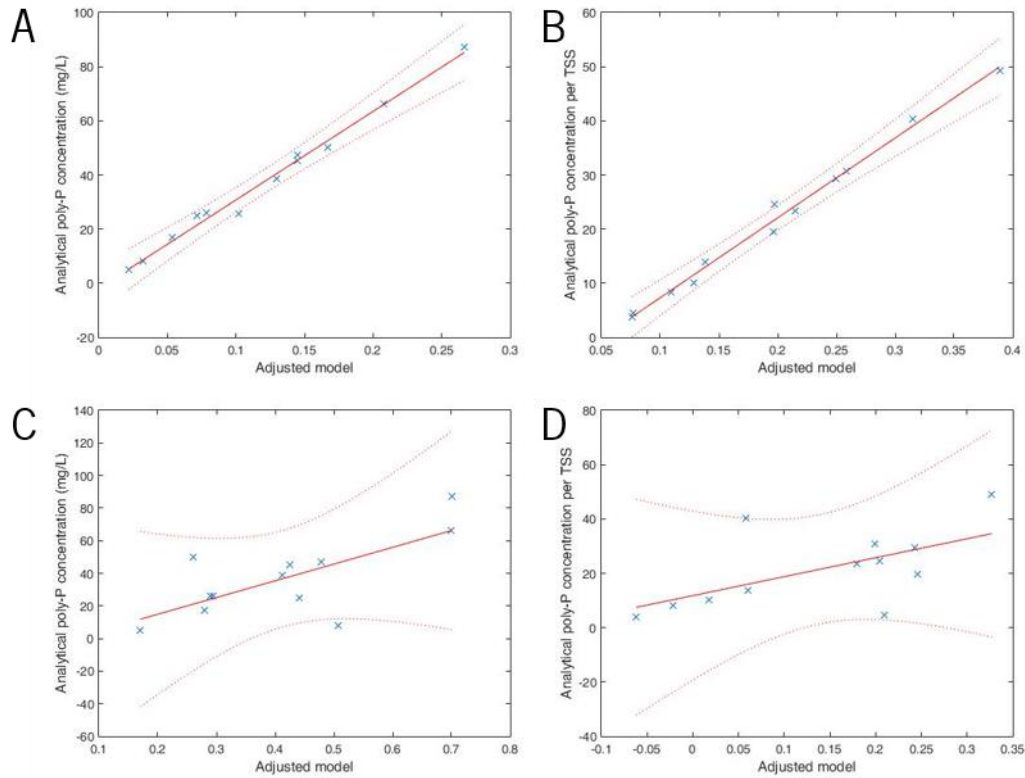


**Figure 33** Linear regression models of HSV treated images using poly-P area per biomass area as predictor variables and analytical intracellular poly-P concentration per day (left side) or per TSS (right side) as response variables. The models were created using different data based on the dilution factor: models **A** and **B** – diluted samples; models **C** and **D** – non-diluted samples. Data is represented as blue stars, the fitted regression line is represented in red and the model's 95% confidence bounds are represented as light red dotted lines. The models' regression coefficients are the following: **A** = 0.218, **B** = 0.429, **C** = 0.706 and **D** = 0.611. The models' p-values are the following: **A** = 0.995, **B** = 0.932, **C** = 0.599 and **D** = 0.757.

The LAB regression models based on dilution factor showed a stark difference between diluted and non-diluted models with smaller confidence bounds in diluted samples. The regression coefficients of these models reflected the difference found in their plots, with diluted models having incredibly high regression coefficients of 0.989 and 0.992, depending on the response variable (**Figure 34A-B**, respectively), and non-diluted samples having meager regression coefficient values of 0.510 and 0.359 (**Figure 34C-D**, respectively).

Considering all these models, diluted data seemed to be a better fit for describing analytical poly-P concentration and analytical poly-P concentration per TSS than non-diluted data, which was expected since the purpose of the dilution was to diminish the amount of biomass and poly-P superimposition for the QIA script to properly recognize poly-P regions, with less chance of hidden poly-P granules. Once again, the HSV color model presented poor results that go against the other color spaces. Also, the RGB color space seemed more consistent, no matter the predictor variables used, while LAB showed good coefficients, but was somewhat inconsistent. From these results, RGB seems to be the most trustworthy

color space. Unfortunately, these linear regression models couldn't be validated, due to an insufficiency of data. While some of these models appeared to have acceptable regression coefficients, which was already a good head start, the lack of validation presented a problem in that the method did not truly show if it properly predicted intracellular poly-P concentration through QIA information.



**Figure 34** Linear regression models of LAB treated images using poly-P area per biomass area as predictor variables and analytical intracellular poly-P concentration per day (left side) or per TSS (right side) as response variables. The models were created using different data based on the dilution factor: models **A** and **B** – diluted samples; models **C** and **D** – non-diluted samples. Data is represented as blue stars, the fitted regression line is represented in red and the model's 95% confidence bounds are represented as light red dotted lines. The models' regression coefficients are the following: **A** = 0.989, **B** = 0.992, **C** = 0.510 and **D** = 0.359. The models' p-values are the following: **A** = 0.007, **B** = 0.005, **C** = 0.872 and **D** = 0.965.





## 5. Conclusion and future work

The use of the `Poly_P_analysis_all.m` script showed great promise in the proper identification of both poly-P and biomass regions through color segmentation. More work could be put into this script to make it even more reliable, perhaps with shape identification of poly-P granules in tandem with color segmentation, or different distance calculation methods, instead of solely Euclidean distance. Not only the script but staining and image acquisition could also be improved. For instance, other methods could be performed to completely prevent biomass and poly-P superimposition, or more complex staining solutions could be used instead of pure dyes such as MB and TB, to provide a higher visual difference between poly-P and biomass regions and facilitate the script's work.

The data obtained from this QIA method showed some positive results in the quantification of intracellular poly-P, with some linear regression models having high regression coefficients (higher than 0.900). There were some visible differences between the three color spaces used, when it came to correlate data to analytical poly-P. RGB data presented good correlation in all models when comparing to the other color spaces, particularly in MB stained images and diluted images. LAB data had worse correlation in most models than RGB. However, MB stained images and diluted images in LAB space presented the highest correlation coefficients among all models. In contrast, HSV models presented unsatisfactory correlation coefficients in all its models. The use of both MB and TB data as variables to test the better image settings (gamma and contrast) proved somewhat unfruitful, with low regression coefficients being found in all models, except RGB model G\_2\_C\_2, meaning that this setting could be ideal for this kind of image treatment. There were also differences between the p-values of all models, with only a few having lower p-values than 0.05, indicating that some models were not fit for further testing. Unfortunately, the lack of validation due to insufficient data prevented proper corroboration of these models.

In conclusion, MB colored samples seemed to be more effectively analyzed by RGB and LAB color spaces, with dilution of the samples providing the same effect, possibly by preventing superimposition of poly-P granules and biomass and more accurately identifying poly-P regions. Also, contrast and gamma values showed some influence in the results and are therefore a factor to take into consideration in this sort of image analysis.



## 6. Literature

- [1] U. Wolfmeier *et al.*, "Water, 1. Properties, Analysis and Hydrological Cycle," *Ullmann's Encycl. Ind. Chem.*, pp. 111–172, 2000.
- [2] P. H. Gleick, *Water in crisis : a guide to the world's fresh water resources*, 1st ed. New York: Oxford University Press, 1993.
- [3] E. . Koncagül, M. Tran, R. Connor, S. Uhlenbrook, A. Renata, and C. Ortigara, "Wastewater the untapped resource. Facts and figures. The United Nations World Water Development Report 2017," p. 12, 2017.
- [4] D. Mulkerrins, A. D. W. Dobson, and E. Colleran, "Parameters affecting biological phosphate removal from wastewaters," *Environ. Int.*, vol. 30, no. 2, pp. 249–259, 2004.
- [5] A. C. D. B. Arts, "Review on Waste Water Treatment Technologies," no. July, 2012.
- [6] S. S. Adav, D. J. Lee, K. Y. Show, and J. H. Tay, "Aerobic granular sludge: Recent advances," *Biotechnol. Adv.*, vol. 26, no. 5, pp. 411–423, 2008.
- [7] R. J. Seviour, T. Mino, and M. Onuki, "The microbiology of biological phosphorus removal in activated sludge systems," *FEMS Microbiol. Rev.*, vol. 27, no. 1, pp. 99–127, 2003.
- [8] D. P. Mesquita *et al.*, "Monitoring intracellular polyphosphate accumulation in enhanced biological phosphorus removal systems by quantitative image analysis," *Water Sci. Technol.*, vol. 69, no. 11, pp. 2315–2323, 2014.
- [9] C. Tarayre *et al.*, "Characterisation of Phosphate Accumulating Organisms and Techniques for Polyphosphate Detection: A Review," *Sensors*, vol. 16, no. 6, p. 797, 2016.
- [10] N. Majed, Y. Li, and A. Z. Gu, "Advances in techniques for phosphorus analysis in biological sources," *Curr. Opin. Biotechnol.*, vol. 23, no. 6, pp. 852–859, 2012.
- [11] P. F. Cooper, "Historical aspects of wastewater treatment," *Decent. Sanit. reuse concepts, Syst. Implement.*, pp. 11–38, 2007.
- [12] P. Wolfe, "History of wastewater," in *World of water 2000: past, present and future*, Tulsa, Oklahoma, USA: Penn Well Corporation, 1999.
- [13] S. Halliday, *The Great Stink of London*. Stroud, Gloucestershire: Sutton publishing, 1999.
- [14] C. Binnie, "Present London," in *World of water 2000: past, present and future*, Tulsa, Oklahoma, USA: Penn Well Corporation, 1999.

- [15] T. Wardle, *On Sewage Treatment and Disposal: For Cities, Towns, Villages, Private Dwellings, and Public Institutions*. Heywood, 1893.
- [16] H. H. Stanbridge, *History of Sewage Treatment in Britain*. Institute of Water Pollution Control, 1976.
- [17] M. V Melosi, *The Sanitary City: Urban Infrastructure in America from Colonial Times to the Present*. Johns Hopkins University Press, 2000.
- [18] E. H. Nicoll, *Small water pollution control works: design and practice*. E. Horwood, 1988.
- [19] M. Samer, "Biological and Chemical Wastewater Treatment Processes," *Wastewater Treat. Eng.*, pp. 1–50, 2015.
- [20] E. Arden and W. T. Lockett, "Experiments on the oxidation of sewage without the aid of filters," *J. Soc. Chem. Ind.*, vol. 33, no. 10, pp. 523–539, 1914.
- [21] Institute of Water Pollution Control, *Unit processes: Activated Sludge*. Maidstone, Kent: Institute of Water Pollution Control, 1987.
- [22] D. Orhon, "Evolution of the activated sludge process: the first 50 years," *J. Chem. Technol. Biotechnol.*, vol. 90, no. 4, pp. 608–640, 2015.
- [23] F. W. Mohlamn, "Twenty-five years of activated sludge," in *Modern Sewage Disposal*, 1938.
- [24] J. L. Barnard, "Biological nutrient removal without the addition of chemicals," *Water Res.*, vol. 9, no. 5–6, pp. 485–490, 1975.
- [25] A. Oehmen *et al.*, "Advances in enhanced biological phosphorus removal: From micro to macro scale," *Water Res.*, vol. 41, no. 11, pp. 2271–2300, 2007.
- [26] J. W. McGrath and J. P. Quinn, "Microbial phosphate removal and polyphosphate production from wastewaters," *Adv. Appl. Microbiol.*, vol. 52, pp. 75–100, 2003.
- [27] R. Hirota, A. Kuroda, J. Kato, and H. Ohtake, "Bacterial phosphate metabolism and its application to phosphorus recovery and industrial bioprocesses," *J. Biosci. Bioeng.*, vol. 109, no. 5, pp. 423–432, 2010.
- [28] B. A. O. Lin-lin *et al.*, "Phosphorus accumulation by bacteria isolated from a continuous- ow two-sludge system," vol. 19, pp. 391–395, 2007.
- [29] J. P. Kern-Jespersen and M. Henze, "Biological phosphorus uptake under anoxic and aerobic conditions," *Water Res.*, vol. 27, no. 4, pp. 617–624, 1993.
- [30] Y. Z. Peng, Y. Y. Wang, M. Ozaki, and M. L. Pan, "Denitrifying Phosphorus Removal in a Continuously-

- Flow A2N Two-Sludge Process," *J. Environ. Sci. Heal. - Part A Toxic/Hazardous Subst. Environ. Eng.*, vol. 39, no. 3, pp. 703–715, 2004.
- [31] N. N. Rao, M. R. Gómez-García, and A. Kornberg, "Inorganic Polyphosphate: Essential for Growth and Survival," *Annu. Rev. Biochem.*, vol. 78, no. 1, pp. 605–647, 2009.
- [32] K. C. Lindrea, E. M. Seviour, R. J. Seviour, L. L. Blackall, and J. A. Soddell, "Practical methods for the examination and characterization of activated sludge BT - The Microbiology of Activated Sludge," R. J. Seviour and L. L. Blackall, Eds. Dordrecht: Springer Netherlands, 1999, pp. 257–300.
- [33] H. C. Schröder and W. E. G. Müller, *Inorganic polyphosphates: biochemistry, biology, biotechnology*. 1999.
- [34] O. Henriët, C. Meunier, P. Henry, and J. Mahillon, "Improving phosphorus removal in aerobic granular sludge processes through selective microbial management," *Bioresour. Technol.*, vol. 211, pp. 298–306, 2016.
- [35] V. R. Hill *et al.*, "Development of a rapid method for simultaneous recovery of diverse microbes in drinking water by ultrafiltration with sodium polyphosphate and surfactants," *Appl. Environ. Microbiol.*, vol. 71, no. 11, pp. 6878–6884, 2005.
- [36] A. M. Freitas and M. M. Sharma, "Detachment of particles from surfaces: An AFM study," *J. Colloid Interface Sci.*, vol. 233, no. 1, pp. 73–82, 2001.
- [37] J. E. Clark, H. Beegen, and H. G. Wood, "Isolation of intact chains of polyphosphate from 'Propionibacterium shermanii' grown on glucose or lactate.," *J. Bacteriol.*, vol. 168, no. 3, pp. 1212–1219, Dec. 1986.
- [38] J. Ahlgren, H. De Brabandere, K. Reitzel, E. Rydin, A. Gogoll, and M. Waldeback, "Sediment Phosphorus Extractants for Phosphorus-31 Nuclear Magnetic Resonance Analysis," *J. Environ. Qual.*, vol. 36, no. 3, p. 892, 2007.
- [39] M. Hupfer, S. Glöss, P. Schmieder, and H.-P. Grossart, "Methods for Detection and Quantification of Polyphosphate and Polyphosphate Accumulating Microorganisms in Aquatic Sediments," *Int. Rev. Hydrobiol.*, vol. 93, no. 1, pp. 1–30, Feb. 2008.
- [40] K. Lindrea, G. Lockwood, and M. Majone, "The distribution and movement of polyphosphate and associated cations in sludges from ndebpr plants in different configurations at pilot scale," *Water Sci. Technol.*, vol. 37, no. 4–5, 1998.
- [41] C. Van Moorlehem, L. Six, F. Degryse, E. Smolders, and R. Merckx, "Effect of Organic P Forms and P Present in Inorganic Colloids on the Determination of Dissolved P in Environmental Samples by the

- Diffusive Gradient in Thin Films Technique, Ion Chromatography, and Colorimetry," *Anal. Chem.*, vol. 83, no. 13, pp. 5317–5323, Jul. 2011.
- [42] A. Z. Gu, L. Liu, J. B. Neethling, H. D. Stensel, and S. Murthy, "Treatability and fate of various phosphorus fractions in different wastewater treatment processes," *Water Sci. Technol.*, vol. 63, no. 4, p. 804, Feb. 2011.
- [43] Y. Baba, N. Yoza, and S. Ohashi, "Simultaneous determination of phosphate and phosphonate by flow injection analysis and high-performance liquid chromatography with a series detection system," *J. Chromatogr. A*, vol. 318, pp. 319–324, Jan. 1985.
- [44] Choi, Hercules, and Houalla, "Characterization of polyphosphates by electrospray mass spectrometry," *Anal. Chem.*, vol. 72, no. 20, pp. 5087–5091, Oct. 2000.
- [45] K. Saito, R. Ohtomo, Y. Kuga-Uetake, T. Aono, and M. Saito, "Direct Labeling of Polyphosphate at the Ultrastructural Level in *Saccharomyces cerevisiae* by Using the Affinity of the Polyphosphate Binding Domain of *Escherichia coli* Exopolyphosphatase," *Applied and Environmental Microbiology*, vol. 71, no. 10, pp. 5692–5701, Oct-2005.
- [46] R. Ohtomo, Y. Sekiguchi, T. Kojima, and M. Saito, "Different chain length specificity among three polyphosphate quantification methods," *Anal. Biochem.*, vol. 383, no. 2, pp. 210–216, Dec. 2008.
- [47] D. Ault-Riche, C. D. Fraley, C. M. Tzeng, and A. Kornberg, "Novel assay reveals multiple pathways regulating stress-induced accumulations of inorganic polyphosphate in *Escherichia coli*," *J. Bacteriol.*, vol. 180, no. 7, pp. 1841–1847, Apr. 1998.
- [48] M. Heldal, "Measurement of elemental content and dry weight of single cells: X-ray microanalysis," in *Handbook of methods in aquatic microbial ecology*, vol. 39, no. 6, Boca Raton, Florida: Lewis Publishers, 1994, pp. 387–394.
- [49] J. Goldberg, H. Gonzalez, T. E. Jensen, and W. A. Corpe, "Quantitative analysis of the elemental composition and the mass of bacterial polyphosphate bodies using STEM EDX," *Microbios*, vol. 106, pp. 177–188, 2001.
- [50] J. Diaz *et al.*, "Marine Polyphosphate: A Key Player in Geologic Phosphorus Sequestration," *Science*, vol. 320, pp. 652–655, 2008.
- [51] J. Diaz *et al.*, "Characterization of phosphorus, calcium, iron, and other elements in organisms at sub-micron resolution using X-ray fluorescence spectromicroscopy," *Limnol. Oceanogr. Methods*, vol. 7, no. 1, pp. 42–51, Jan. 2009.
- [52] S. Günther, I. Röske, and T. Bley, "Population Structure and Dynamics of Polyphosphate Accumulating

- Organisms in a Communal Wastewater Treatment Plant,” Technischen Universität Dresden, Dresden, Germany, 2011.
- [53] J. M. Diaz and E. D. Ingall, “Fluorometric quantification of natural inorganic polyphosphate,” *Environ. Sci. Technol.*, vol. 44, no. 12, pp. 4665–4671, 2010.
- [54] M. Streichan, J. R. Golecki, and G. SchÄ¶n, “Polyphosphate-accumulating bacteria from sewage plants with different processes for biological phosphorus removal,” *FEMS Microbiol. Lett.*, vol. 73, no. 2, pp. 113–124, Feb. 1990.
- [55] R. Aschar-Sobbi *et al.*, “High Sensitivity, Quantitative Measurements of Polyphosphate Using a New DAPI-Based Approach,” *J. Fluoresc.*, vol. 18, no. 5, pp. 859–866, Sep. 2008.
- [56] N. Majed, T. Chernenko, M. Diem, and A. Z. Gu, “Identification of Functionally Relevant Populations in Enhanced Biological Phosphorus Removal Processes Based On Intracellular Polymers Profiles and Insights into the Metabolic Diversity and Heterogeneity,” *Environ. Sci. Technol.*, vol. 46, no. 9, pp. 5010–5017, May 2012.
- [57] J. W. Bartholomew, “Stains for Microorganisms in Smears,” in *Staining Procedures*, 4th editio., G. Clark, Ed. Baltimore: Williams and Wilkins, 1981, pp. 375–440.
- [58] L. S. Serafim, P. C. Lemos, C. Levantesi, V. Tandoi, H. Santos, and M. A. M. Reis, “Methods for detection and visualization of intracellular polymers stored by polyphosphate-accumulating microorganisms,” *J. Microbiol. Methods*, vol. 51, no. 1, pp. 1–18, 2002.
- [59] R. Murray, R. Doetsch, and C. Robinow, “Determinative and cytological light microscopy,” in *Methods for general and molecular bacteriology*, P. Gerhardt, R. Murray, W. Wood, and N. Krieg, Eds. Washington D.C.: American Society of Microbiology, 1994, pp. 21–41.
- [60] E. Meijering and G. Cappellen, “Quantitative biological image analysis,” *Imaging Cell. Mol. Biol. Funct.*, vol. 2, pp. 45–70, 2007.
- [61] R. C. . Gonzalez and R. E. Woods, “Digital image processing,” *Nueva Jersey*. p. 976, 2008.
- [62] R. Kuehni, “Color spaces,” *Scholarpedia*, vol. 5, no. 3, p. 9606, 2010.
- [63] G. Hasting and A. Rubin, “Colour spaces-a review of historic and modern colour models\*,” *African Vis. Eye Heal.*, vol. 71, no. 3, pp. 133–143, 2012.
- [64] P. C. Coding, “Chapter 2 Basic Color Theory.”
- [65] C. Poynton, “Frequently asked questions about gamma,” *Rapp. Tech. janvier*, pp. 1–12, 1998.
- [66] A. Braga, D. P. Mesquita, A. L. Amaral, E. C. Ferreira, and I. Belo, “Quantitative image analysis as a tool

- for *Yarrowia lipolytica* dimorphic growth evaluation in different culture media,” *J. Biotechnol.*, vol. 217, pp. 22–30, 2016.
- [67] D. P. Mesquita, A. L. Amaral, C. Leal, A. Oehmen, M. A. M. Reis, and E. C. Ferreira, “Polyhydroxyalkanoate granules quantification in mixed microbial cultures using image analysis: Sudan Black B versus Nile Blue A staining,” *Anal. Chim. Acta*, vol. 865, no. 1, pp. 8–15, 2015.
- [68] D. P. Mesquita, O. Dias, A. M. A. Dias, A. L. Amaral, and E. C. Ferreira, “Correlation between sludge settling ability and image analysis information using partial least squares,” *Anal. Chim. Acta*, vol. 642, no. 1–2, pp. 94–101, 2009.
- [69] R. Jenne, E. N. Banadda, I. Y. Smets, A. Bamelis, L. Verdickt, and J. F. Van Impe, “Activated sludge image analysis system: monitoring settleability and effluent clarity,” *Water Sci. Technol.*, vol. 52, no. 10–11, pp. 193–199, 2005.
- [70] M. da Motta, M. N. Pons, and N. Roche, “Automated monitoring of activated sludge in a pilot plant using image analysis,” *Water Sci. Technol.*, vol. 43, no. 7, pp. 91–96, 2001.
- [71] M. da Motta, M.-N. Pons, N. Roche, and H. Vivier, “Characterisation of activated sludge by automated image analysis,” *Biochem. Eng. J.*, vol. 9, no. 3, pp. 165–173, 2001.
- [72] A. L. Amaral and E. C. Ferreira, “Activated sludge monitoring of a wastewater treatment plant using image analysis and partial least squares regression,” *Anal. Chim. Acta*, vol. 544, no. 1, pp. 246–253, 2005.
- [73] D. P. Mesquita, A. L. Amaral, and E. C. Ferreira, “Activated sludge characterization through microscopy: A review on quantitative image analysis and chemometric techniques,” *Anal. Chim. Acta*, vol. 802, pp. 14–28, 2013.
- [74] A. Oehmen, A. M. Saunders, M. T. Vives, Z. Yuan, and J. Keller, “Competition between polyphosphate and glycogen accumulating organisms in enhanced biological phosphorus removal systems with acetate and propionate as carbon sources,” *J. Biotechnol.*, vol. 123, no. 1, pp. 22–32, 2006.
- [75] G. J. F. Smolders, J. Van Der Meij, M. C. M. Van Loosdrecht, and J. J. Heijnen, “Model of the Anaerobic Metabolism of the Biological Phosphorus Removal Process: Stoichiometry and pH Influence,” vol. 43, pp. 461–470, 1994.
- [76] L. S. Clesceri, A. E. Greenbaerg, and A. D. Eaton, “Standard Methods for Examination of Water and Wastewater (Standard Methods for the Examination of Water and Wastewater),” *Am. Public Heal. Assoc. Washington, DC, USA*, vol. 552, pp. 5–16, 1998.
- [77] D. A. Freedman, *Statistical models: Theory and practice*, 2nd ed. Cambridge University Press, 2009.



- [78] J. Aldrich, "Doing least squares: Perspectives from Gauss and Yule," *Int. Stat. Rev.*, vol. 66, no. 1, pp. 61–81, 1998.
- [79] M. C. Lovell, "A Simple Proof of the FWL Theorem," *J. Econ. Educ.*, vol. 39, no. 1, pp. 88–91, 2008.