S41-Monday-1130-Alvarez

DATASET · MAY 2013

7 AUTHORS, INCLUDING:

Jorge Díez
University of Oviedo
49 PUBLICATIONS **335** CITATIONS
SEE PROFILE

Juan José del Coz
University of Oviedo
51 PUBLICATIONS **297** CITATIONS
SEE PROFILE

Automated classification techniques targeted to improve the precision of biomass estimates

Eva Álvarez

Ángel López-Urrutia

Enrique Nogueira

Rafael González-Quirós

Pablo González

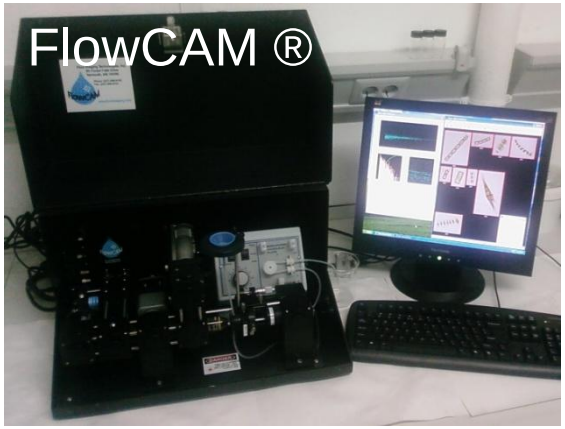
Jorge Díez

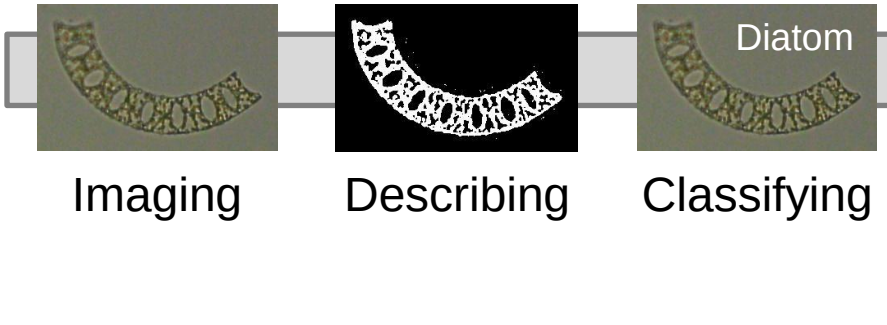
Juan José del Coz

Instituto Español de Oceanografía
Centro Oceanográfico de Gijón
Asturias, Spain

Artificial Intelligence Center
University of Oviedo
Asturias, Spain

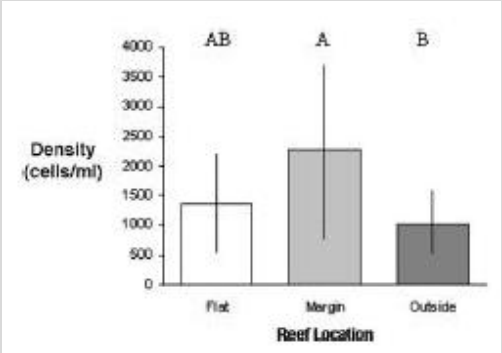
Automatic sampling and classification techniques





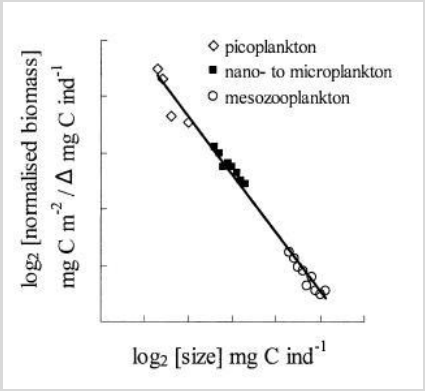
OBJECTIVES

Abundance per taxonomic group



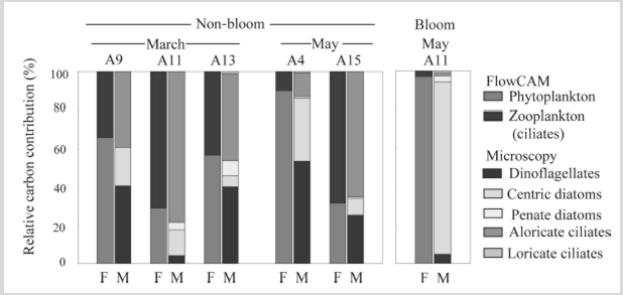
Littman, R.A. et al.
J. Exp. Mar. Biol. Ecol. (2008)

Community size-structure



San Martin, E. et al.
Limnol. Oceanogr. (2006)

Biomass per taxonomic or functional group



Ide, K.et al.
J. Plankton Res. (2007)

1

Community size-structure → Improvement of the biovolume calculation

Projected area based method vs shape based method

2

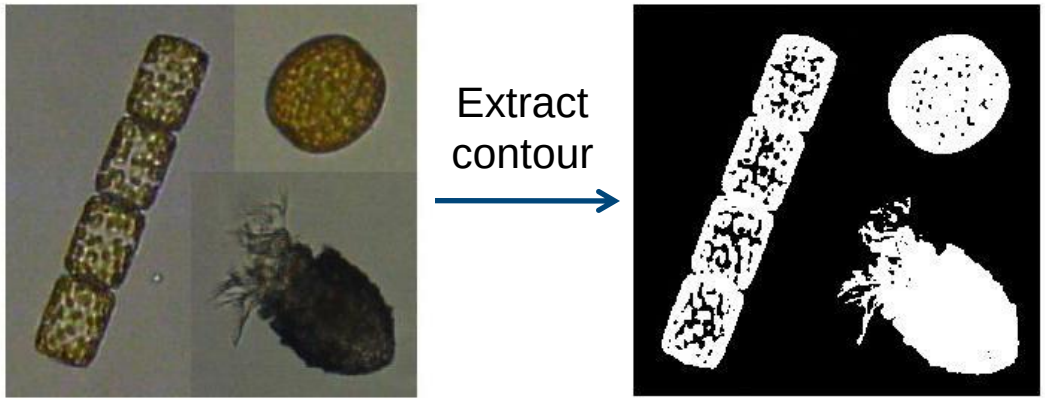
Biomass per group → Biomass-oriented formulation of the classification algorithm

Abundance oriented method vs biomass oriented method

1. Biovolume calculation

Biovolume based on projected area

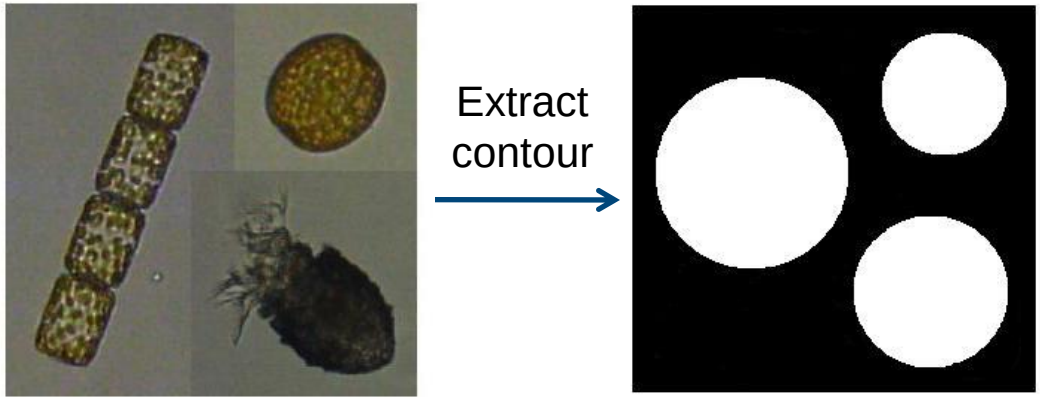
Automatic sampling devices based on image-analysis measure particle volume from the Equivalent Spherical Diameter (ESD)



1. Biovolume calculation

Biovolume based on projected area

Automatic sampling devices based on image-analysis measure particle volume from the Equivalent Spherical Diameter (ESD)

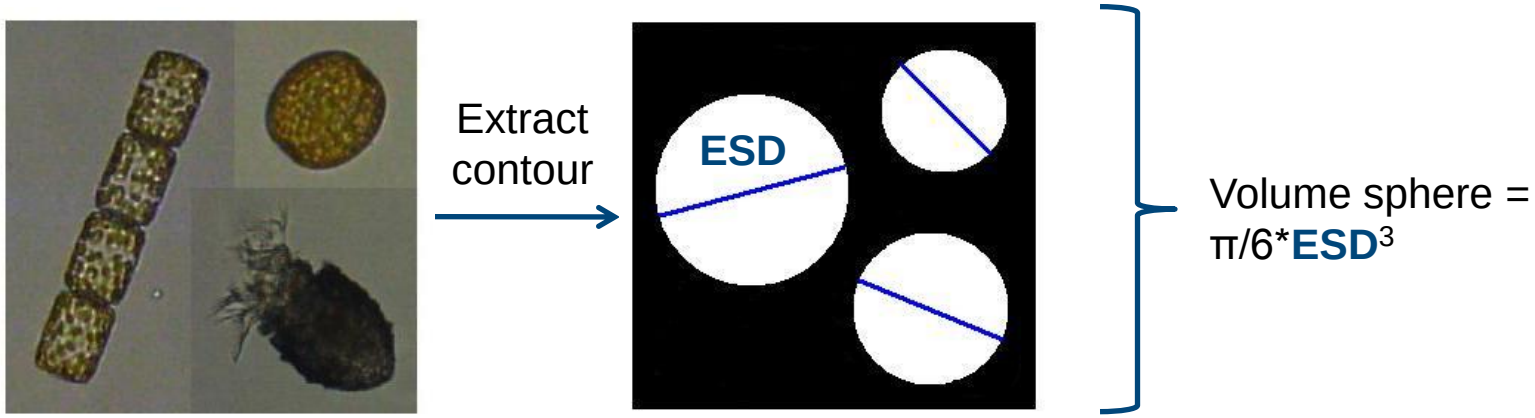


considering every particle as an sphere independently of its shape.

1. Biovolume calculation

Biovolume based on projected area

Automatic sampling devices based on image-analysis measure particle volume from the Equivalent Spherical Diameter (ESD)

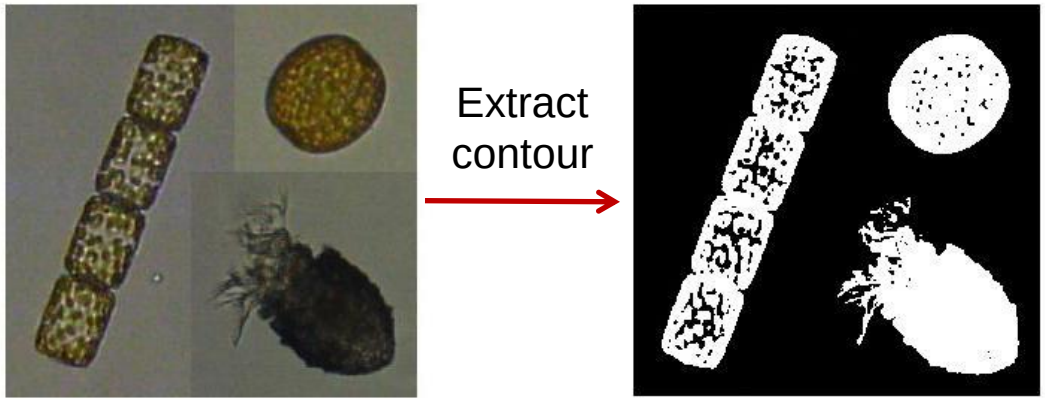


considering every particle as an sphere independently of its shape.

1. Biovolume calculation

Biovolume based on shape

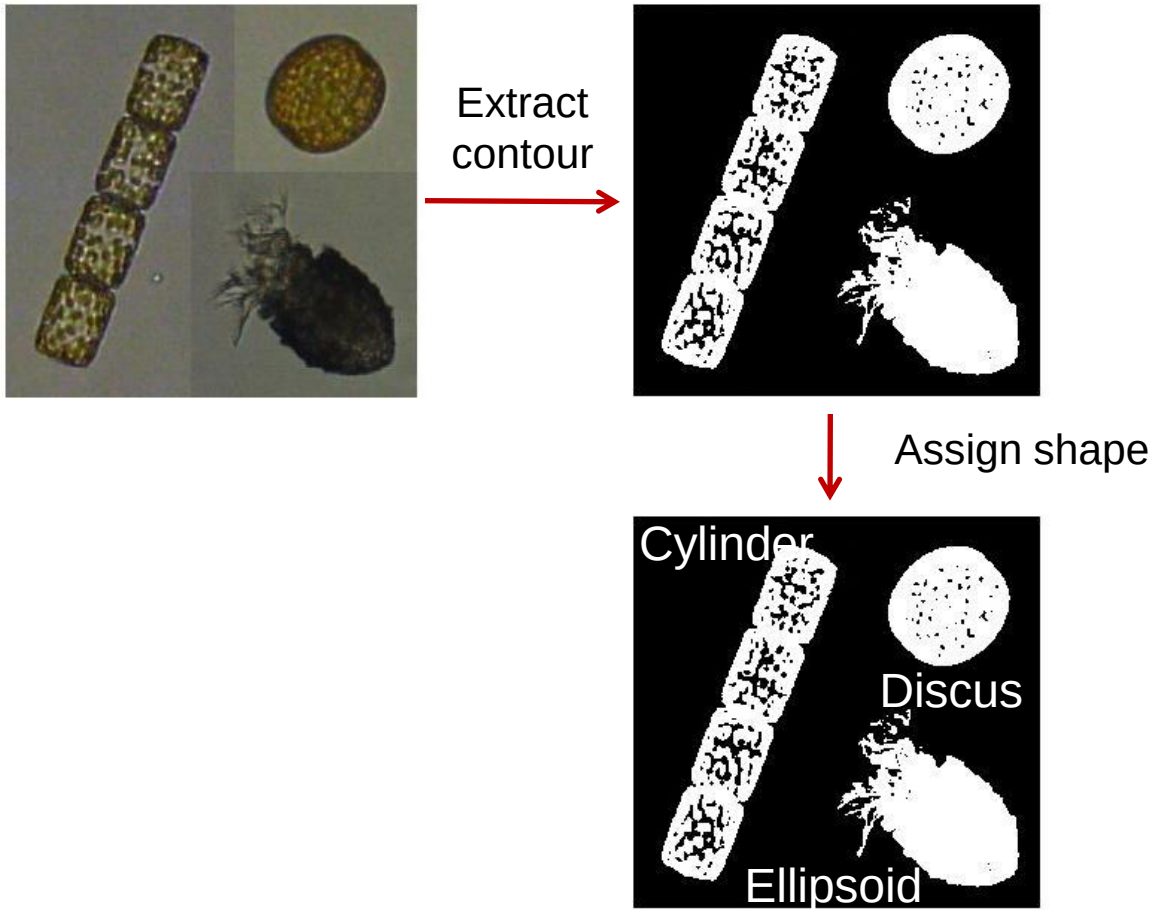
Traditionally biovolume has been measured assigning shape to the cells and calculating the volume accordingly.



1. Biovolume calculation

Biovolume based on shape

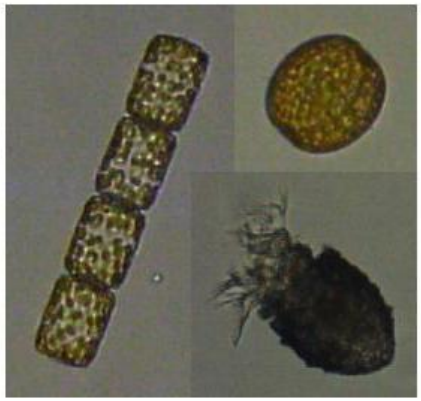
Traditionally biovolume has been measured assigning shape to the cells and calculating the volume accordingly.



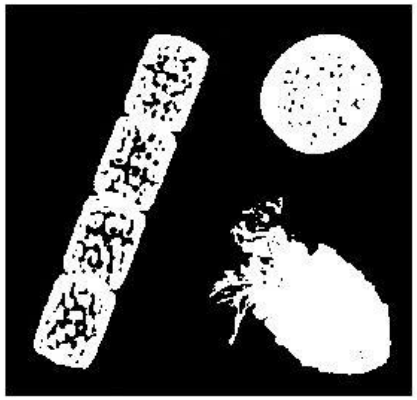
1. Biovolume calculation

Biovolume based on shape

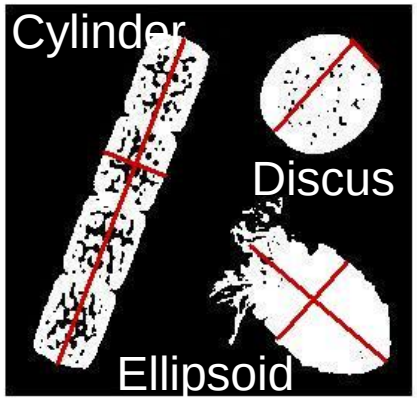
Traditionally biovolume has been measured assigning shape to the cells and calculating the volume accordingly.



Extract contour



Assign shape



→ Volume discus = π/4*$\mathbf{length}^2$*$\mathbf{width}$

→ Volume cylinder = π/4*$\mathbf{width}^2$*$\mathbf{length}$

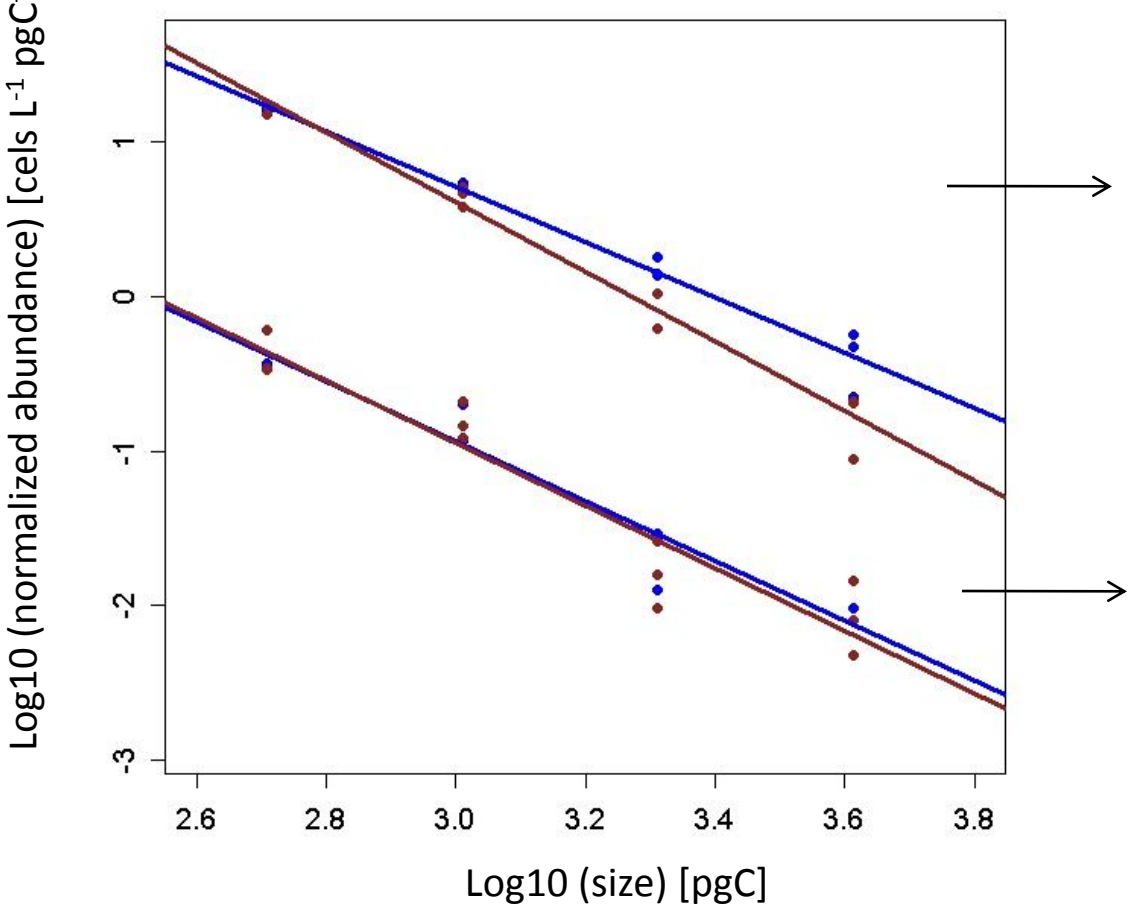
→ Volume ellipsoid = π/4*$\mathbf{length}^2$*$\mathbf{width}$

1. Biovolume calculation

The way we measure the volume can modify the size-spectra slope depending on the morphological composition of the sample.

Biovolume based on projected area

Biovolume based on shape



Coastal station

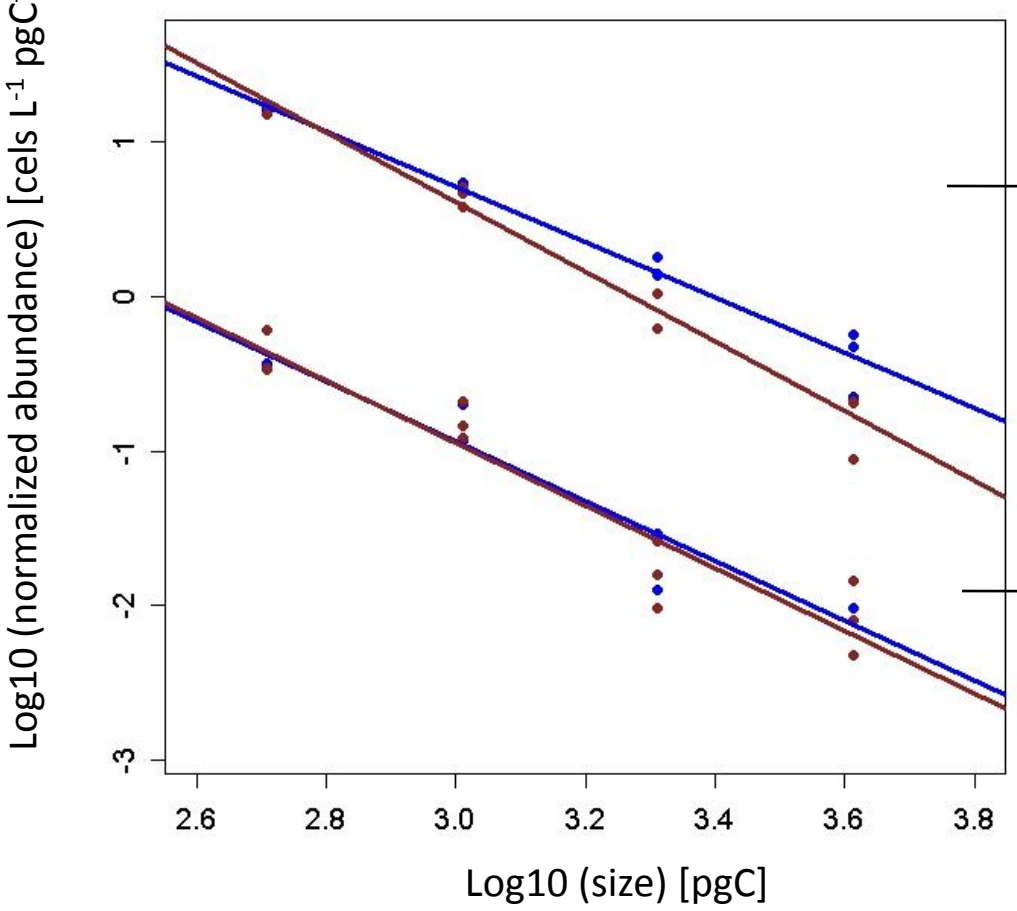
Oceanic station

1. Biovolume calculation

The way we measure the volume can modify the size-spectra slope depending on the morphological composition of the sample.

Biovolume based on projected area

Biovolume based on shape



Coastal station

Sphere	Ellipsoid	Cylinder
8 %	11 %	81 %

T-test p-value: 0.004

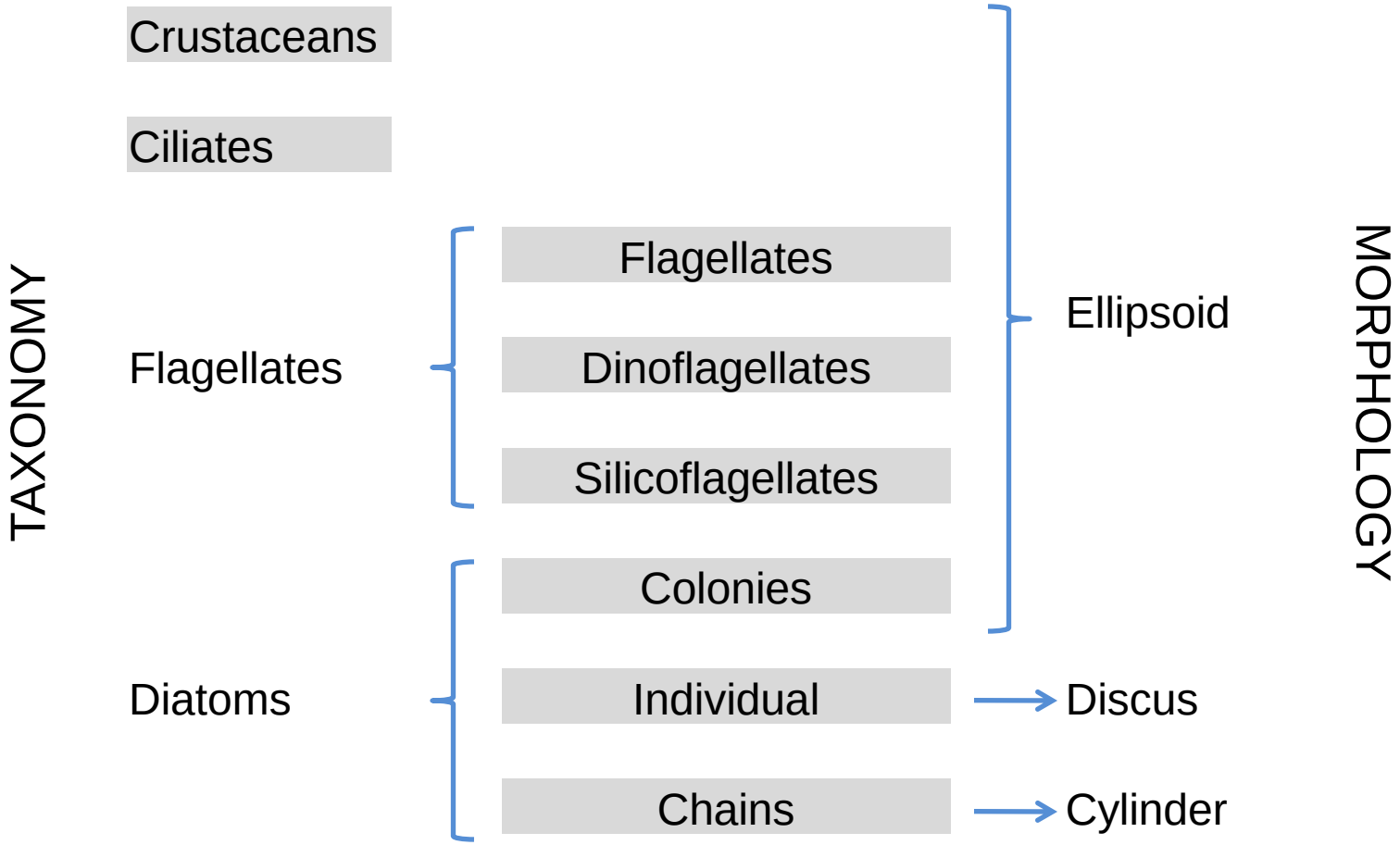
Oceanic station

Sphere	Ellipsoid	Cylinder
25 %	62 %	14 %

T-test p-value: 0.773

1. Biovolume calculation

To know automatically the particle shape we need to classify taxonomically in morphologically homogeneous classes.



1. Biovolume calculation

Abundance-oriented Support Vector Machine

Training set

$$S = \{(x_1, y_1, c_1), (x_2, y_2, c_2) \ldots (x_n, y_n, c_n)\}$$

1. Biovolume calculation

Abundance-oriented Support Vector Machine

Training set

$$S = \{(x_1, y_1, c_1), (x_2, y_2, c_2) \ldots (x_n, y_n, c_n)\}$$

x	Var1	Var2	Var3	…
1				
2				
3				
4				
…				

5500 images

170 image features

1. Biovolume calculation

Abundance-oriented Support Vector Machine

Training set

$$S = \{(x_1, y_1, c_1), (x_2, y_2, c_2) \ldots (x_n, y_n, c_n)\}$$

x	Var1	Var2	Var3	…	y
1					Diatom
2					Ciliate
3					Detritus
4					Crustacean
…					…

5500 images

9 classes

170 image atributes

1. Biovolume calculation

Abundance-oriented Support Vector Machine

Training set

$$S = \{(x_1, y_1, c_1), (x_2, y_2, c_2) \ldots (x_n, y_n, c_n)\}$$

Model

$$h: S' \longrightarrow Y'$$

1. Biovolume calculation

Abundance-oriented Support Vector Machine

Training set

$$S = \{(x_1, y_1, c_1), (x_2, y_2, c_2) \dots (x_n, y_n, c_n)\}$$

Model

$$h: \; S' \longrightarrow Y'$$

Loss function

$$\Delta_{LOSS}(h, S') = \frac{\sum_{x_i \in S'} c_i [h(x_i) \neq y_i]}{\sum_{x_i \in S'} c_i}$$

1. Biovolume calculation

Abundance-oriented Support Vector Machine

Training set

$$S = \{(x_1, y_1, c_1), (x_2, y_2, c_2) \ldots (x_n, y_n, c_n)\}$$

Model

$$h: S' \longrightarrow Y'$$

Loss function

$$\Delta_{LOSS}(h, S') = \frac{\sum_{x_i \in S'} c_i [h(x_i) \neq y_i]}{\sum_{x_i \in S'} c_i}$$

An error occurs when the predicted class does not mach the actual class.

1. Biovolume calculation

Abundance-oriented Support Vector Machine

Training set

$$S = \{(x_1, y_1, c_1), (x_2, y_2, c_2) \ldots (x_n, y_n, c_n)\}$$

Model

$$h: S' \longrightarrow Y'$$

Loss function

$$\Delta_{LOSS}(h, S') = \frac{\sum_{x_i \in S'} c_i [h(x_i) \neq y_i]}{\sum_{x_i \in S'} c_i}$$

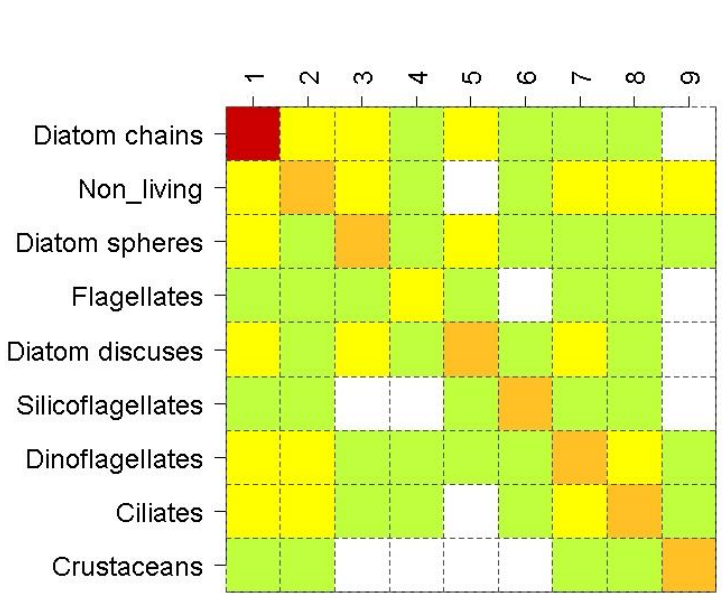
An error occurs when the predicted class does not mach the actual class.

In an abundance-oriented classification the cost of error (c) is 1 for all the examples.

1. Biovolume calculation

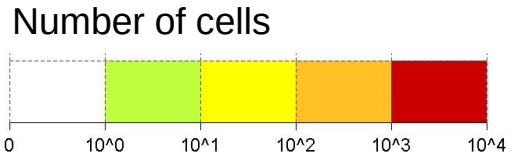
Abundance-oriented Support Vector Machine

Abundance estimates
Accuracy = 0.89



Probability of detection

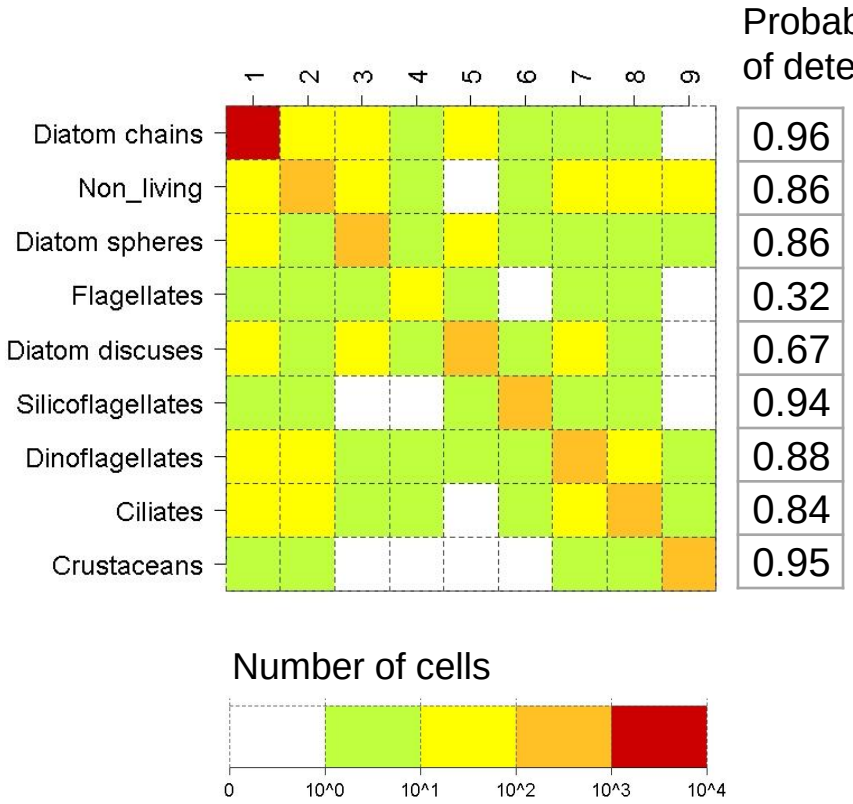
0.96
0.86
0.86
0.32
0.67
0.94
0.88
0.84
0.95



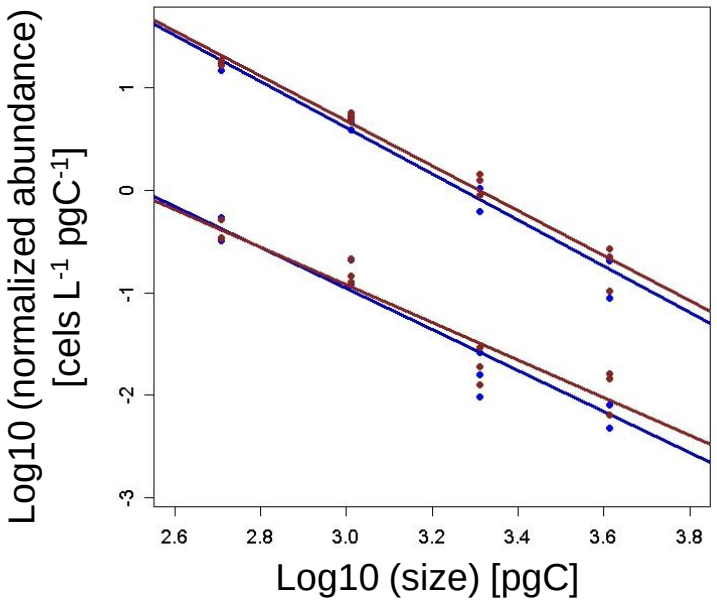
1. Biovolume calculation

Abundance-oriented Support Vector Machine

Abundance estimates
Accuracy = 0.89



Biovolume based on manual shape
Biovolume based on automatic shape



2. Biomass-oriented classification

To improve the biomass estimates the SVM is built considering the error a function of biomass.

Loss function

$$\Delta_{LOSS}(h, S') = \frac{\sum_{x_i \in S'} c_i [h(x_i) \neq y_i]}{\sum_{x_i \in S'} c_i}$$

2. Biomass-oriented classification

To improve the biomass estimates the SVM is built considering the error a function of biomass.

Loss function

$$\Delta_{LOSS}(h, S') = \frac{\sum_{x_i \in S'} c_i [h(x_i) \neq y_i]}{\sum_{x_i \in S'} c_i}$$

The cost of misclassifying one example (c_i) is the value of biomass for this example.

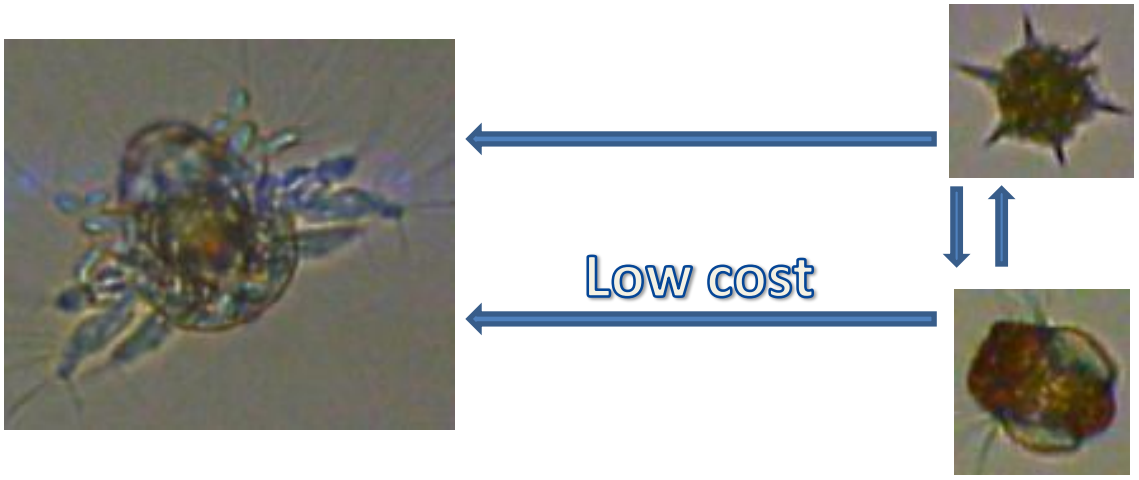
2. Biomass-oriented classification

To improve the biomass estimates the SVM is built considering the error a function of biomass.

Loss function

$$\Delta_{LOSS}(h, S') = \frac{\sum_{x_i \in S'} c_i [h(x_i) \neq y_i]}{\sum_{x_i \in S'} c_i}$$

The cost of misclassifying one example (c_i) is the value of biomass for this example.



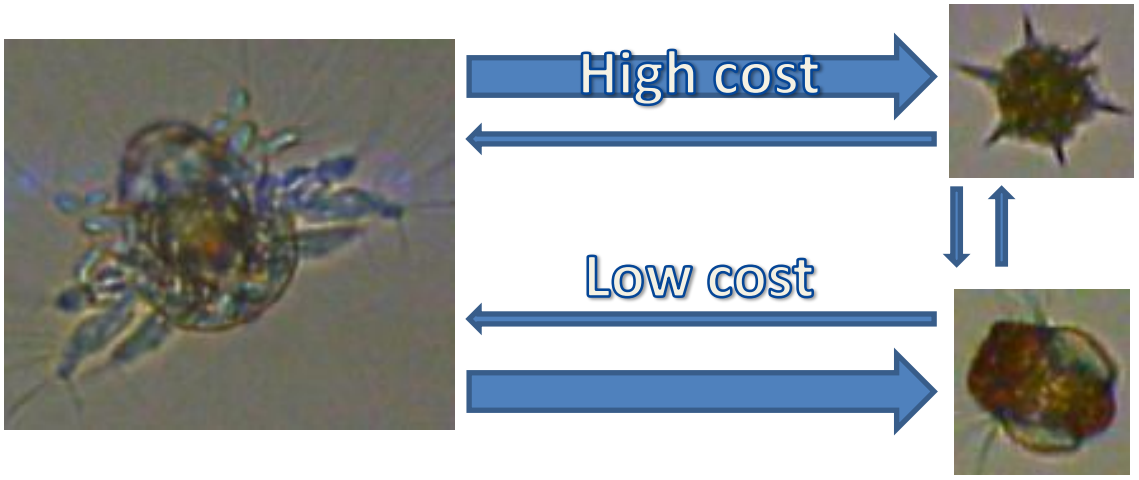
2. Biomass-oriented classification

To improve the biomass estimates the SVM is built considering the error a function of biomass.

Loss function

$$\Delta_{LOSS}(h, S') = \frac{\sum_{x_i \in S'} c_i [h(x_i) \neq y_i]}{\sum_{x_i \in S'} c_i}$$

The cost of misclassifying one example (c_i) is the value of biomass for this example.

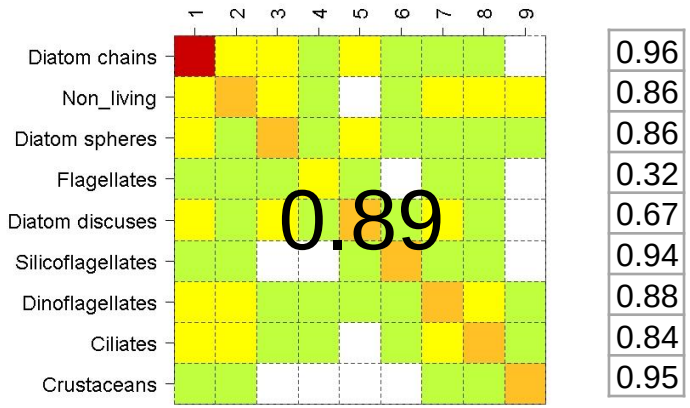


The accuracy of the biomass-oriented SVM on biomass estimations increases a 3%

Abundance estimates

Biomass estimates

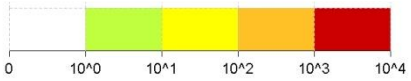
Abundance oriented SVM



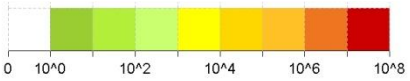
0.96
0.86
0.86
0.32
0.67
0.94
0.88
0.84
0.95

Biomass oriented SVM

Number of cells



pg Carbon

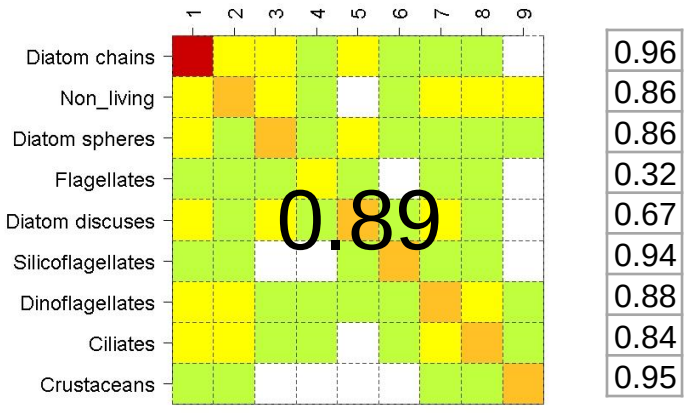


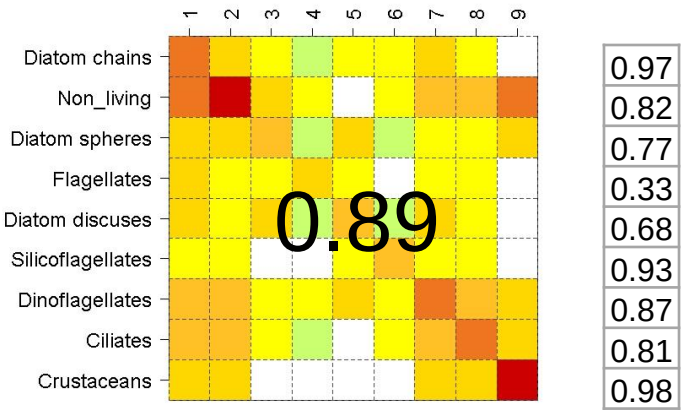
The accuracy of the biomass-oriented SVM on biomass estimations increases a 3%

Abundance estimates

Biomass estimates

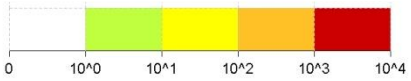
Abundance oriented



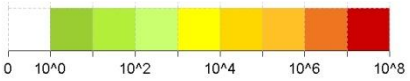


Biomass oriented

Number of cells



pg Carbon

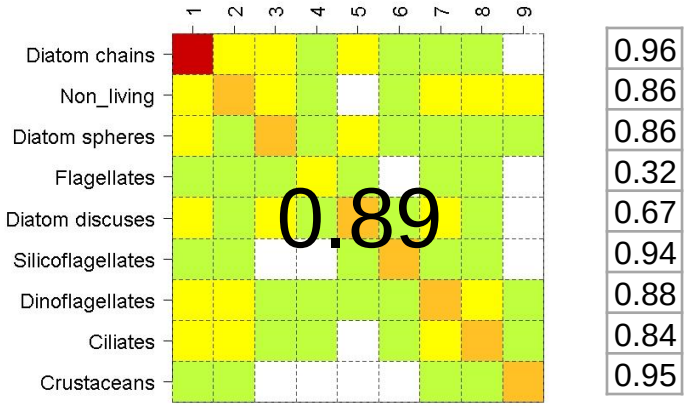


The accuracy of the biomass-oriented SVM on biomass estimations increases a 3%

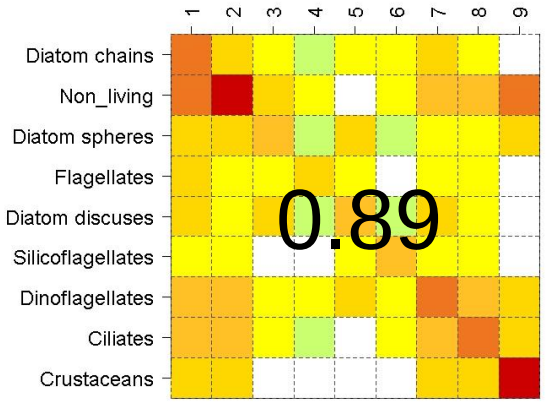
Abundance estimates

Biomass estimates

Abundance oriented

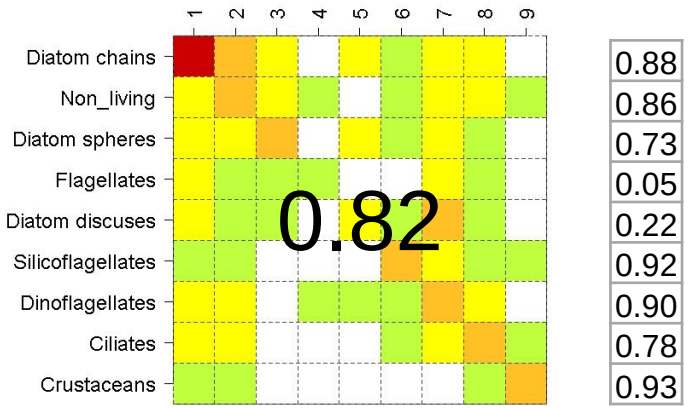


0.96
0.86
0.86
0.32
0.67
0.94
0.88
0.84
0.95



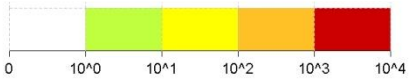
0.97
0.82
0.77
0.33
0.68
0.93
0.87
0.81
0.98

Biomass oriented

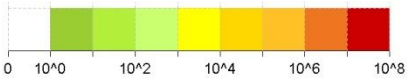


0.88
0.86
0.73
0.05
0.22
0.92
0.90
0.78
0.93

Number of cells



pg Carbon

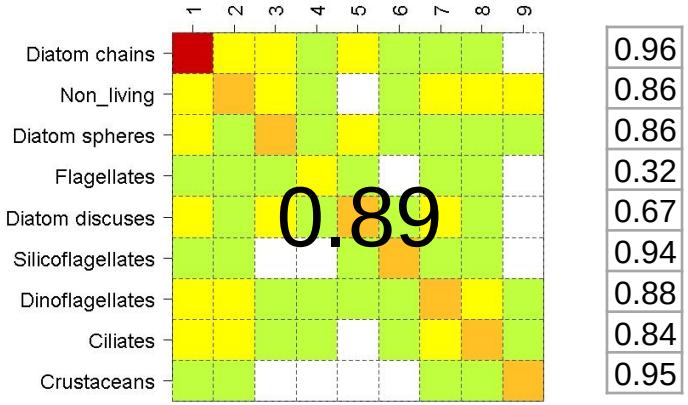


The accuracy of the biomass-oriented SVM on biomass estimations increases a 3%

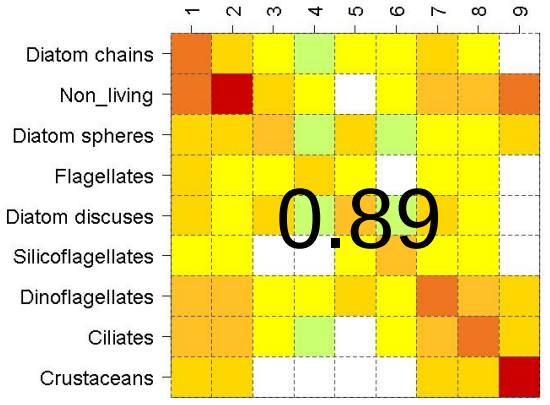
Abundance estimates

Biomass estimates

Abundance oriented

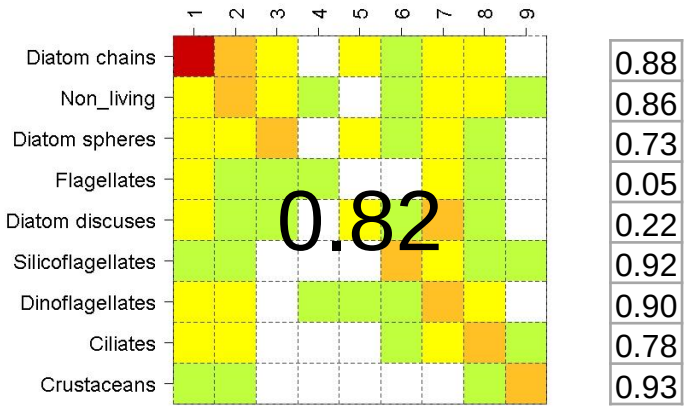


0.96
0.86
0.86
0.32
0.67
0.94
0.88
0.84
0.95

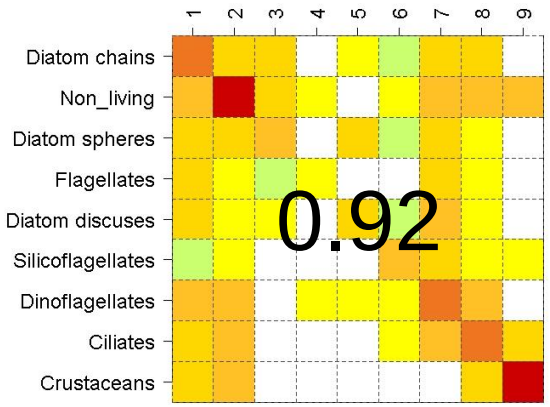


0.97
0.82
0.77
0.33
0.68
0.93
0.87
0.81
0.98

Biomass oriented

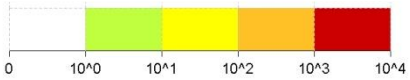


0.88
0.86
0.73
0.05
0.22
0.92
0.90
0.78
0.93

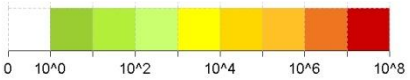


0.93
0.92
0.73
0.05
0.22
0.88
0.89
0.78
0.96

Number of cells



pg Carbon



Conclusions

- Size spectra calculated with projected area-based volume or shape-based volume can be different depending on the morphological composition of the sample.

- To improve the biomass estimates the SVM can be designed considering the error a function of biomass.