The influence of processing methods and environmental factors on the composition of sargassum biomass

Introduction

This report will explore the elemental and biochemical composition of sargassum biomass, a type of seaweed found washed up on the shores of Jamaica. In April 2021 there was a volcanic eruption in St Vincent, leading to the sargassum being exposed to volcanic ash for approximately 50 days (Machado et al., 2024). We will perform agglomerative hierarchical clustering and principal component analysis on the dataset to investigate the influence of processing methods and potential seasonal variation. We will also explore the potential effect of ash exposure on the sargassum biomass.

Data and Methods

The dataset contains the biochemical and elemental composition of sargassum, sampled three times a month from May to October 2021 using different processing methods. The samples were either freeze-dried or shade-dried. The dataset was taken from the initial report conducted (Machado et al., 2024), with the values entered as “bd” being changed to 0. As some of the variables in the dataset are much larger than others, the data was first scaled to unit variance to prevent these from dominating the analysis.

Principal Components Analysis (PCA) was performed using R Studio. PCA is an unsupervised method for multivariate analysis that is used for exploratory data analysis. This method involves defining new variables called principal components that are linear combinations of the original variables. These principal components are uncorrelated, and ordered in terms of how much variance they account for. The loadings of the principal components show which variables contribute most to this variance. PCA is an appropriate method to use here because it can show the elements responsible for variation in the data, and help visualise groupings.

Agglomerative hierarchical cluster analysis was also performed using R Studio. Cluster analysis is an unsupervised method used in exploratory data analysis to find inherent groupings in the data. It reveals associations and structures in the dataset, and can suggest hypotheses about relationships. The degree of association between elements in the same cluster is high, meaning they’re similar and have properties in common. To perform agglomerative hierarchical clustering, first a distance matrix must be created using a chosen metric. Here the euclidean metric has been chosen using the complete-linkage method, which calculates the distance between the objects in the cluster that are furthest apart. Then each element of the dataset is taken as an individual cluster, and successively merged into new clusters based on the elements that are closest together. The elements that cluster together quickly represent the natural groupings in the data.

Results and Discussion

The biplot displayed in Figure 1 is the result of performing PCA on the whole biochemical dataset. It can be seen that the freeze-dried and shade-dried samples divide into two

clusters, separated along PC1. The variables most responsible for variance in this principal component are alginate_mono and fucoxanthin. This suggests that processing methods have an influence on the biochemical composition of sargassum samples, especially on the levels of fucoxanthin, which are higher in the freeze-dried samples.

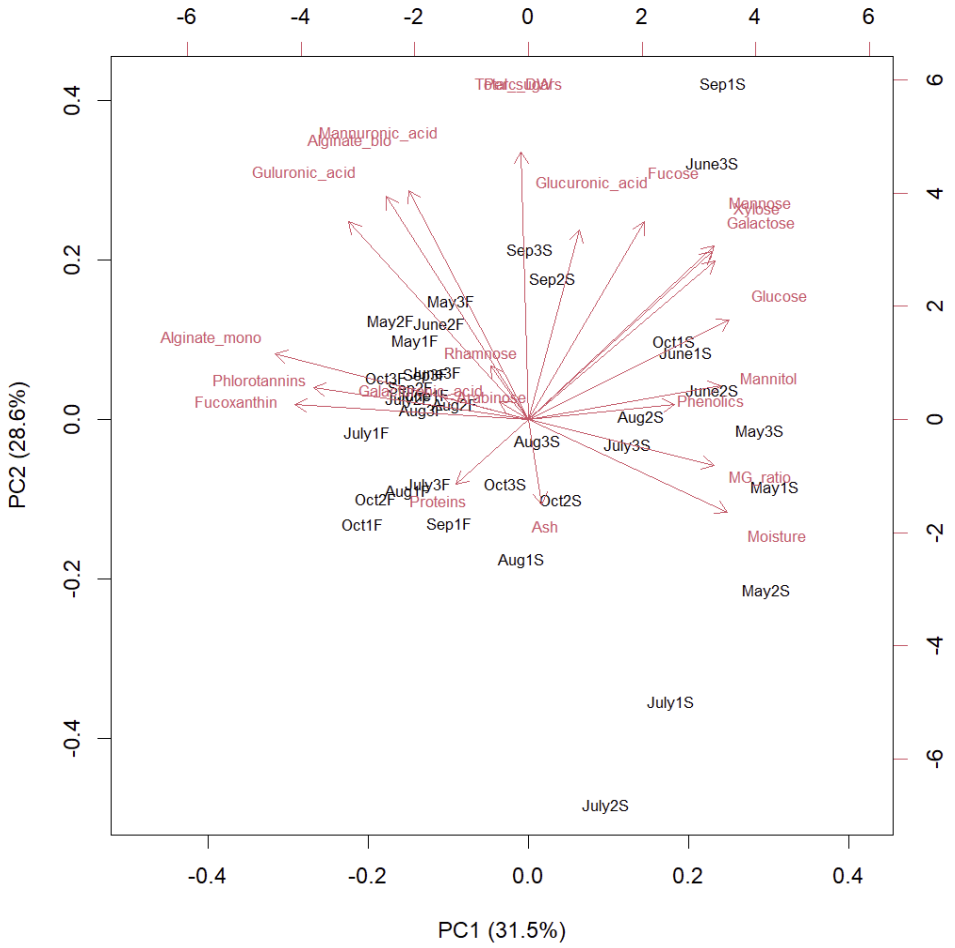

Figure 1: A biplot displaying the results of PCA performed the whole biochemical dataset. An S at the end of the label indicates the sample was shade-dried, an F indicates freeze-dried. Two clusters emerge, separated by processing method.

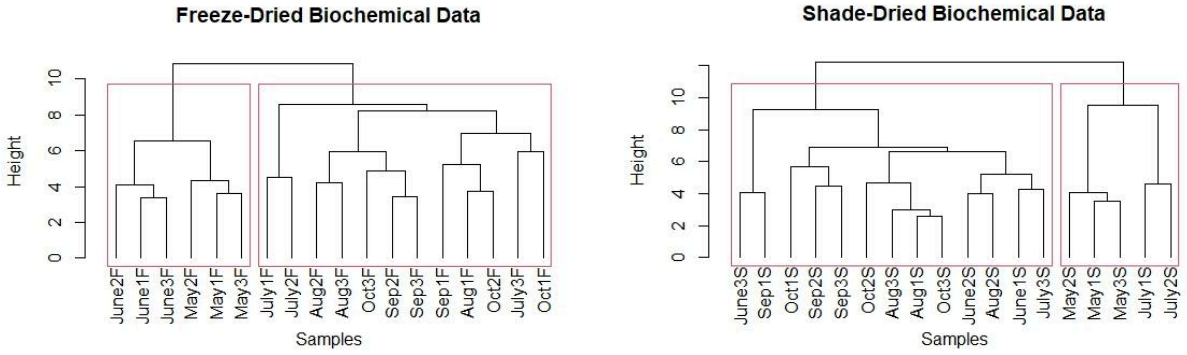

Figure 2: Dendrograms showing the clustering of freeze-dried and shade-dried samples in the biochemical dataset. The red boxes outline the clusters that have emerged.

To further investigate the biochemical composition of the seaweed, the data can be divided by processing method. Performing agglomerative hierarchical cluster analysis on the freeze-dried subset of the biochemical dataset gives the first dendrogram in Figure 2. Two clusters emerge, separating the samples taken before ash exposure (May and June) from the samples taken after possible ash exposure (July-October). This suggests that ash exposure has had an effect on the biochemical composition of the seaweed. In contrast, the dendrogram displaying the results for shade-dried samples shows a cluster between May and some samples in July. This suggests that it's harder to spot seasonal variation in the shade-dried samples, showing the influence that the processing method can have.

The biplot in Figure 4 displays the results of performing PCA on the freeze-dried biochemical subset, which shows that the samples from months before potential ash exposure are separated from the others along PC1. The variables most responsible for the variance in this principal component are the acids in the monosaccharides (Figure 3A), which are higher in the samples before ash exposure. Specifically, these variables are glucuronic acid, guluronic acid, mannuronic acid, the percentage of biomass accounted for by alginate and the percentage of dry weight accounted for by sugars. This suggests that exposure to ash has led to a decrease of alginates in the sargassum.

Performing PCA on the shade-dried samples gave slightly different results. The biplot in Figure 5 shows perhaps some separation along PC2 between the months before and after potential ash exposure, but it's less clear than the results from the freeze-dried samples. The variables responsible for this separation are arabinose, galactose, glucose and xylose, seen in the loadings in Figure 3B. These variables decrease in the months after potentially coming into contact with volcanic ash. Though this is again showing a change in monosaccharides after ash exposure, they are different ones to the freeze-dried results. This further shows the influence of processing methods. This difference could be due to degradation in the shade-dried samples. As shade drying takes longer, the samples could lose chemicals to the environment over time. The results from the freeze-dried samples are clearer, so this may be a better processing method to use.

A)

	PC1	PC2
Moisture	-0.017507288	-0.08536792
Ash	0.006921633	-0.16918000
Proteins	-0.111279694	0.19768795
Phenolics	0.064390795	-0.18830334
Phlorotannins	-0.178862642	0.04631399
Fucoxanthin	-0.137306607	0.18467788
Mannitol	0.226434068	-0.07682211
Fucose	0.008277925	0.35087393
Arabinose	-0.052419463	0.23334088
Rhamnose	-0.173911091	0.20057610
Galactose	-0.073944985	0.40770743
Glucose	0.193620153	0.30331231
Xylose	0.126612432	0.38182209
Mannose	0.200676638	0.33310615
Galacturonic_acid	0.214973177	-0.02032924
Glucuronic_acid	0.321927408	-0.02251139
Guluronic_acid	0.306583554	0.03822347
Mannuronic_acid	0.333405347	-0.03542631
Total_sugars	0.339553544	0.07649245
Perc_DW	0.339537675	0.07651127
Alginate_bio	0.330153716	-0.01631819
Alginate_mono	0.030362413	-0.25733966
MG_ratio	0.230586100	-0.18112232

B)

	PC1	PC2
Moisture	0.18190506	-0.011473783
Ash	0.13900860	0.186549841
Proteins	0.10250624	-0.070759874
Phenolics	-0.01547865	0.211249374
Phlorotannins	-0.10382349	-0.244005504
Fucoxanthin	-0.05444929	-0.145773389
Mannitol	-0.09209474	-0.006071737
Fucose	-0.29383437	0.122169833
Arabinose	0.01169422	0.355788153
Rhamnose	-0.12757572	0.092299447
Galactose	-0.23755531	0.283515966
Glucose	-0.10212785	0.392701874
Xylose	-0.21777750	0.308596531
Mannose	-0.23648117	0.257674388
Galacturonic_acid	0.07783565	0.271215737
Glucuronic_acid	-0.28328759	-0.125149032
Guluronic_acid	-0.30528764	-0.160542163
Mannuronic_acid	-0.31862303	-0.097391648
Total_sugars	-0.33753709	0.014636211
Perc_DW	-0.33755447	0.014650116
Alginate_bio	-0.32030512	-0.119121643
Alginate_mono	-0.15151513	-0.337899880
MG_ratio	0.08480950	0.187298326

Figure 3: A) The loadings for the principal components of the biochemical freeze-dried subset B) The loadings for the principal components of the biochemical shade-dried subset

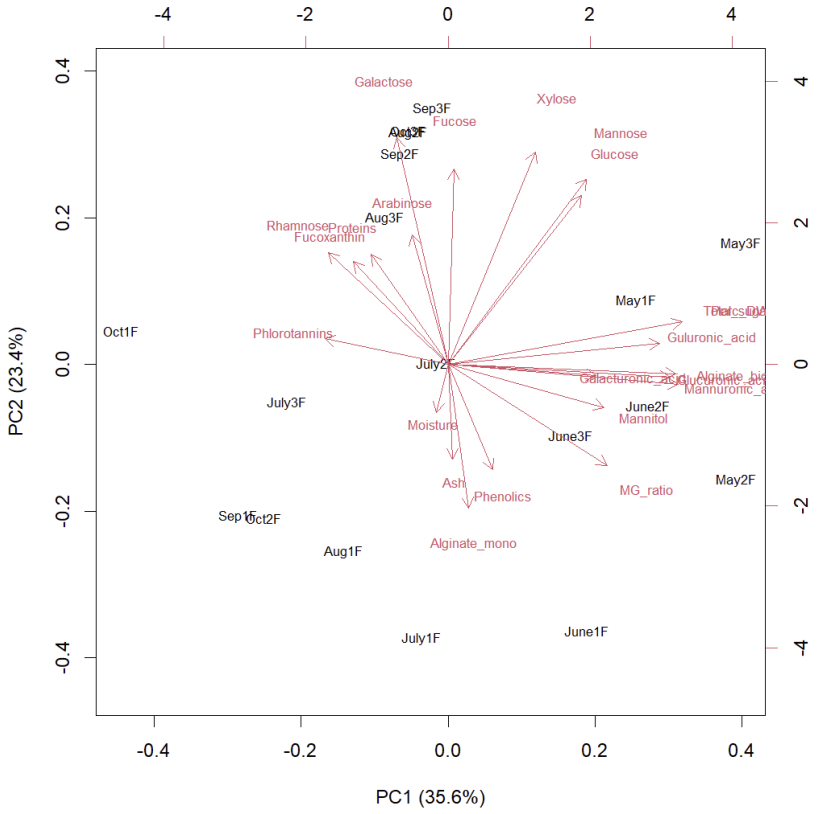

Figure 4: A biplot displaying results of PCA on the freeze-dried samples in the biochemical dataset.

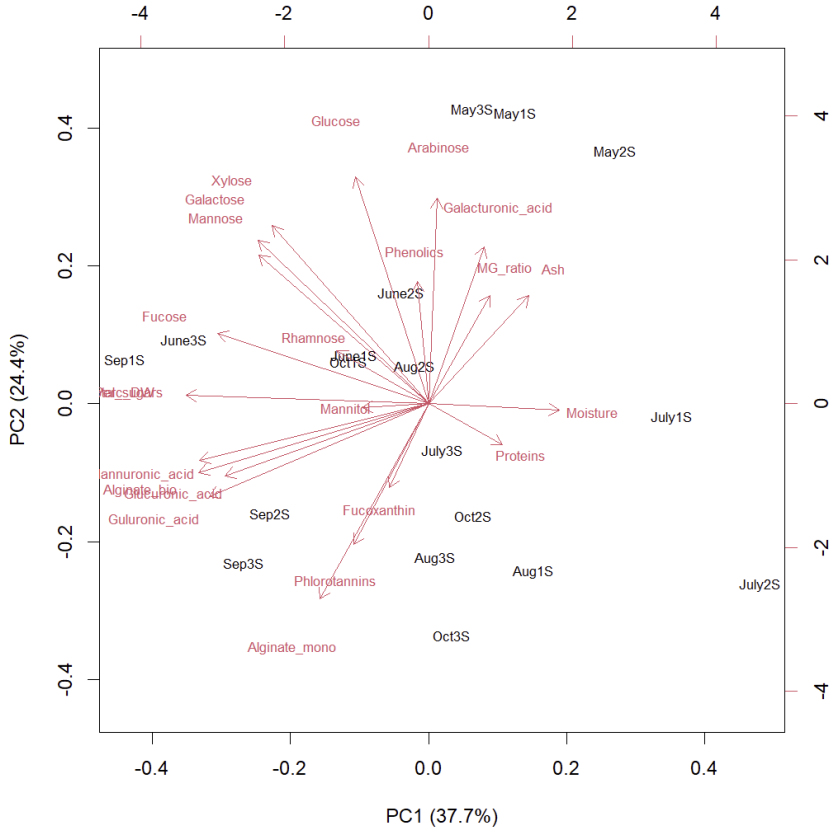

Figure 5: A biplot showing the results of PCA on the shade-dried samples in the biochemical dataset.

Performing PCA on the whole element dataset gave the biplot displayed in Figure 6. In contrast to the biochemical composition results, here the samples tend to cluster by month, rather than processing method. This suggests there's seasonal variation in the elemental composition of the seaweed. The samples from months before the potential ash exposure tend to have higher levels of Na, As and K, but conclusive results can't be drawn from this biplot. Though there doesn't appear to be much variation due to the processing method used, it may make things analysis easier to separate the dataset into processing type again.

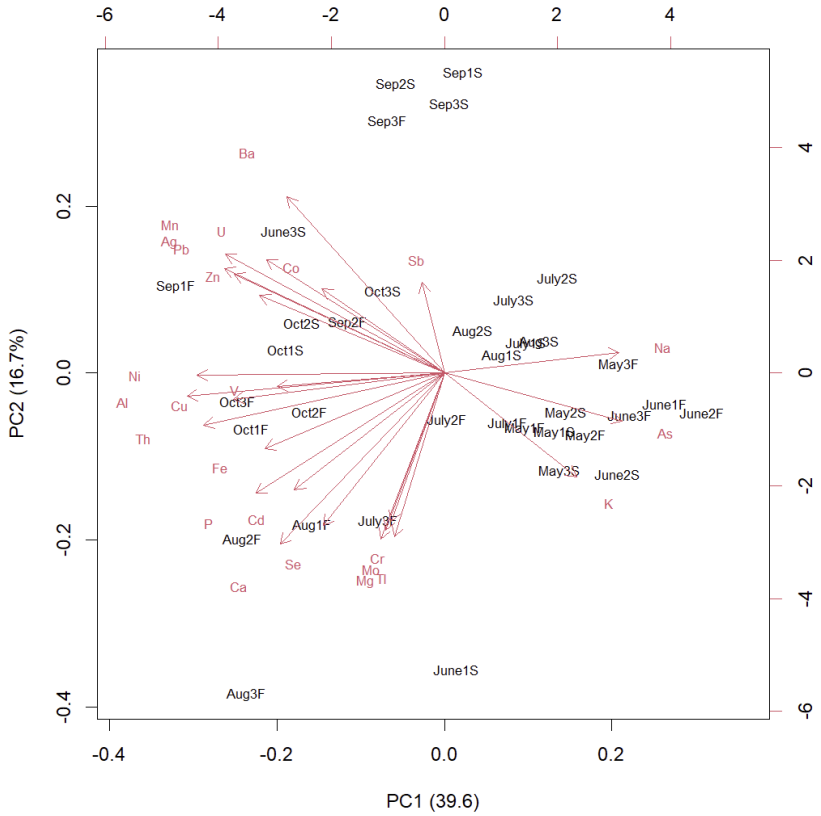

Figure 6: A biplot displaying the results of PCA performed the whole element dataset. The samples cluster by month rather than sample type.

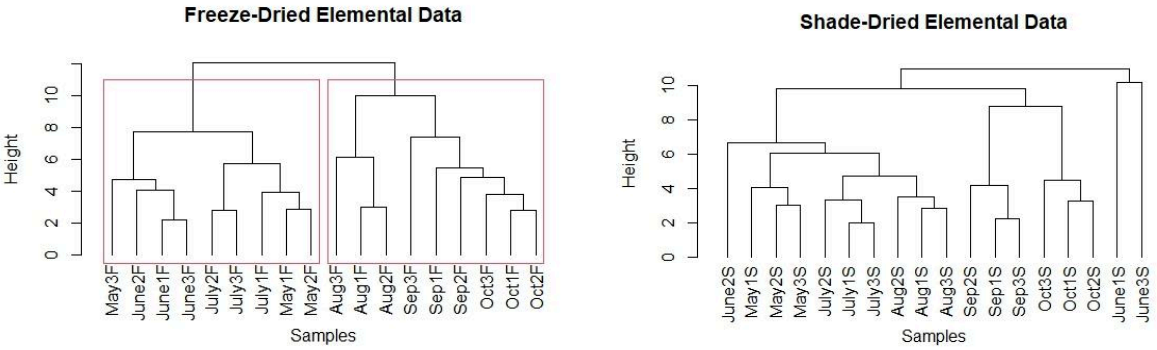

Figure 7: Dendrogram displaying the results of agglomerative hierarchical clustering of the freeze-dried and shade-dried elemental data. The red boxes outline clusters found.

The dendrograms in Figure 7 display the results of performing hierarchical cluster analysis on the freeze-dried and shade-dried subsets of the element dataset. In the freeze-dried

subset the samples from the same months generally tend to cluster together, and then form two larger clusters in 3 month blocks. This suggests there is strong seasonal variation in the elemental composition of the seaweed. The cluster analysis for the shade-dried subset again shows clustering of samples from months close to each other, but to a lesser degree. There are two samples from June that don't cluster with the others. This suggests that the processing method used does have an effect on the elemental composition of the sargassum samples, as the results here are different. The samples from June could also be outliers.

A)

	PC1	PC2
Na	0.20790169	-0.13042891
Mg	-0.10048826	0.28668364
Al	-0.26136188	0.01141101
P	-0.23150676	0.17358264
K	0.17468238	-0.17506047
Ca	-0.20233151	0.27775707
V	-0.16994270	-0.13642322
Cr	-0.06157714	0.30866284
Mn	-0.24443068	-0.18632137
Fe	-0.25481056	0.03718042
Co	-0.22866959	-0.22893390
Ni	-0.25429000	0.05231491
Cu	-0.21594197	-0.10378249
Zn	-0.21448345	-0.26749453
As	0.21554180	-0.03952500
Se	-0.18210208	0.24052994
Mo	-0.06365003	0.31928912
Ag	-0.24838575	-0.07943738
Cd	-0.18032629	0.12193555
Sb	-0.06002646	-0.18909751
Ba	-0.17267974	-0.26954225
Tl	-0.15210809	0.26163145
Pb	-0.21165409	-0.28747260
Th	-0.25541227	0.05119915
U	-0.20472485	-0.17500271

B)

	PC1	PC2
Na	0.08382472	0.28194369
Mg	0.02386956	0.33627050
Al	-0.32140361	0.12675685
P	-0.02439673	0.12205998
K	0.17335424	0.27831891
Ca	-0.01291261	0.25315322
V	-0.22025798	0.16780202
Cr	-0.06740364	0.31918441
Mn	-0.28196966	-0.20897972
Fe	-0.17406352	0.31636279
Co	-0.24090753	0.11301953
Ni	-0.31528725	-0.02011208
Cu	-0.28533699	0.20490897
Zn	-0.17810524	-0.03577981
As	0.11517365	0.24399060
Se	0.08678133	0.06304921
Mo	-0.04511496	0.22764381
Ag	-0.28107835	-0.15161434
Cd	-0.09352542	0.10916122
Sb	-0.14741504	0.05359888
Ba	-0.27404600	-0.17503814
Tl	0.07295040	0.17058324
Pb	-0.32492120	0.07516529
Th	-0.24412719	0.16867520
U	-0.21665030	-0.24027838

Figure 8: A) The loadings for the principal components of the freeze-dried element subset B) The loadings for the principal components of the shade-dried element subset

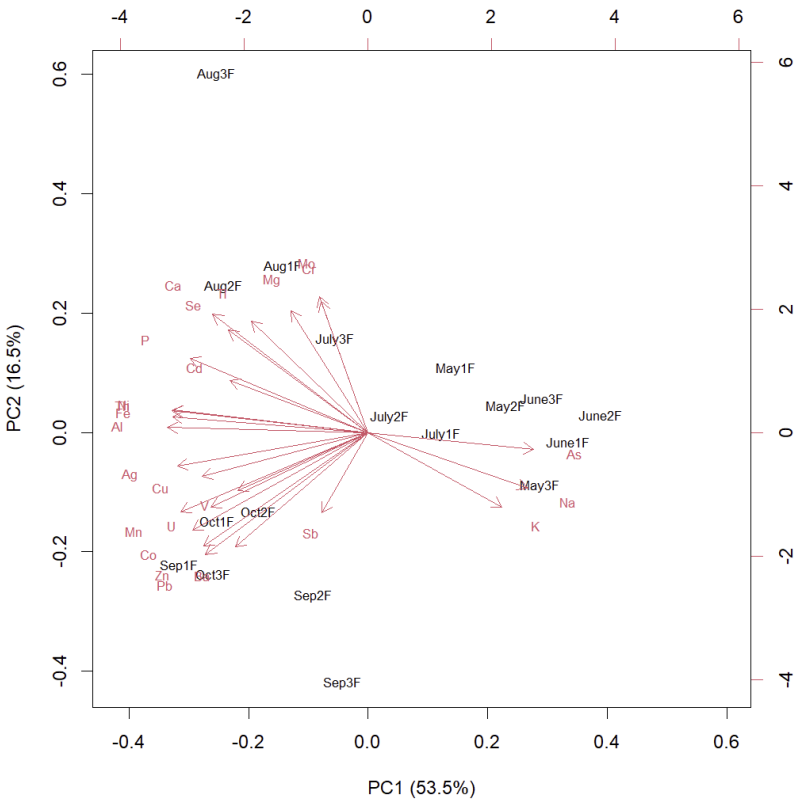

Figure 9: A biplot displaying the results of PCA performed on the freeze-dried subset of elemental data

The biplot in Figure 9 shows the results of performing PCA on the freeze-dried subset of the element data. Here there are seasonal variations in the data, with levels of As, Na and K higher earlier in the year. The variation splitting May-July apart from the other months is along PC1. The principal component loadings in Figure 8A shows that the elements most responsible for variation along this principal component are Al, Fe, Th, Ni, Mn, which suggests the levels of these elements have increased in the sargassum samples exposed to ash. This could be due to the seaweed absorbing elements from the ash.

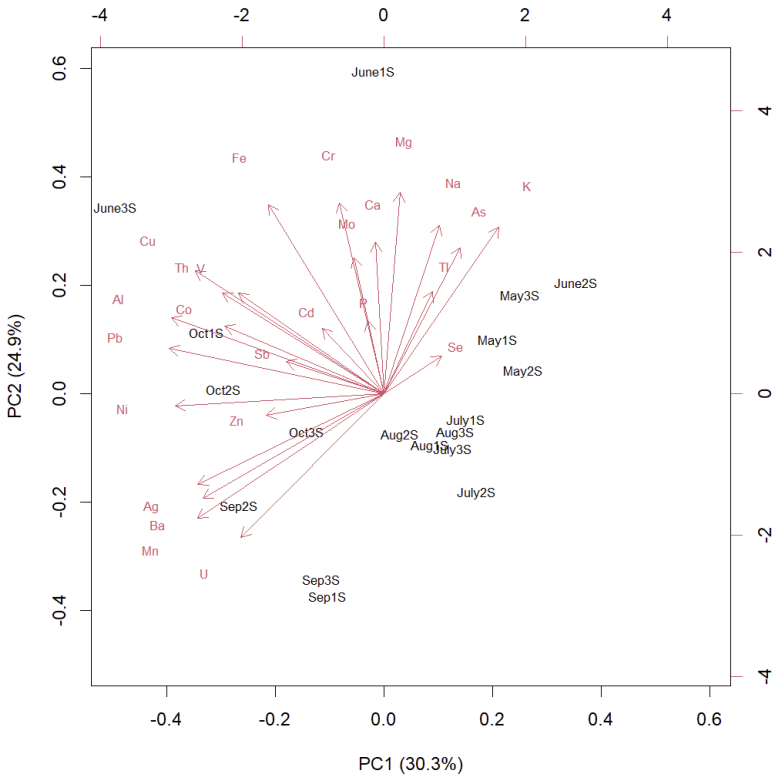

Figure 10: A biplot displaying the results of PCA performed on the shade-dried subset of elemental data

Once again, the PCA biplot for the shade-dried samples in Figure 10 displays less conclusive results than the freeze-dried one. It does still show that the samples before potential ash exposure have higher levels of As, Na and K, along with many other elemental variations. This could be due to the seaweed samples being affected by the environment as they dried, which could affect the elemental composition. Overall, the results from the freeze-dried samples are more useful for detecting the potential influence of ash exposure.

Conclusion

This report found that processing methods had a large influence on the biochemical composition of the sargassum samples, with freeze-dried samples having higher levels of fucoxanthin. The freeze-dried samples also displayed the potential impact of ash exposure on the sargassum, as the samples taken after exposure had lower levels of alginates according to the principal component analysis performed. The shade-dried samples also displayed a decrease in monosaccharides after ash exposure.

Cluster analysis on the elemental composition of the sargassum showed it had strong seasonal variations. There also appeared to be potential influence from ash exposure, as the freeze-dried samples showed lower levels of Na, K and As, and higher levels of Al, Fe, Th, Ni, Mn after mixing with the ash. This suggests that the sargassum took up some elements from the ash and lost others.

Overall the effect of the ash exposure was easier to see in the freeze-dried samples. This could be due to the shade-dried samples getting slightly corrupted in the time it took them to dry. This suggests that freeze-drying may be a better processing method, though it is more expensive.

Further research:
The next step would be to collect samples from a year without a volcanic eruption to provide a control dataset. This would help to differentiate the effects of natural seasonal variation from the effects of ash exposure.

References

Machado, C. B., Marsh, R., Hargreaves, J. K., Oxenford, H. A., Maddix, G. M., Webber, D. F., & Tonon, T. (2024). Changes in holopelagic Sargassum spp. biomass composition across an unusual year. Proceedings of the National Academy of Sciences, 121(23), e2312173121.

Appendix: Code

```
##### Begin with the biochemical dataset #####

#Read in the biochemical data, which has both column names and row names
bio_data = read.csv("Biochemical.csv", header = T,  row.names= 1)

##### PCA on the whole biochemical dataset #####

#perform pca, scaling the data to unit variance
#start by looking at all the biochemical data together
bio_pca = prcomp(bio_data[,4:26], scale = TRUE)
#view summary
summary(bio_pca)
#The proportion of variance shows that pc1 and pc2 are a lot more important
#than the others so can just focus on these.
#view loadings
bio_pca$rotation
#make a biplot
biplot(bio_pca, xlab="PC1 (31.5%)", ylab="PC2 (28.6%)", cex = 0.7)
#cex reduces the text size. Added the proportion of variance to x and
#y labels.

##### Separating the sample types ######

#Now look at the processing methods individually
#Separate the data into freeze dried and shade dried
bio_copy = bio_data #create a copy of the data before splitting in case it
#affects the original dataset
bio_frozen = subset(bio_copy, Samples=="Frozen")
#subset with just frozen samples
bio_shade = subset(bio_copy, Samples=="Shade")

##### Cluster analysis on the freeze-dried biochemical samples #####

#Perform cluster analysis on the frozen subset of the biochemical data
bf_data = scale(bio_frozen[,4:26]) #scale the data to unit variance
bf_d = dist(bf_data) #create a distance matrix with euclidean metric
bf_hc = hclust(bf_d) #cluster using the complete linkage method

#plot a dendrogram
install.packages("dendextend")
library(dendextend) #need this package to create nice a dendrogram from the
#clustering

bf_dend = as.dendrogram(bf_hc) #turn it into a dendrogram
plot(bf_dend, main = "Freeze-Dried Biochemical Data", xlab = "Samples",
    ylab = "Height") #plot the dendrogram
rect.hclust(bf_hc, 2) #draw a rectangle around the 2 clusters found

##### PCA on the freeze-dried biochemical samples #####
```

```
#Perform pca on the freeze-dried subset
bf_pca = prcomp(bio_frozen[,4:26], scale = TRUE)
#view summary
summary(bf_pca)
#view loadings
bf_pca$rotation
#make a biplot
biplot(bf_pca, xlab="PC1 (35.6%)", ylab="PC2 (23.4%)", cex = 0.7)

##### Cluster analysis on the shade-dried biochemical samples #####

bs_data = scale(bio_shade[,4:26])
bs_d = dist(bs_data)
bs_hc = hclust(bs_d)
bs_dend = as.dendrogram(bs_hc) #turn it into a dendrogram
plot(bs_dend, main = "Shade-Dried Biochemical Data", xlab = "Samples",
    ylab = "Height")
rect.hclust(bs_hc, 2)

##### PCA on the shade-dried biochemical samples #####

#Perform pca on the shade-dried subset
bs_pca = prcomp(bio_shade[,4:26], scale = TRUE)
#view summary
summary(bs_pca)
#view loadings
bs_pca$rotation
#make a biplot
biplot(bs_pca, xlab="PC1 (37.7%)", ylab="PC2 (24.4%)", cex = 0.7)

##### Next look at the elemental dataset #####

#Read in the elemental data, which has both column names and row names
element_data = read.csv("Elements.csv", header = T,  row.names= 1)

##### Perform PCA on the whole elemental dataset ####

element_pca = prcomp(element_data[,3:27], scale = TRUE)
#excluding the total elements column
#view summary
summary(element_pca)
#view loadings
element_pca$rotation
#make a biplot
biplot(element_pca, xlab="PC1 (39.6)", ylab="PC2 (16.7%)", cex = 0.7)

##### Separate the sample types ######

#Now look at the processing methods individually
#Separate the data into freeze dried and shade dried
element_copy = element_data #create a copy of the data before splitting in case it
#affects the original dataset
```

```
element_frozen = subset(element_copy, Samples=="Frozen")
#subset with just frozen samples
element_shade = subset(element_copy, Samples=="Shade-dried")

##### Cluster analysis on the freeze-dried elemental samples #####

ef_data = scale(element_frozen[,3:27])
ef_d = dist(ef_data)
ef_hc = hclust(ef_d, method = "complete")
ef_dend = as.dendrogram(ef_hc) #turn it into a dendrogram
plot(ef_dend, main = "Freeze-Dried Elemental Data", xlab = "Samples",
    ylab = "Height")
rect.hclust(ef_hc, 2) #draw a rectangle around the 2 clusters found

##### PCA on the freeze-dried elemental samples #####

#Perform pca on the shade-dried subset
ef_pca = prcomp(element_frozen[,3:27], scale = TRUE)
#view summary
summary(ef_pca)
#view loadings
ef_pca$rotation
#make a biplot
biplot(ef_pca, xlab="PC1 (53.5%)", ylab="PC2 (16.5%)", cex = 0.7)

##### Cluster analysis on the shade-dried elemental samples #####

es_data = scale(element_shade[,3:27])
es_d = dist(es_data)
es_hc = hclust(es_d)
es_dend = as.dendrogram(es_hc) #turn it into a dendrogram
plot(es_dend, main = "Shade-Dried Elemental Data", xlab = "Samples",
    ylab = "Height")
rect.hclust(es_hc, 2)

##### PCA on the shade-dried elemental samples #####

#Perform pca on the shade-dried subset
es_pca = prcomp(element_shade[,3:27], scale = TRUE)
#view summary
summary(es_pca)
#view loadings
es_pca$rotation
#make a biplot
biplot(es_pca, xlab="PC1 (30.3%)", ylab="PC2 (24.9%)", cex = 0.7)
```