Bioplastic degradation

Abstract

Bioplastic are polymers produced by micro-organisms. An example of a bioplastic is PHBV. PHBV is short for Poly(3-hydroxybutyrate-co-3-hydroxyvalerate). PHBV is a thermoplastic polymer, and most important, is renewable.

This project is about processing data form a flow cytometer and develop with machine learning an analysis pipeline that can distinguish PHBV particles from bacteria. Also the degradation rate is calculated over time to see which bacteria is better in processing plastic debris.

So far, we have analysed and studied the difficult dataset from the flow cytometer, and translated this into plots and graphs. At the end, we have now a workable dataset to be used in machine learning.

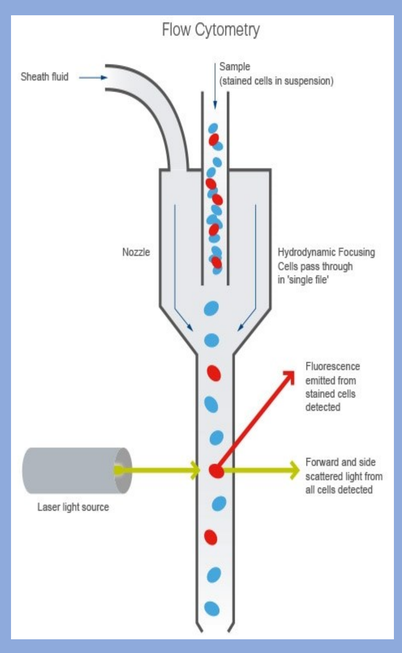
Conclusion

Many challenges arised when we first got the dataset. It was messy and it was not clear what the organisation of the files was. By doing many analysis of the data and talk with the author of the dataset, our knowledge of the dataset grew and it is now

Material and method

For analyzing the dataset we have used primary the programing language R. To be specific: R version 4.1.2 (2021-11-01) -- "Bird Hippie" Platform: x86_64-w64-mingw32/x64 (64-bit).

The data was retrieved form a flow cytometer, which works with lasers to scatter light of a practical. Ad catches those scatter values.



The data it self contains of 4 folders with different measurement techniques and different samples. Each sample had also different rates of flow cytometer cycles, which made it so difficult. Each file 14 columns with information, and rows with data of one particle that was passed trough the flow cytometer.

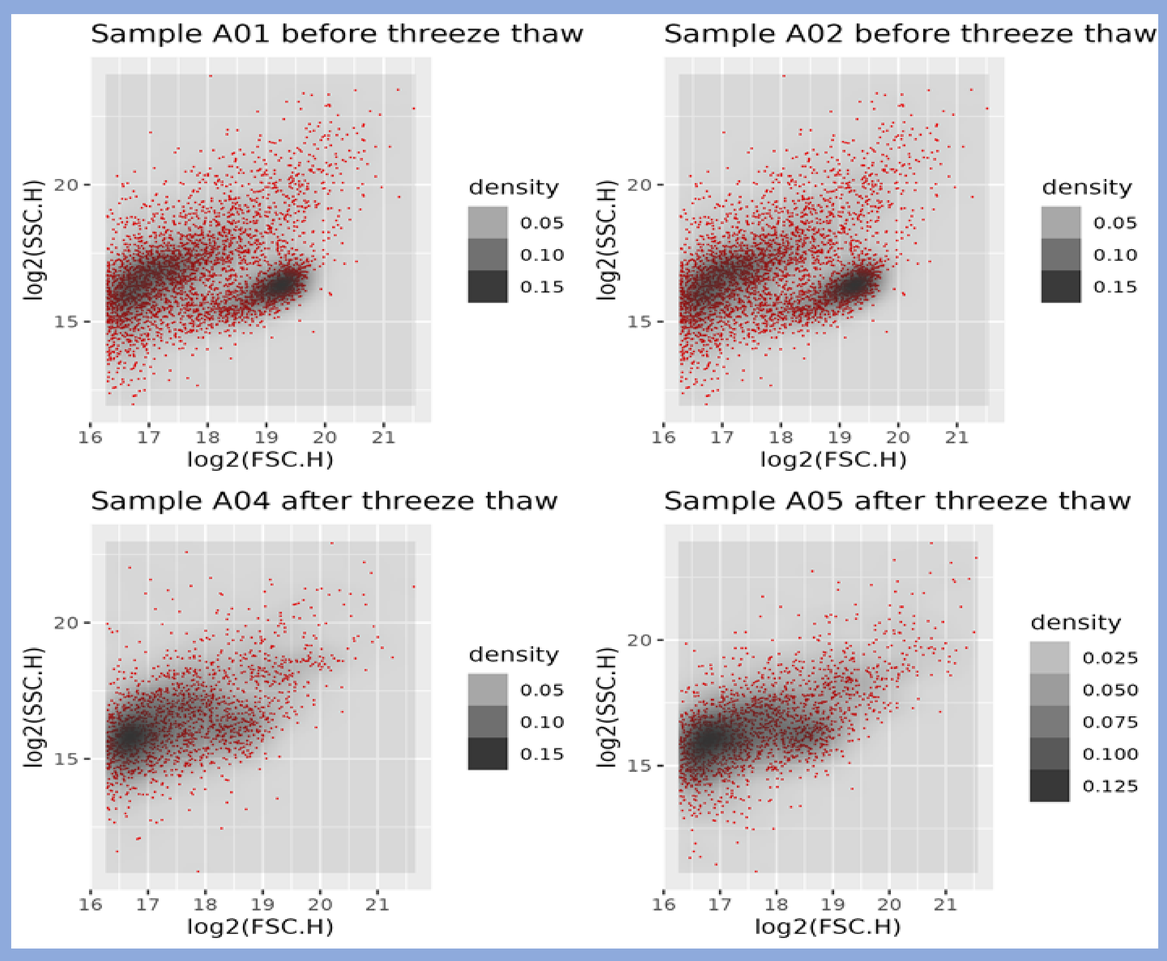
The most important column variables where the Forward scatter (FSC), Sideward scatter (SSC) and the width. These variables tell something about the area and surface complexity of the particle. Below a example of how the data looks in R.

FSC.A <int>	SSC.A <int>	FL1.A <int>	FL2.A <int>	FL3.A <int>	FL4.A <int>
183714	83123	92	109	94	249
135029	50422	125	219	284	467
1152415	963584	136	51	382	354
471739	125265	4	27	87	238
432272	103531	127	113	259	463
158793	59836	185	99	363	290
107982	23163	136	132	278	232
450685	78495	54	62	117	432
111624	28196	88	54	0	331
126381	58323	187	116	319	103

Results

Freeze – thaw:

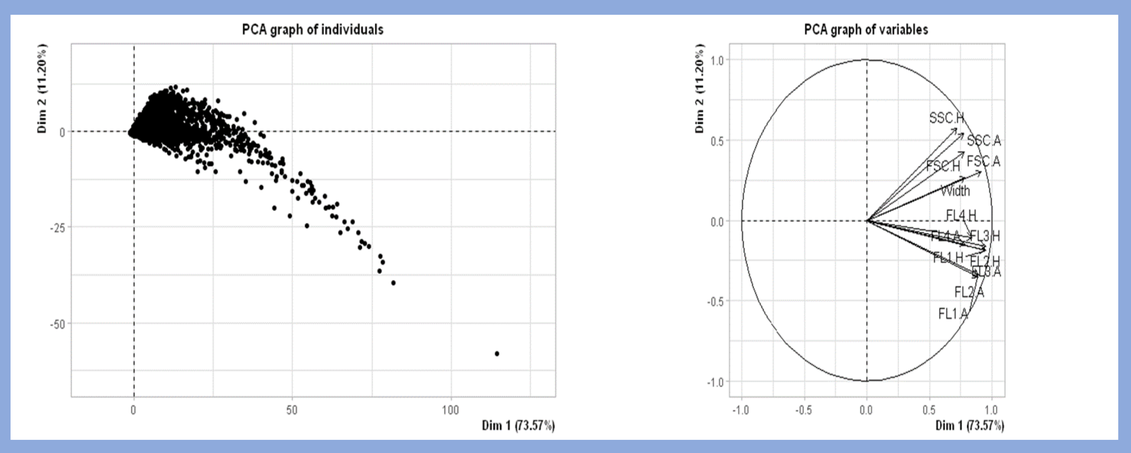
A method used in the flow cytometer data us Freeze thaw. Where bacteria of different samples where killed.



Here you can clearly see that before Freeze thaw , there were 2 clusters. And after freeze thaw only one cluster. This was very useful to identify the bacteria and plastics.

PCA:

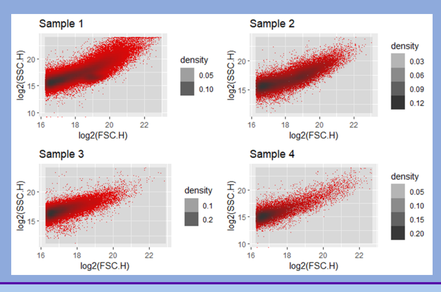
Pca is a useful statical analysis technique to find new correlation between variables. PCA was done on a part of the data to sneak peek into it, to see if there are any correlations.

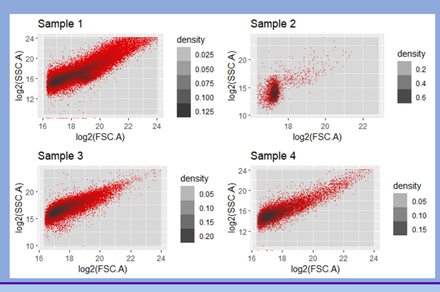


When comparing PC1 against PC2 in this example, there is not a defined cluster to see, but the biplot let us see that certain variables are strong correlated to each other. Which makes this helpful.

Density

The density was also plotted of determined samples. But although there were differences, these differences where not the same across the data.





Hanzehogeschool Groningen
University of Applied Sciences

Pascal Visser
p.visser@st.hanze.nl
BFV3

Rob Meulenkamp
R.meulenkamp@st.hanze.nl
BFV3

Rienk Heins
R.heins@st.hanze.nl
BFV3