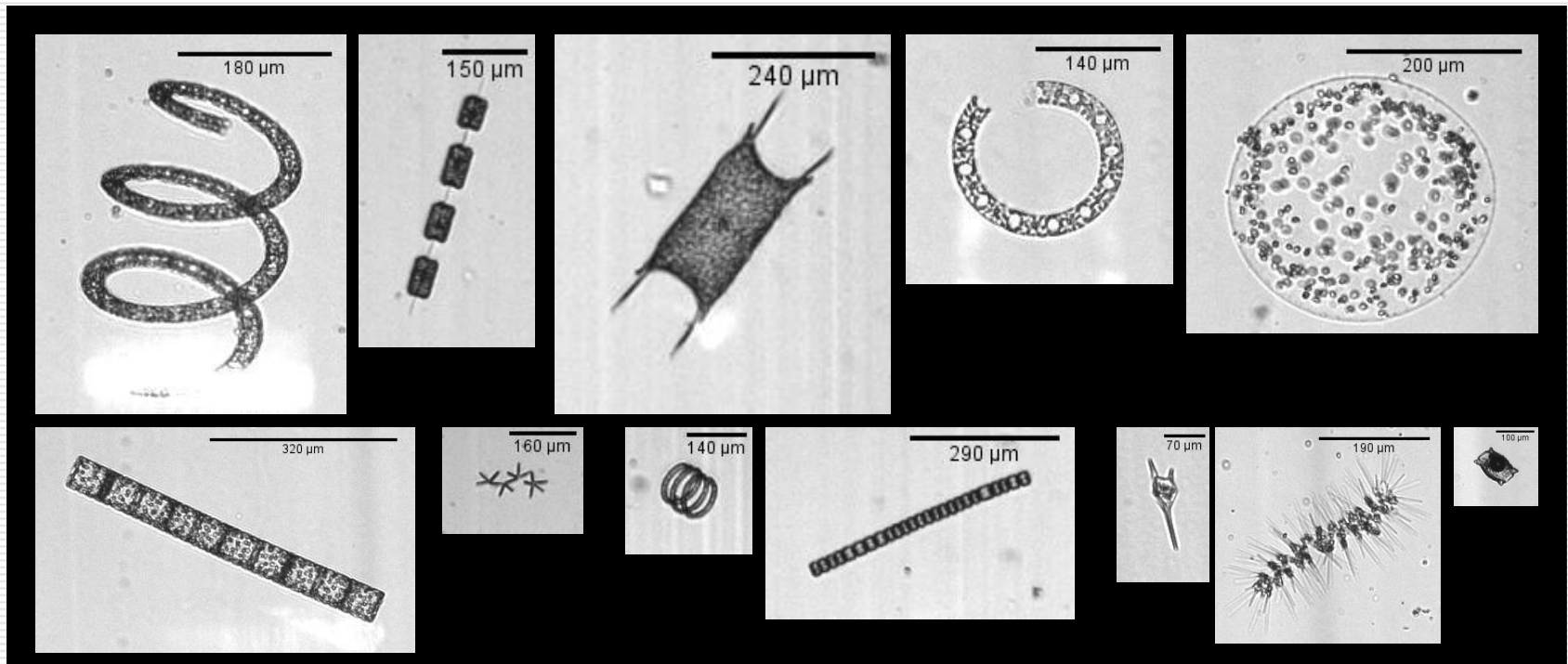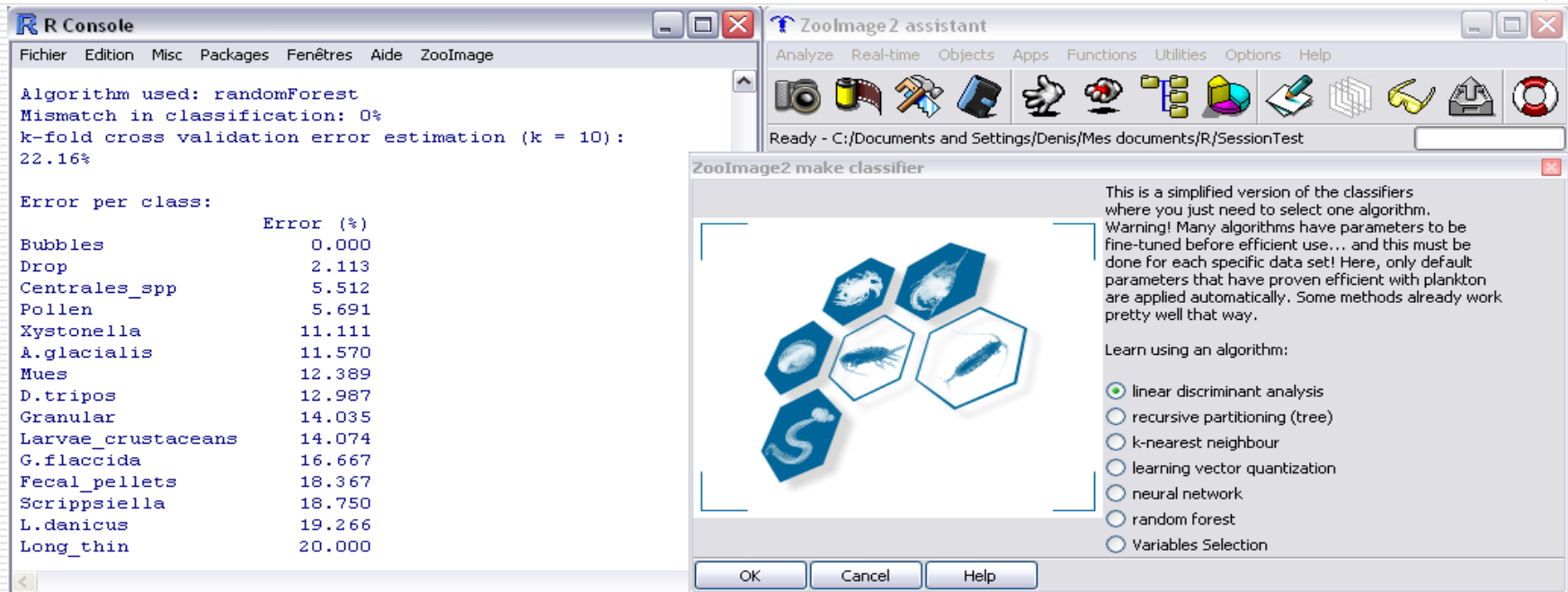# Semi-automatic analysis of FlowCAM phytoplankton data using Zoo/PhytoImage

**Grosjean Ph., K. Denis, G. Wacquet, V. Rousseau, J.-Y. Parent, Ch. Lancelot, D. Hamad, F. Artigas, A. Lefebvre, N. Naud-Masson, D. Maurer, A. Tunin-Ley, F. Colas, C. Belin**

Ifremer

# Zoo/PhytoImage



✔ https://cran.r-project.org/web/packages/zooimage/

✔ Free (open source) software written in R and Java specialized to classify zoo- and phytoplankton digital images

✔ Machine learning (supervised) classification

✔ Adaptable to analyze *any* plankton image, e.g., images from the FlowCAM...

# Zoo/PhytoImage version 3

Install latest R

(http://cran.r-project.org) then in R :

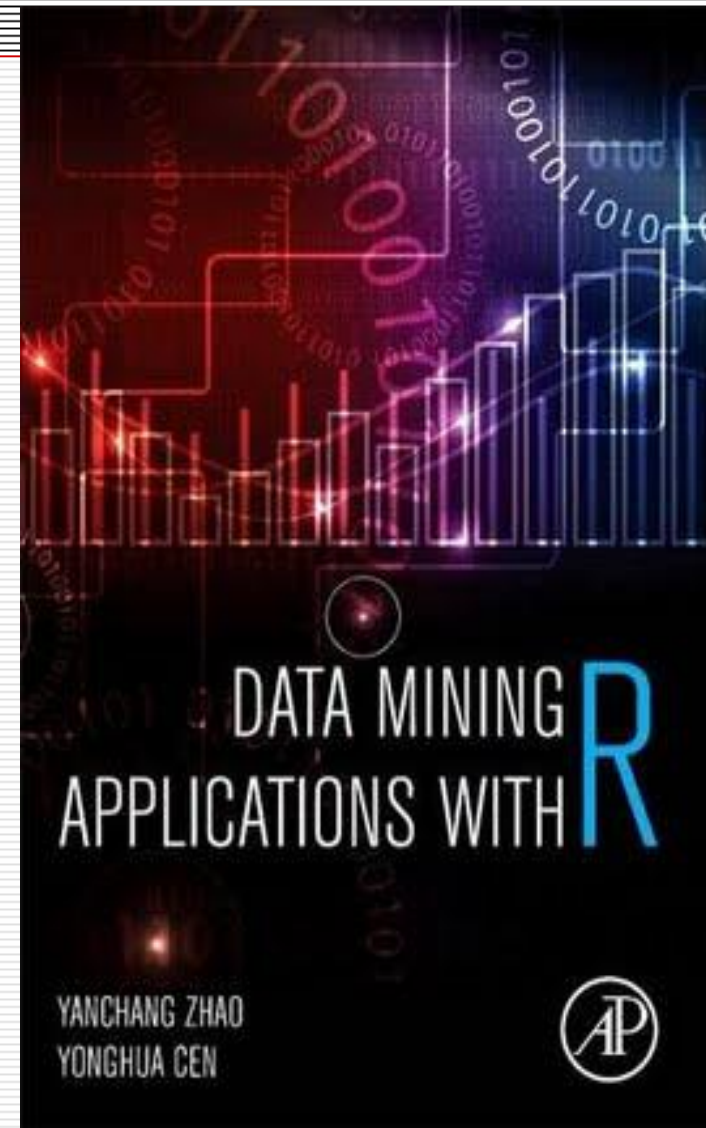> install.packages("zooimage")

> library(zooimage)

See : Data mining application with R

ISBN 978-0124115118,
December 2013.
Academic Press, Elsevier.

Chapter 12 is complete
description and tutorial
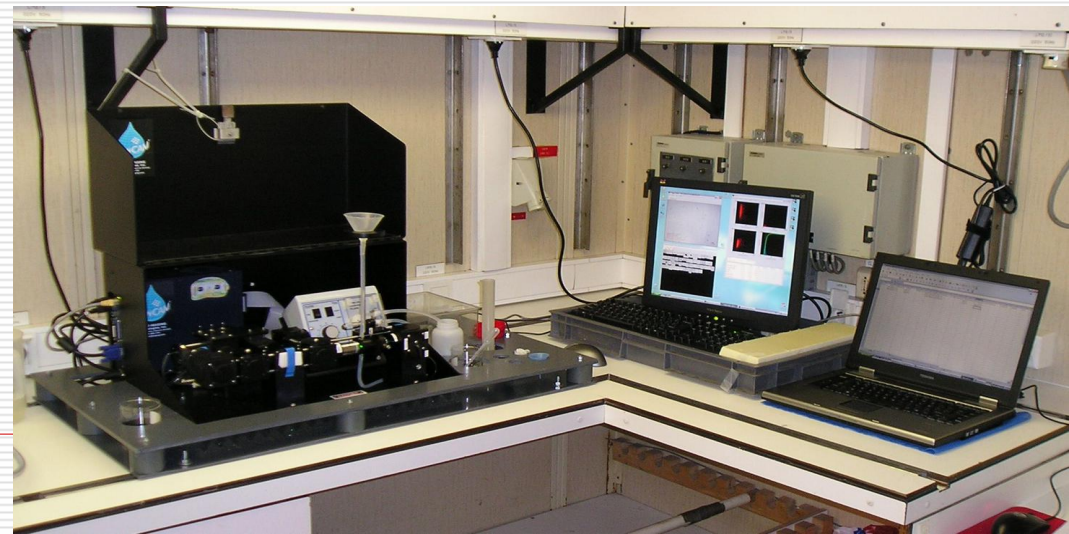of  Zoo/PhytoImage v.3.

Latest version 5 available from:

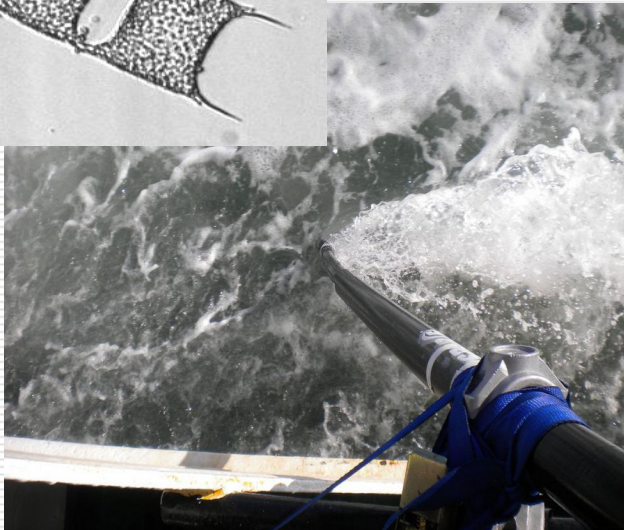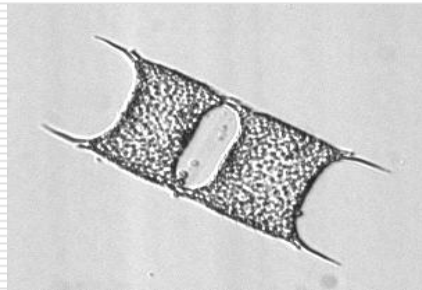https://r-forge.r-project.org/projects/zooimage/
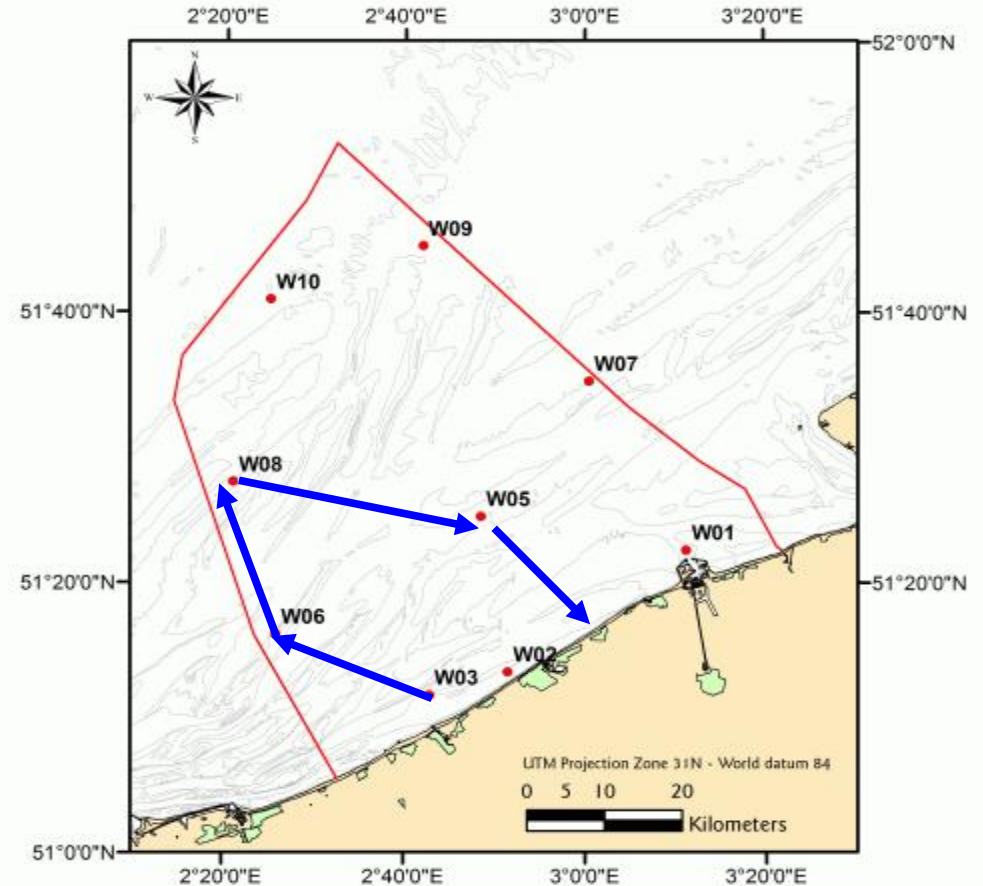
# Application in real-time

- Combination of the FlowCAM with Zoo/PhytoImage

- Aboard the 'Belgica' (Belgian oceanographic ship)

- 25 groups discriminated in real-time (incl. 18 phytoplankton groups)

Application example: a one-day cruise in the Belgian Coastal Zone (BCZ).

Small particles dominating near the coast and the Schelde estuary.

Notice the influence of sandbanks

(no interpolation!)

## Small Particles

log(part./mL)
- 0
- 0.5
- 1
- 1.5
- 2
- 2.5
- 3
- 3.5+

0  5  10  20
Kilometers

Latitude

Longitude
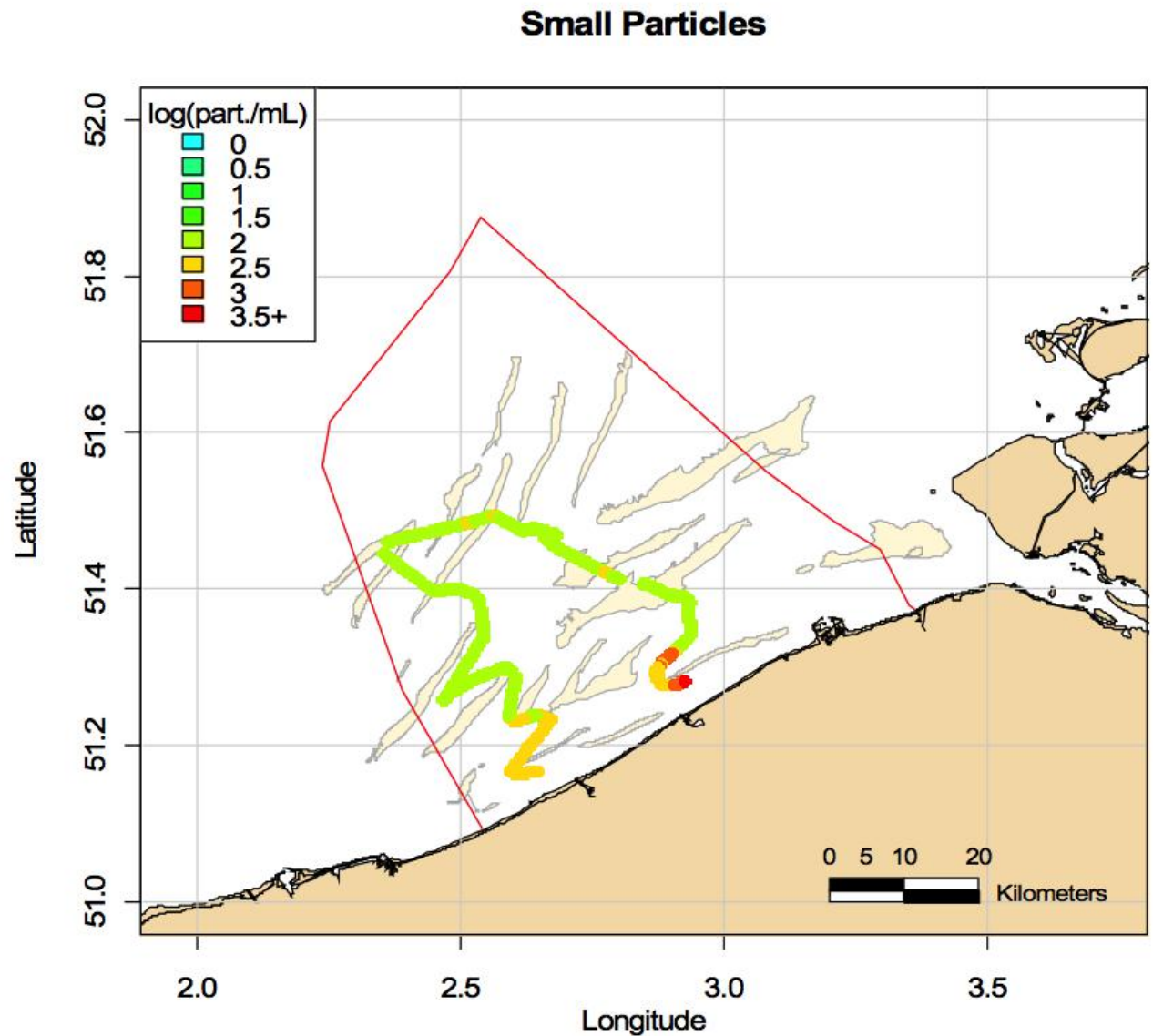
A neritic
diatoms :
*Chaetoceros sp.*



**Chaetoceros sp**
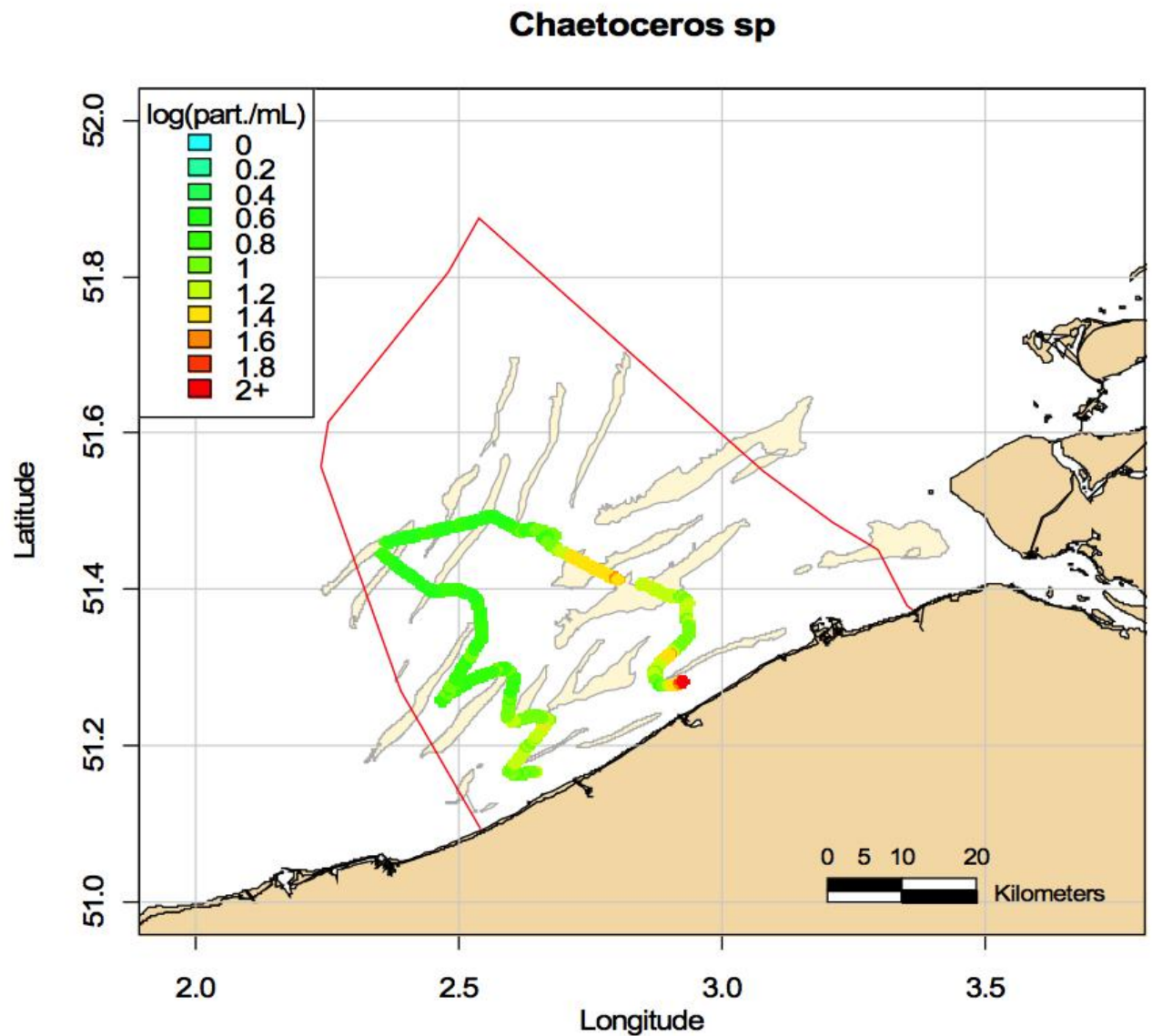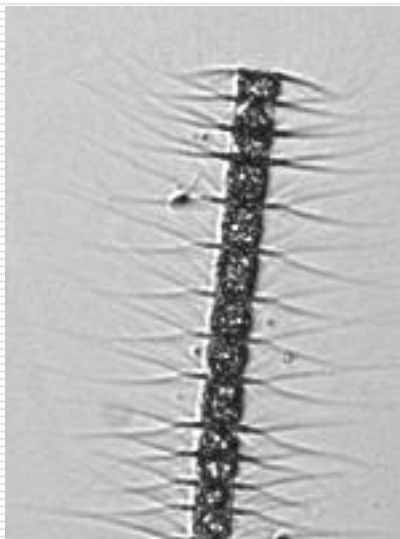
*Rhizosolenia shrubsolei* with a very different distribution pattern.



### R. shrubsolei
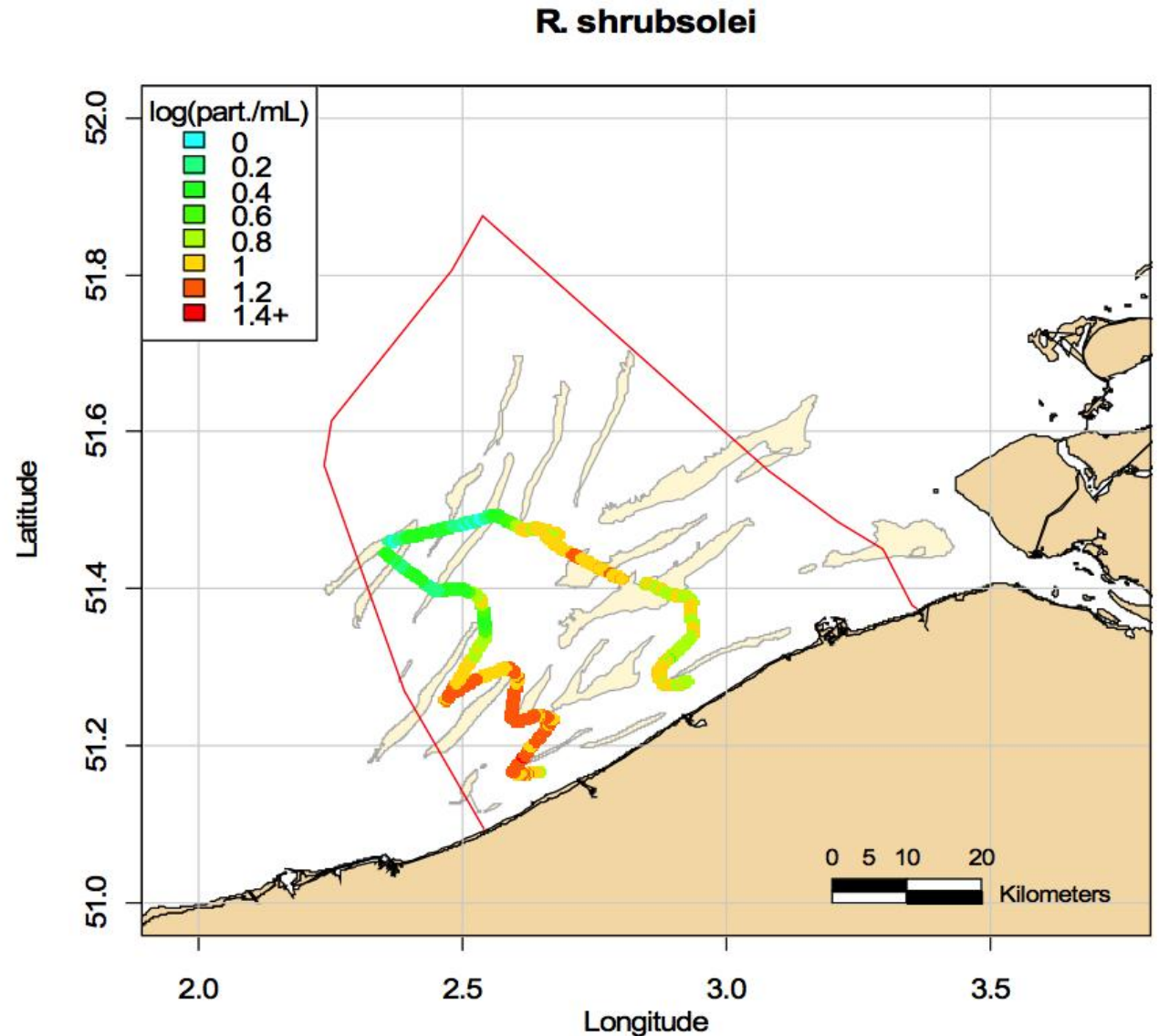
Yet another distribution pattern : *Pseudo-nitzschia*.



**Pseudo-Nitzschia**
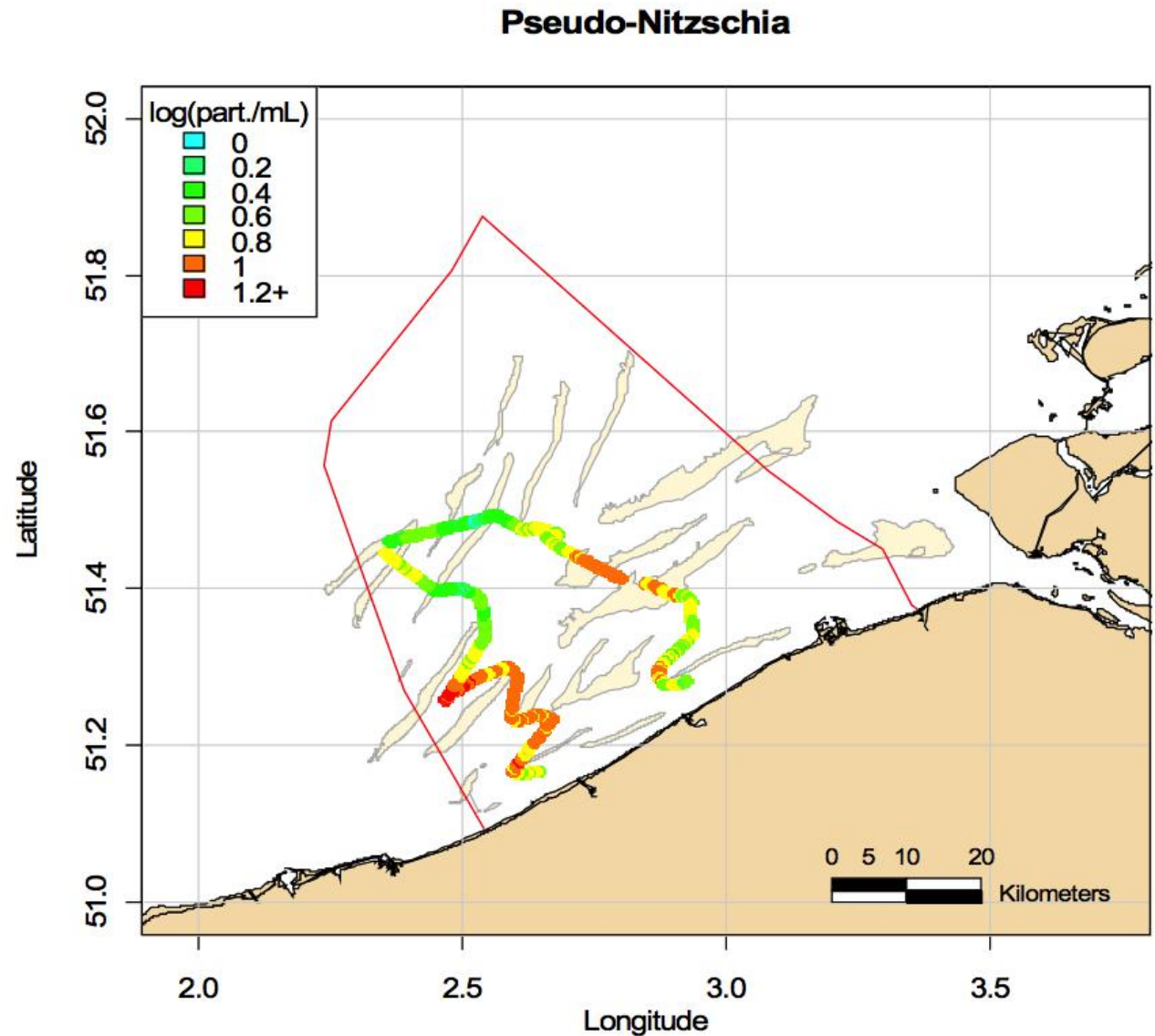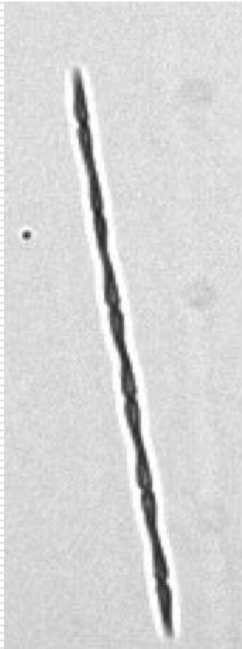
What about a higher taxonomic discrimination ?

Challenges with the IFREMER Rephy network

# IFREMER Rephy

**Lieux de surveillance actifs REPHY**
Inventaire cartographique

rephy

- French coasts

- All the dots on the map are stations included in the survey

- Starting to use the FlowCAM + Zoo/PhytoImage v.5 since mid 2014

(map from IFREMER)

Copyright © Ifremer
IGN,BD Carthage,OIEau,SHOM,

100 km
100 mi

# Typical results with > 40 classes

**Global error :  26 %**

**Error per class:**

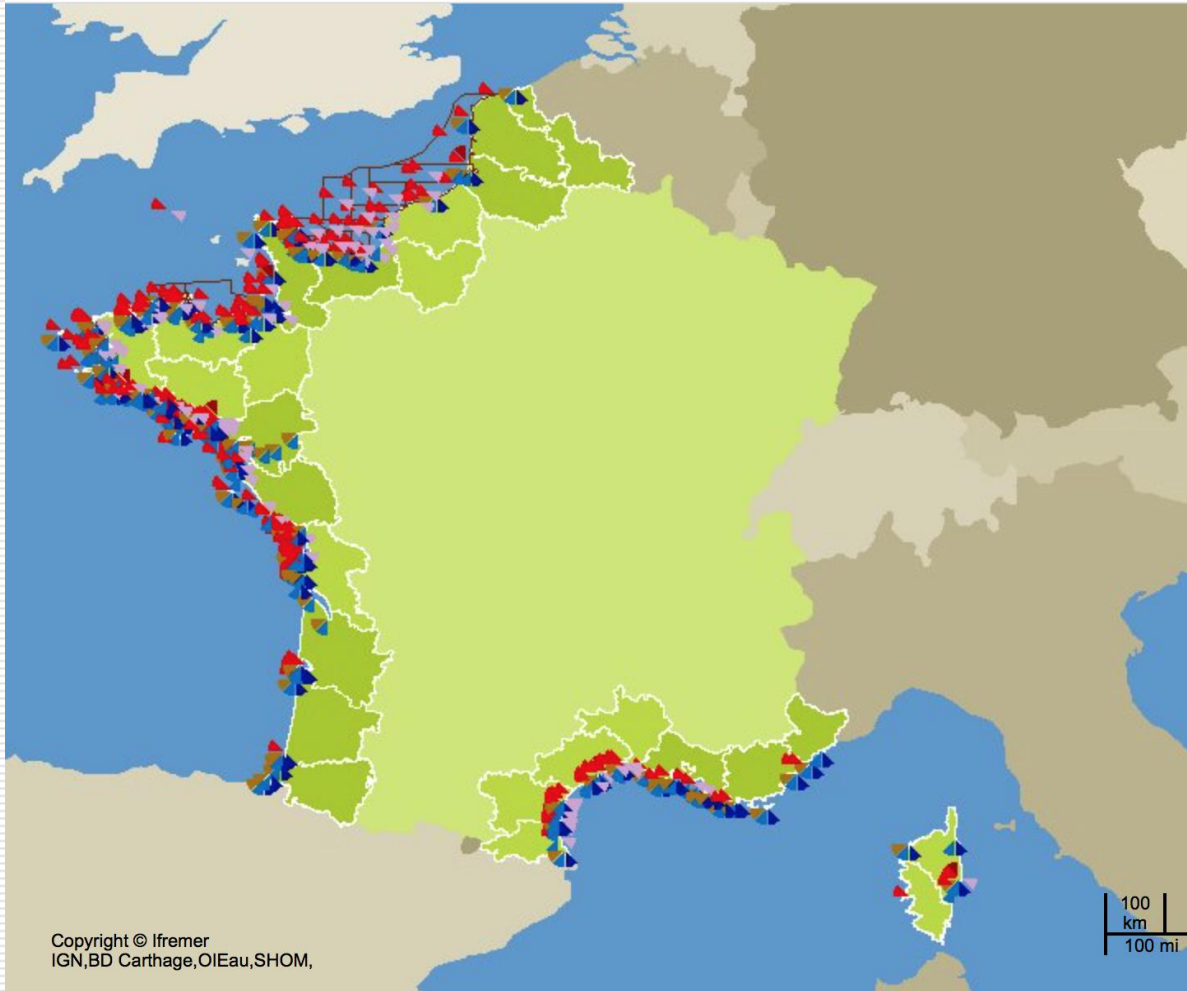| | Error (%) | | Error (%) |
|---|---|---|---|
| Drop | 7 | Lauderia_Schroederella | 24 |
| Pollen | 7 | Thalassiosira_spp | 25 |
| C.fusus | 8 | Fecal_pellets | 25 |
| L.danicus | 9 | Long_thin | 26 |
| L.undulatum | 12 | Dark | 32 |
| Thalassiosira_spp_cells | 13 | D.fragilissimus | 33 |
| G.flaccida | 14 | D.brightwelli | 35 |
| Black_opaque | 15 | Aggregates | 36 |
| D.tripos | 17 | P.alata_indica | 40 |
| Centrales_spp | 17 | G.delicatula | 46 |
| Mues | 18 | Membranous | 48 |
| G.striata | 18 | R.imbricata_styli | 50 |
| Pseudo-nitzschia | 19 | Fibers | 50 |
| Euglenophyceae | 20 | Dictyochophyceae | 53 |
| T.subtilis | 20 | Protoperidinum_spp | 68 |
| Short_thin | 22 | Chaetoceros_spp | 69 |
| Clear | 23 | Larvae_crustaceans | 71 |
| N.longissima_Cylindrotheca | 23 | Ceratium_spp | 79 |
| A.glacialis | 23 | P.alata | 93 |
| Bubbles | 23 | C.danicus | 95 |
| Granular | 24 | C.decipiens | 97 |

- On average, ¼ wrongly classified

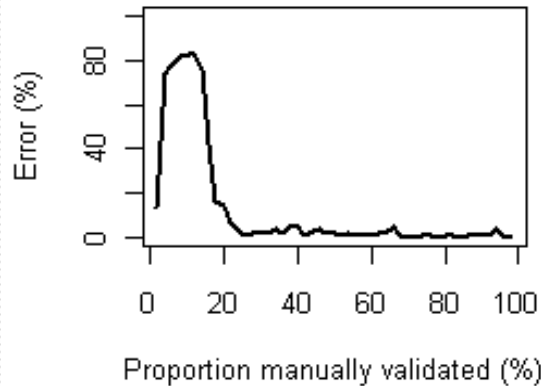- Large discrepancy of results: some classes are OK, other are completely wrong

**Manual validation is required to lower the error down to acceptable levels for all classes**
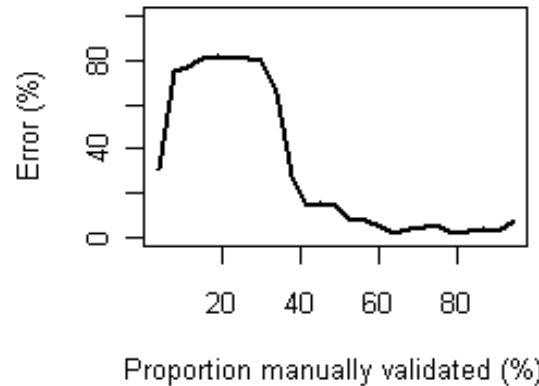
(from Tunin-Ley et al, 2011)
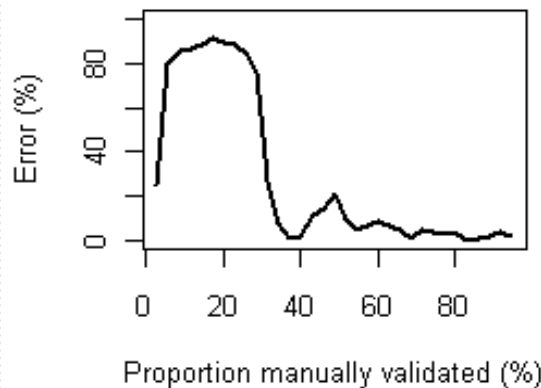
# Detection of suspect particles
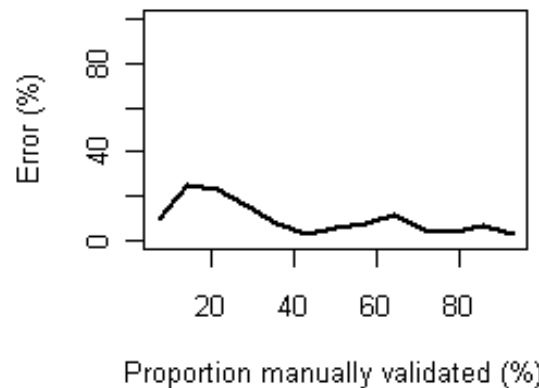


W06 : Error in manual validation

W07 : Error in manual validation

W08 : Error in manual validation

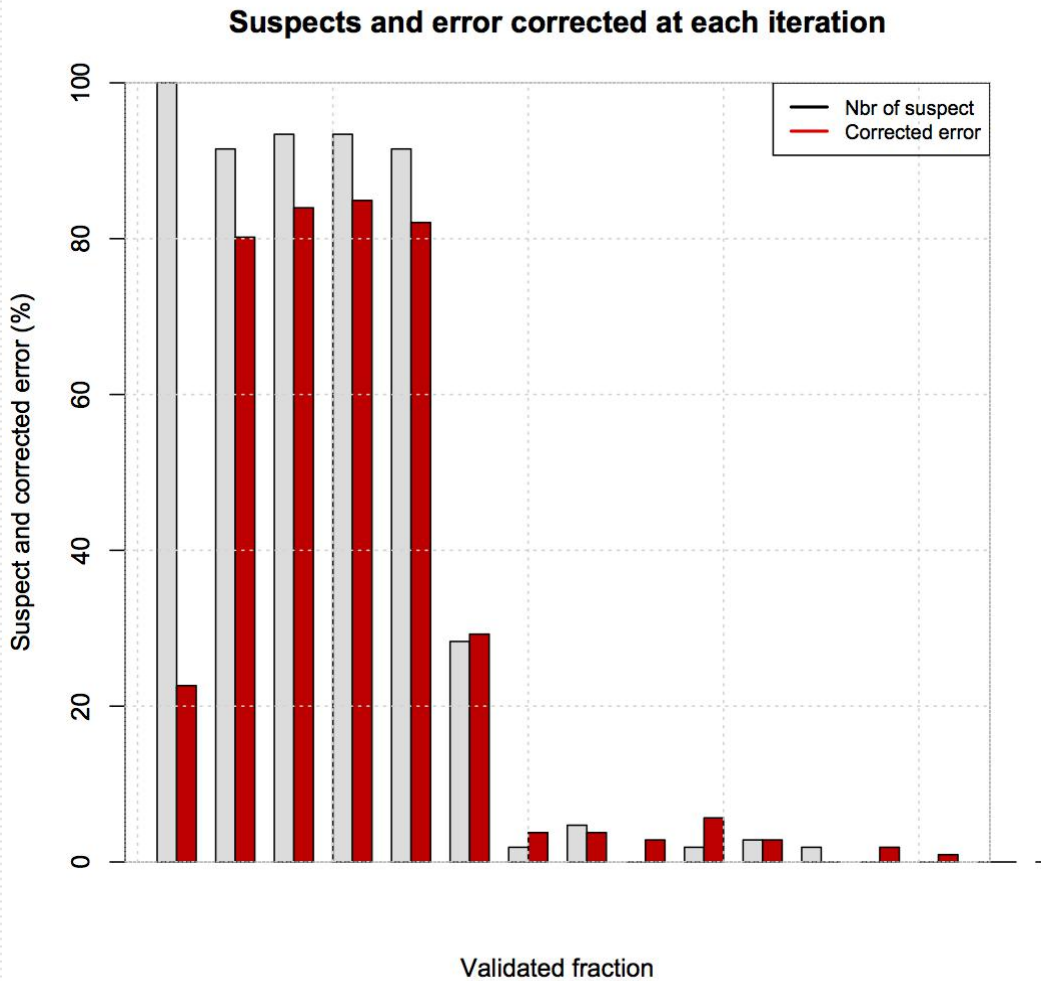T34 : Error in manual validation

In addition to the classification, is it possible to estimate a probability of misclassification ?

**Yes : suspect detection**
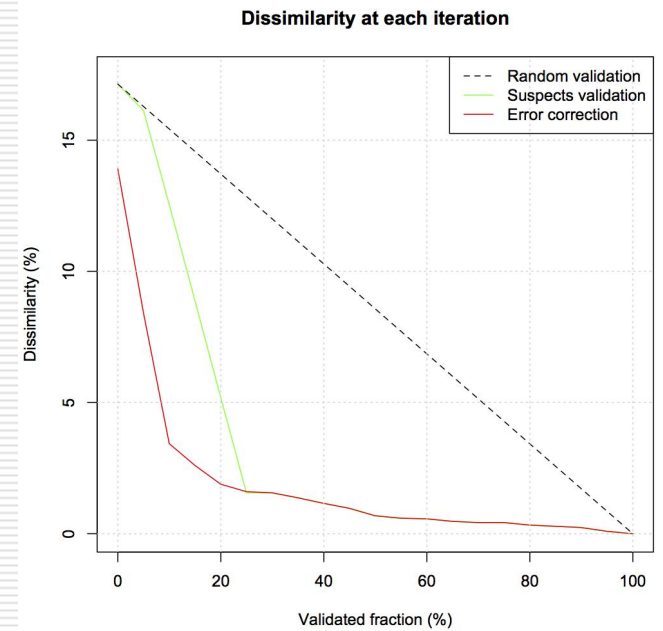
Application on four different samples

**Suspect particles contain a large fraction of the error in most cases**

Suspects and error corrected at each iteration

Gray bars = suspects

Red bars = error corrected after manual validation
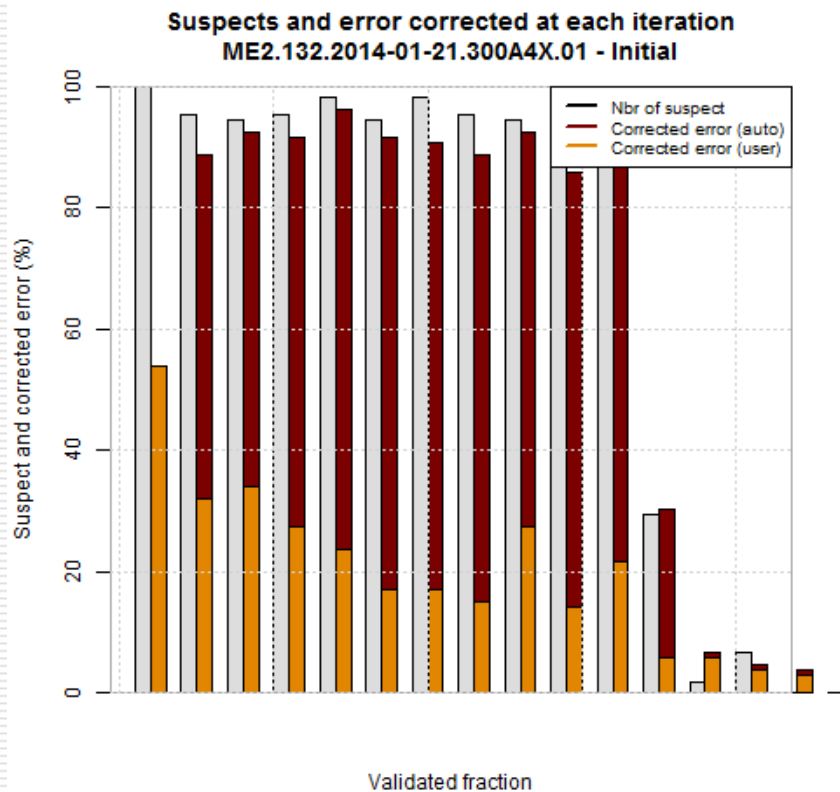


Dissimilarity at each iteration

# Active learning and statistical error correction
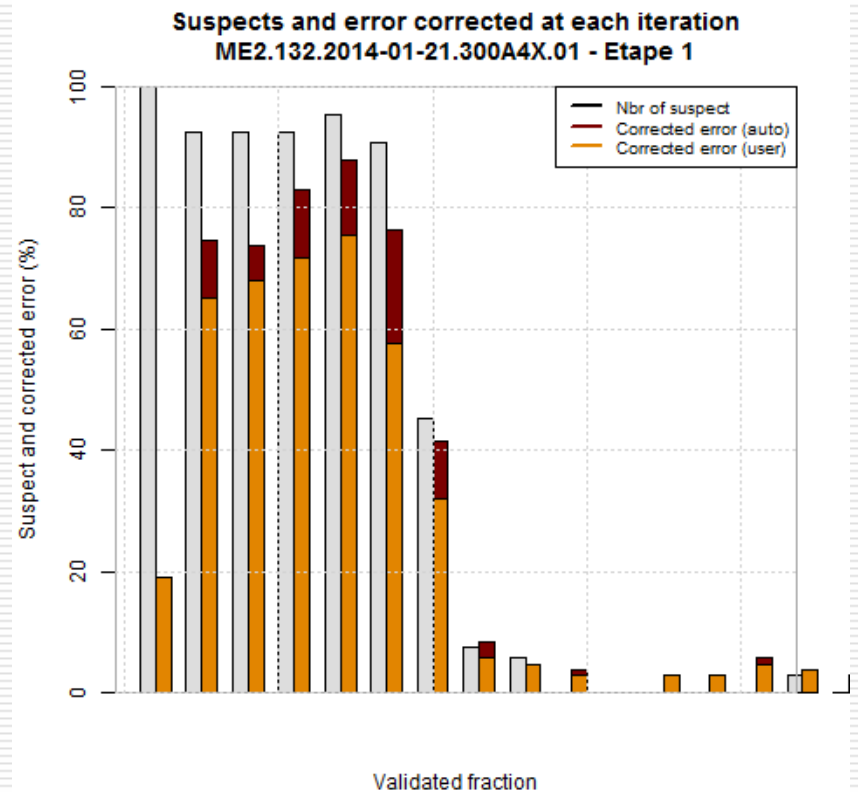
Brown bars = error statistically corrected

Orange bars = error that the user must correct

Without adaptative training set

With adaptative training set

# Conclusions & perspectives

• Image analysis (FlowCAM) combined with supervised analysis usable in real-time for coarse classification of phytoplankton

• Deployment in routine survey network with higher discrimination needs manual validation to lower the error (e.g., IFREMER - Rephy)

• Validation of suspects combined with statistical error correction greatly reduces the number of items to validate for a given error rate
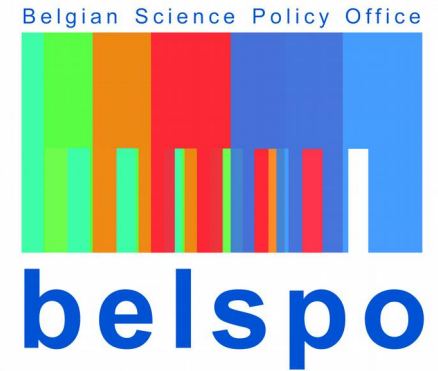
Study in collaboration with *IFREMER*, also funded by the *Belgian Science Policy*