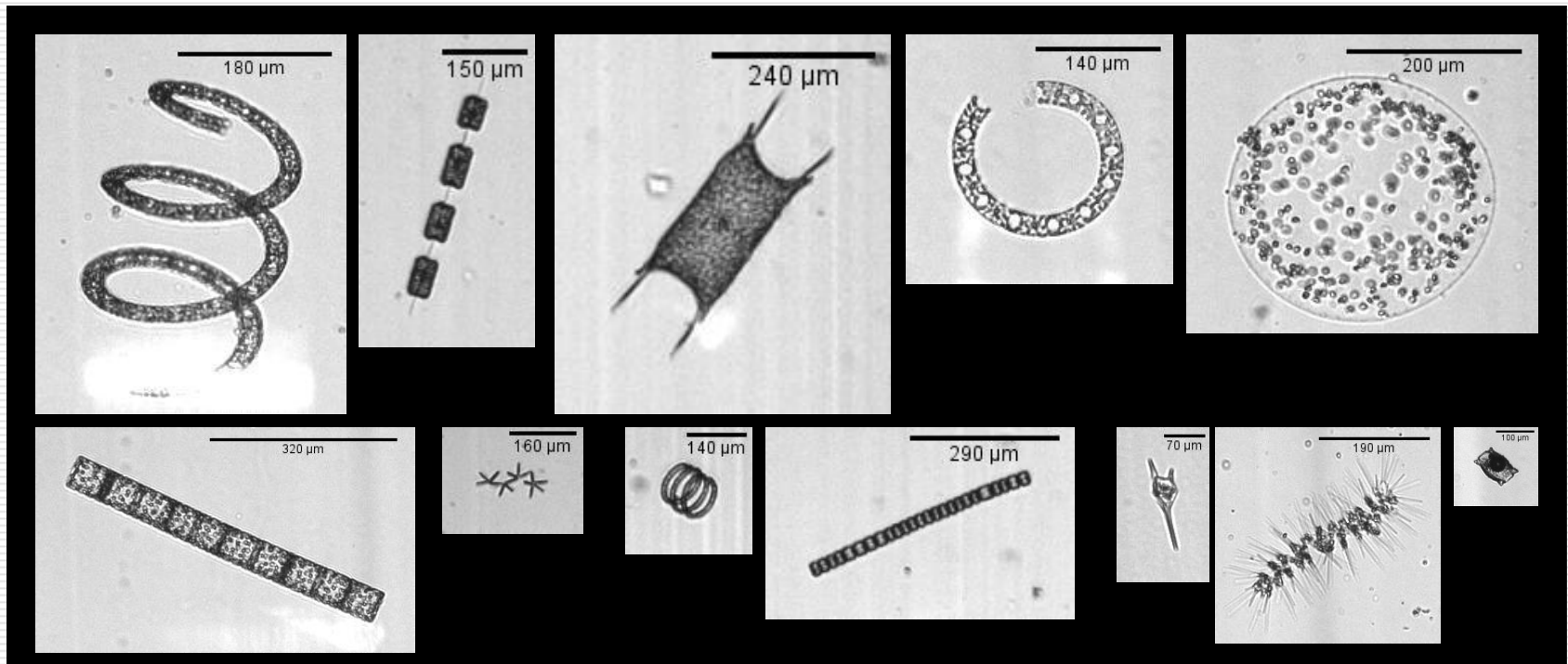
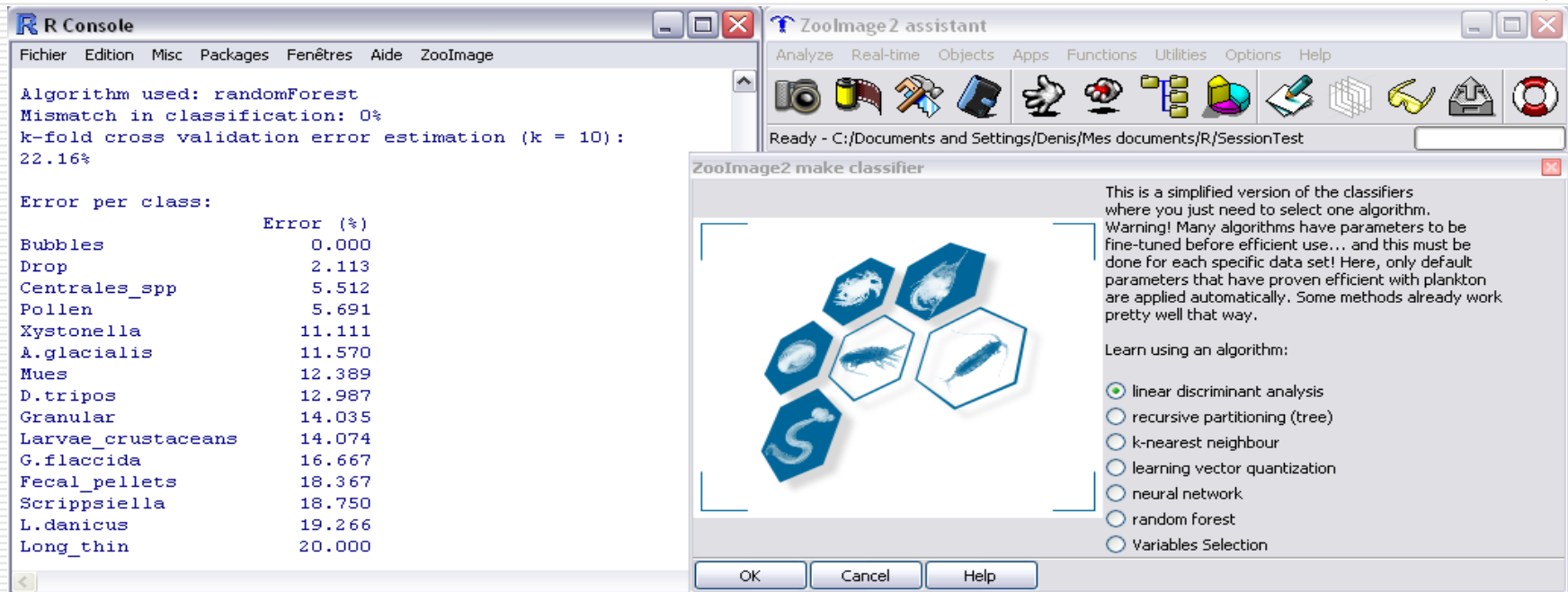


Towards a better classification of plankton digital images : suspect detection, error correction and real-time



Grosjean Ph., K. Denis, G. Wacquet, N. Ali, V. Rousseau, J.-Y. Parent,
Ch. Lancelot, D. Hamad, F. Artigas, A. Lefebvre, C. Belin

What is Zoo/PhytoImage ?



The screenshot displays two windows from the ZooImage2 software. The 'R Console' window shows the results of a classification using the randomForest algorithm. The mismatch in classification is 0%, and the k-fold cross validation error estimation (k = 10) is 22.16%. A table lists the error percentage for various classes.

Error per class:	Error (%)
Bubbles	0.000
Drop	2.113
Centrales_spp	5.512
Pollen	5.691
Xystonella	11.111
A.glacialis	11.570
Mues	12.389
D.tripos	12.987
Granular	14.035
Larvae_crustaceans	14.074
G.flaccida	16.667
Fecal_pellets	18.367
Scrippsiella	18.750
L.danicus	19.266
Long_thin	20.000

The 'ZooImage2 assistant' window shows a 'ZooImage2 make classifier' dialog box. It contains a preview of a plankton image and a list of algorithms to choose from. The 'linear discriminant analysis' option is selected.

This is a simplified version of the classifiers where you just need to select one algorithm. Warning! Many algorithms have parameters to be fine-tuned before efficient use... and this must be done for each specific data set! Here, only default parameters that have proven efficient with plankton are applied automatically. Some methods already work pretty well that way.

Learn using an algorithm:

- linear discriminant analysis
- recursive partitioning (tree)
- k-nearest neighbour
- learning vector quantization
- neural network
- random forest
- Variables Selection

- ✓ 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")
```

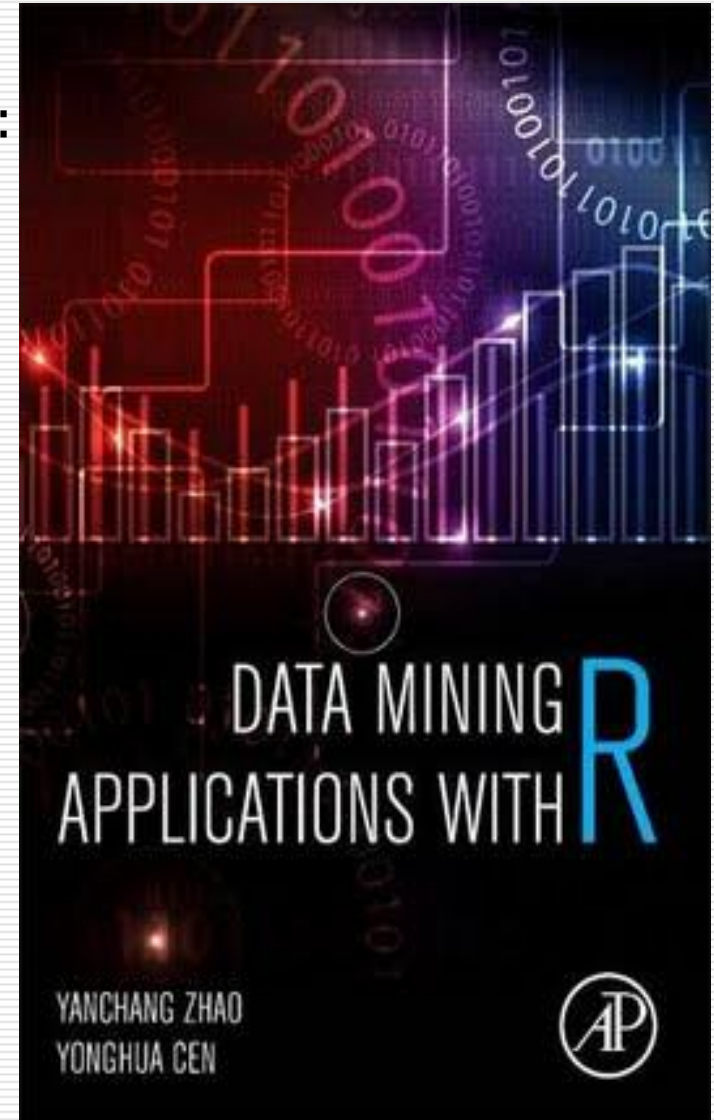
```
> library(zooimage)
```

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.

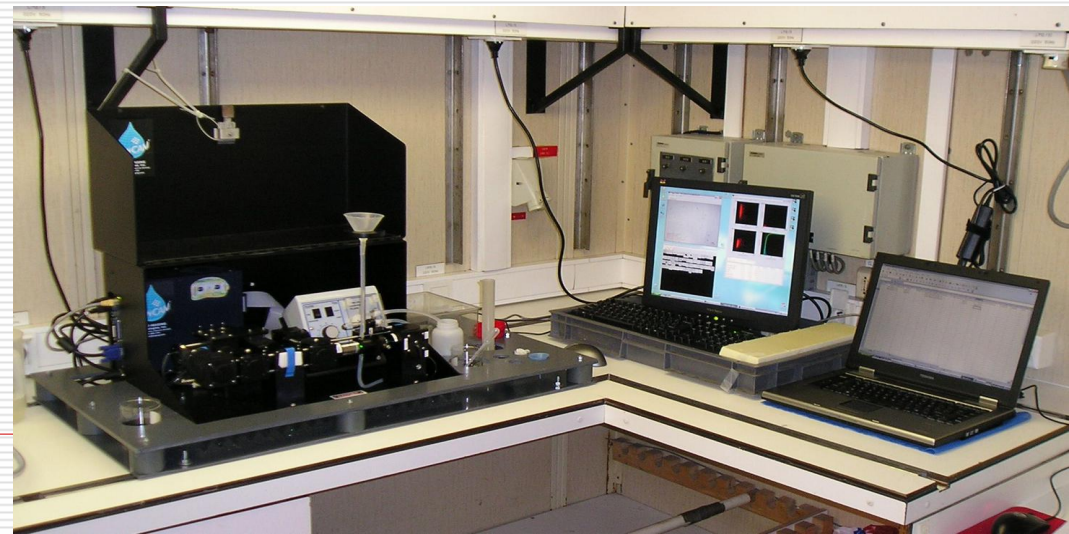
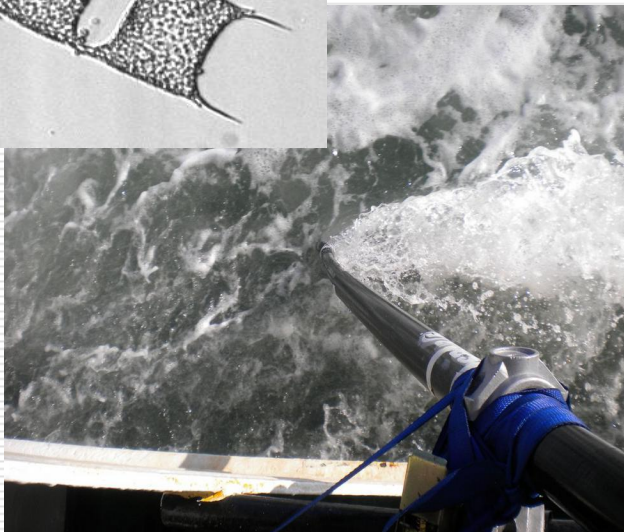
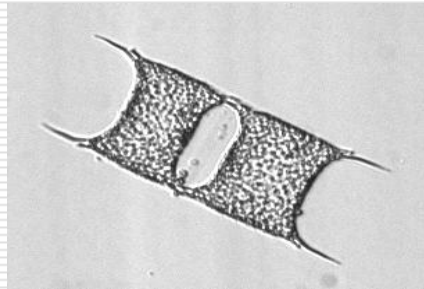
Version 5 near the end of this year



Application in real-time

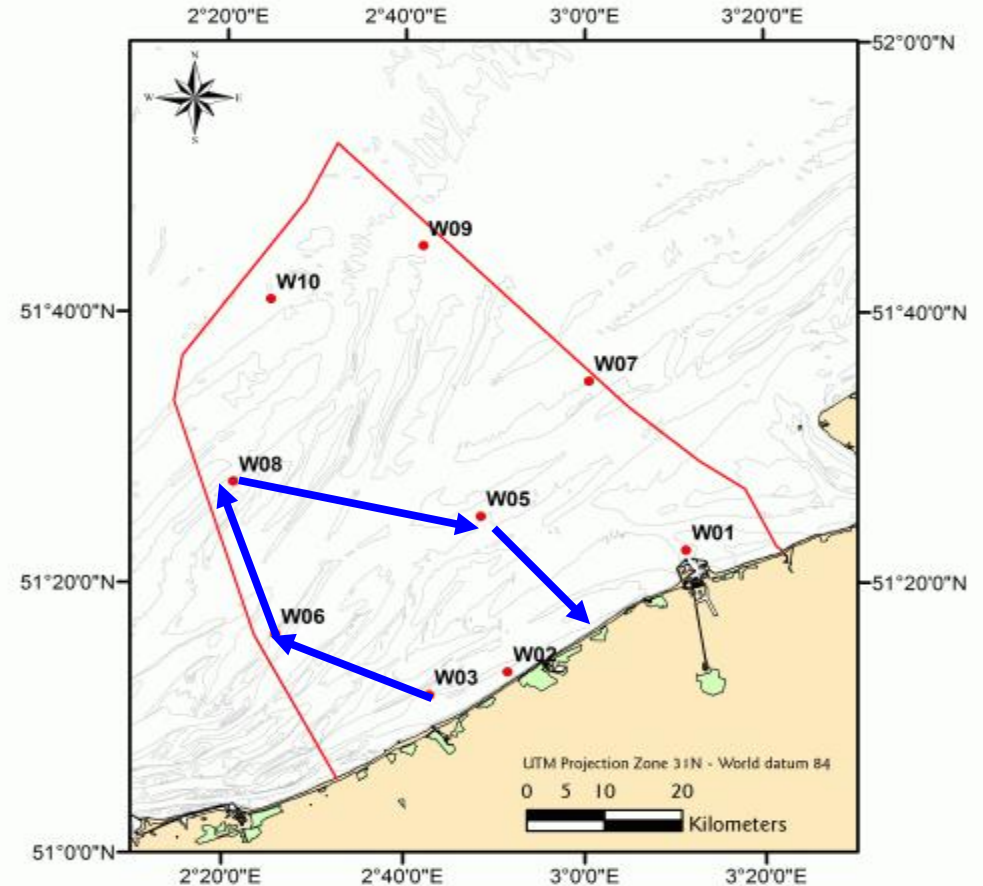


- Combination of the FlowCAM with Zoo/PhytoImage
- Aboard the 'Belgica' (Belgian oceanographic ship)
- 25 groups discriminated in real-time (incl. 18 phytoplankton groups)
- Done a couple of years ago... but no more money for further work since then



Real-time monitoring of North Sea phytoplankton

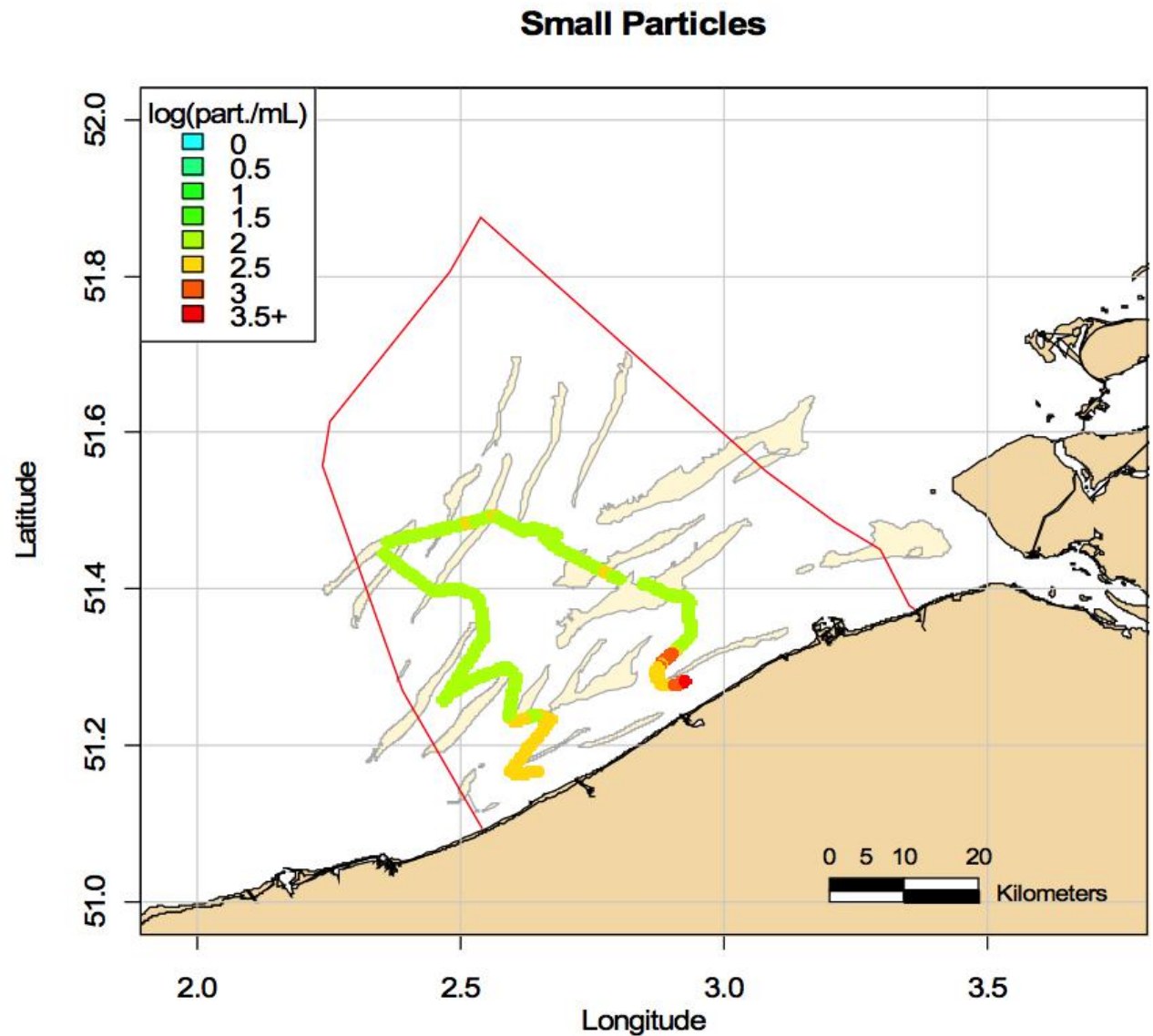
Application example: a one-day cruise in the Belgian Coastal Zone (BCZ).



Real-time monitoring - results

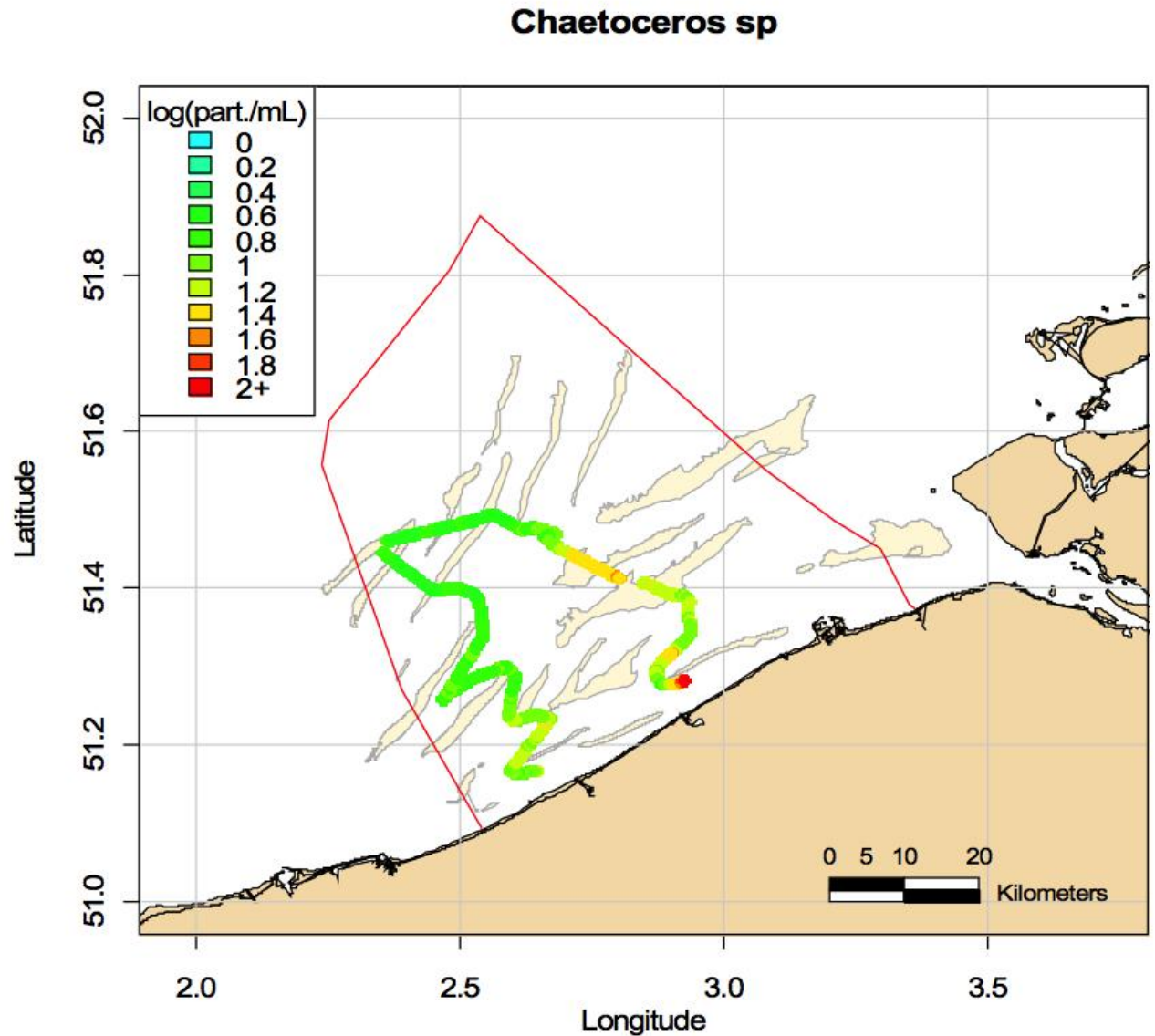
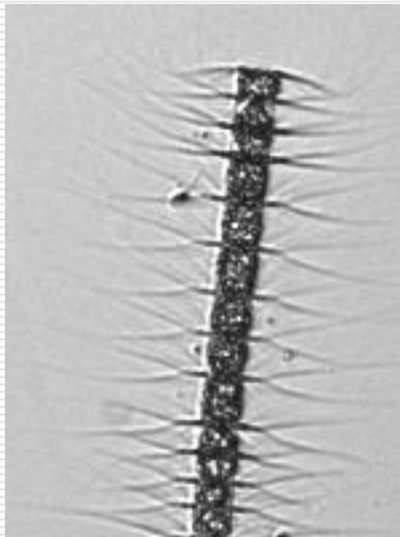
Small particles dominating near the coast and the Schelde estuary.

Notice the influence of sandbanks (no interpolation!)



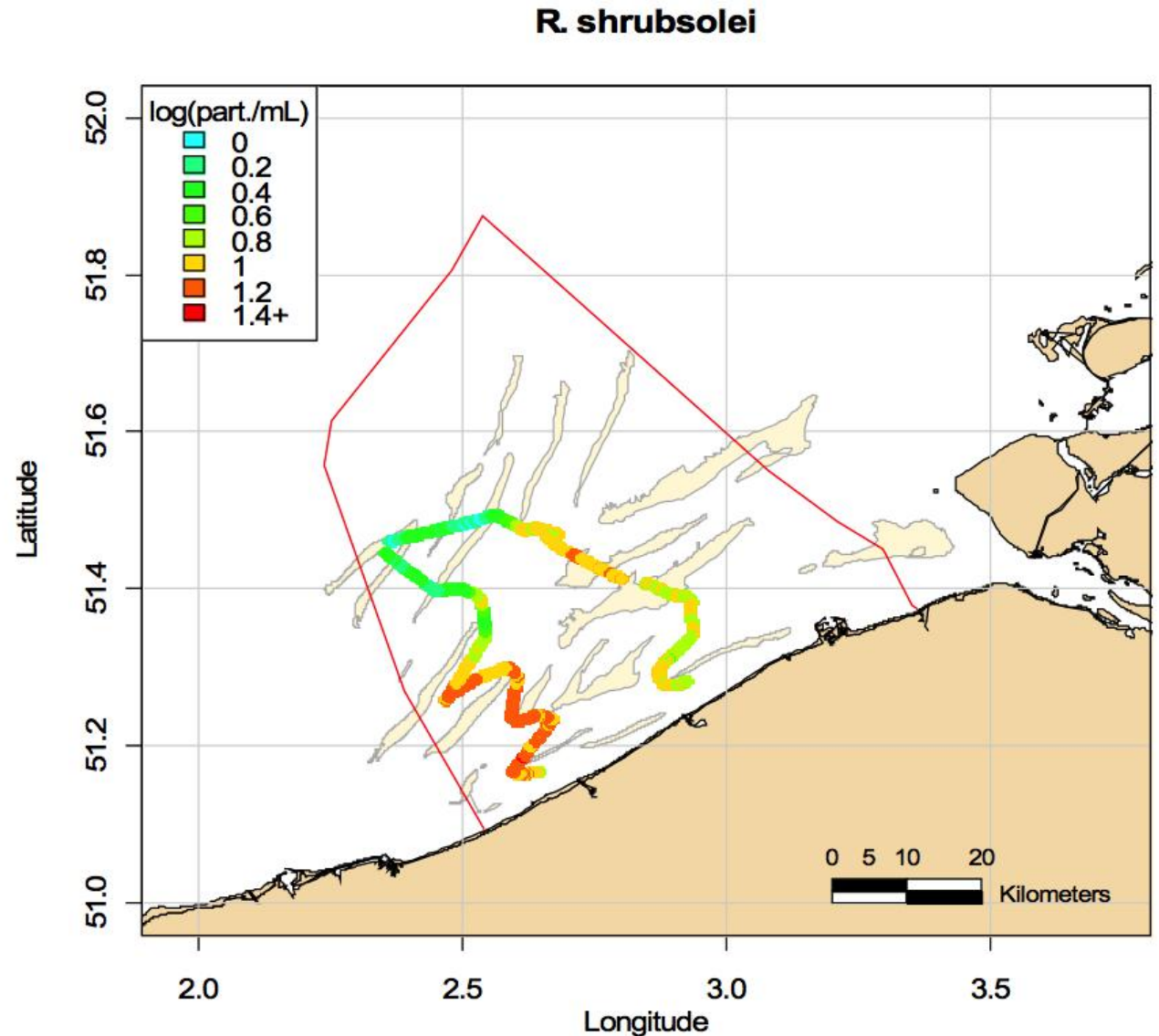
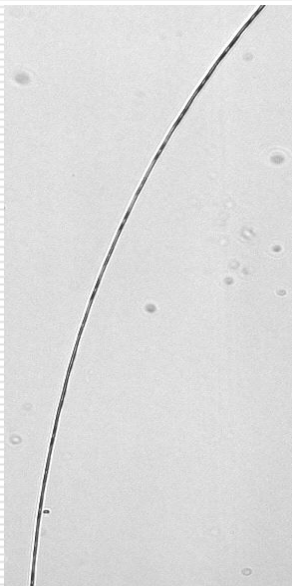
Real-time monitoring - results

A neritic
diatoms :
Chaetoceros sp.



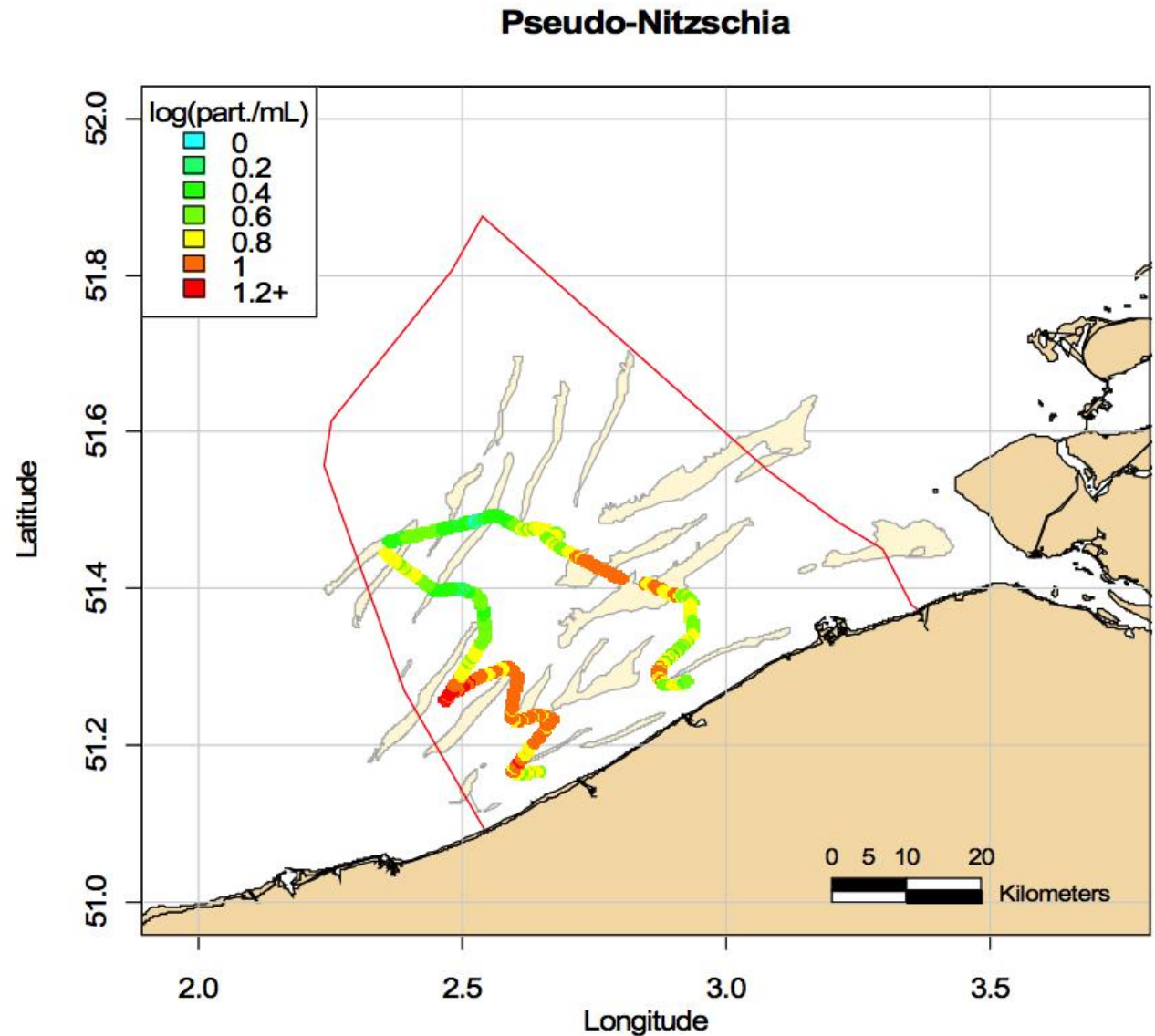
Real-time monitoring - results

*Rhizosolenia
shrubsolei*
with a very
different
distribution
pattern.



Real-time monitoring - results

Yet another
distribution
pattern :
Pseudo-
nitzschia.





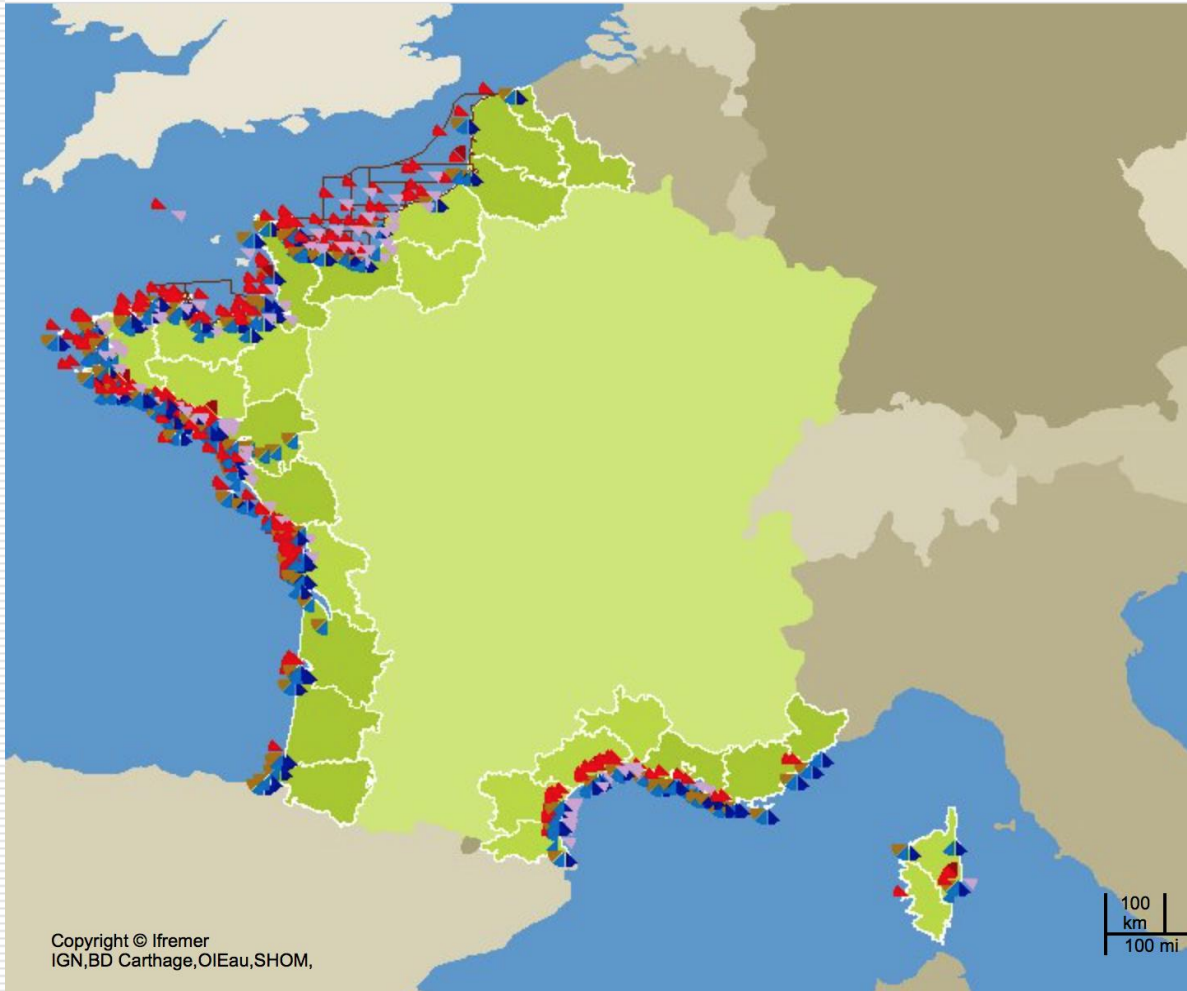
What about a higher taxonomic discrimination ?

Routine use in the IFREMER Rephy network

IFREMER Rephy

Lieux de surveillance actifs REPHY

Inventaire cartographique



- 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)

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	Protopteridinium_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, $\frac{1}{4}$ wrongly classified

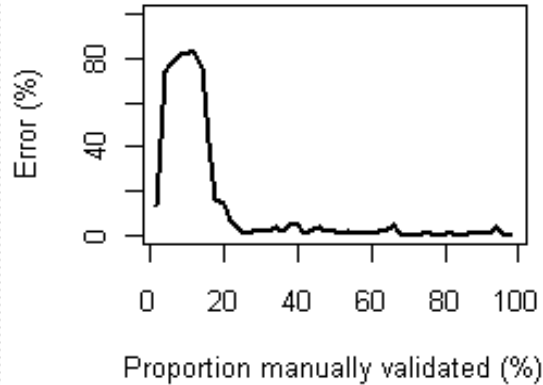
- Large discrepancy of results: some classes are OK, other are completely wrong

Manual validation is required to lower the error down to acceptable levels for all classes

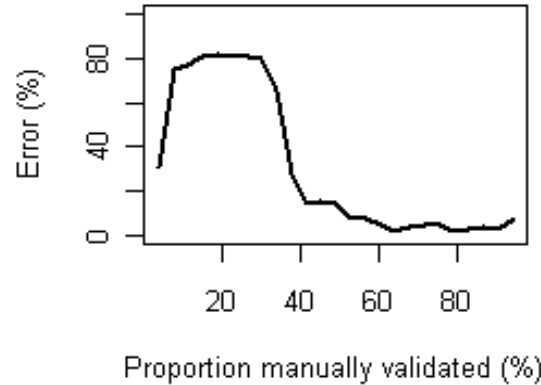
(from Tunin-Ley et al, 2011)

Detection of suspect particles

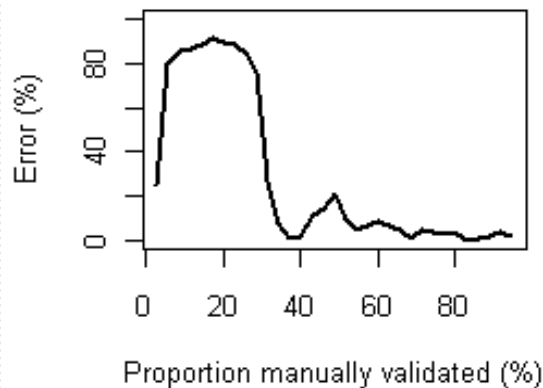
W06 : Error in manual validation



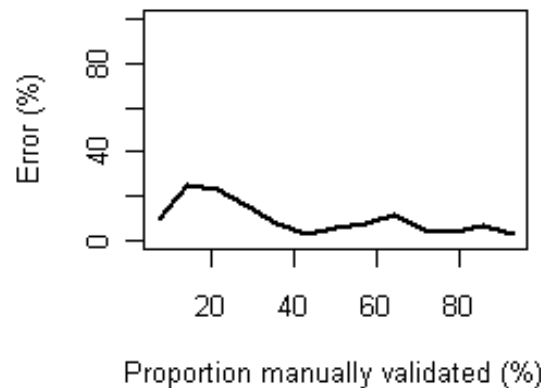
W07 : Error in manual validation



W08 : Error in manual validation



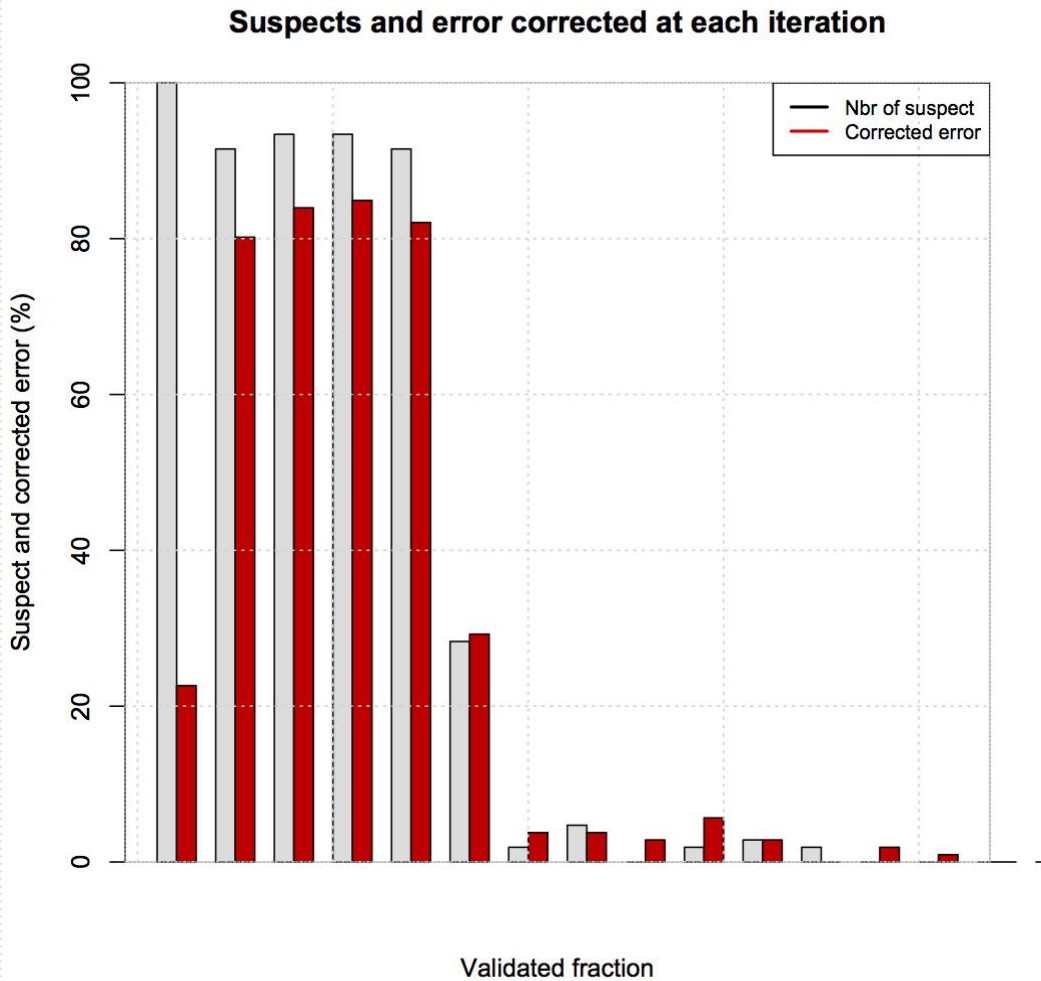
T34 : Error in manual validation



Application on four different samples

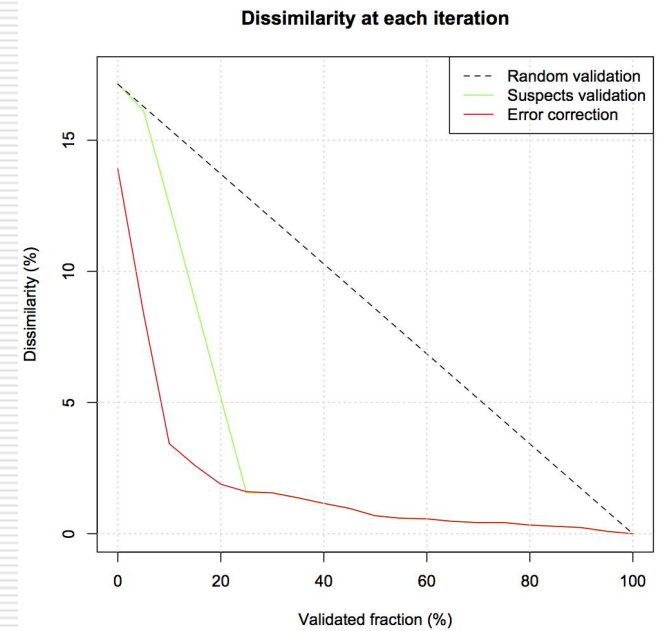
Suspect particles contain a large fraction of the error in most cases

Step by step validation of the suspect items



Gray bars = suspects

Red bars = error corrected after manual validation

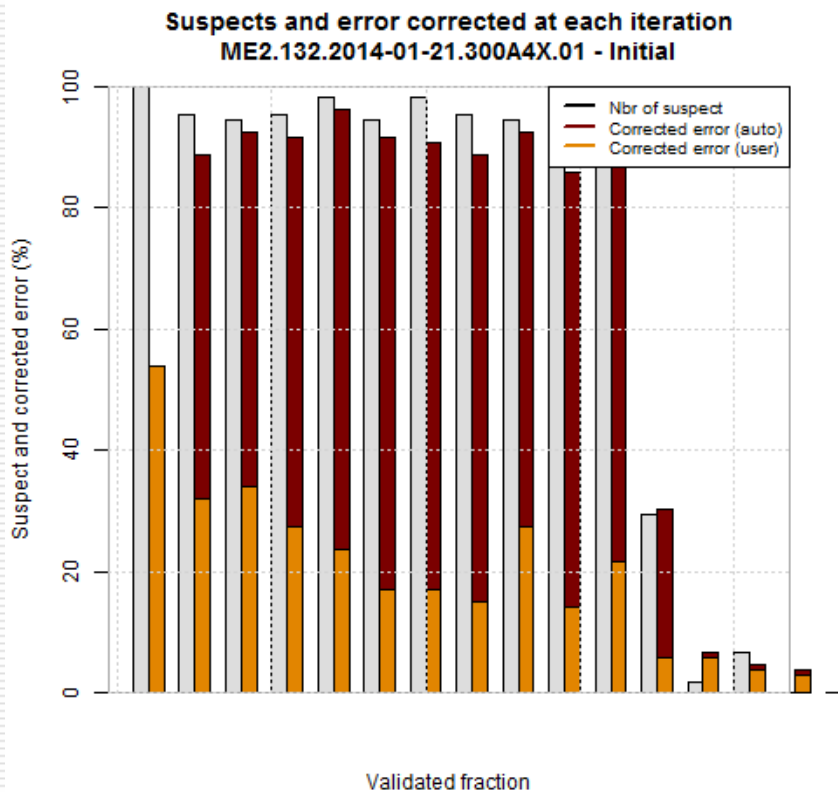


Active learning and statistical error correction

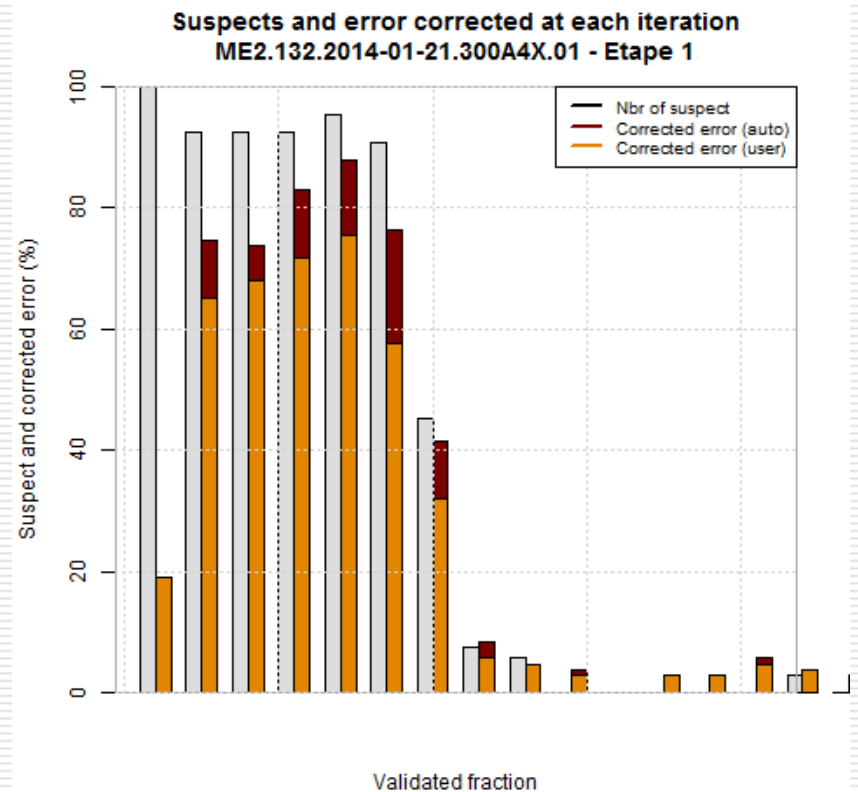
Brown bars = error statistically corrected

Orange bars = error that the user must correct

Without adaptative training set



With adaptative training set



Conclusions & perspectives

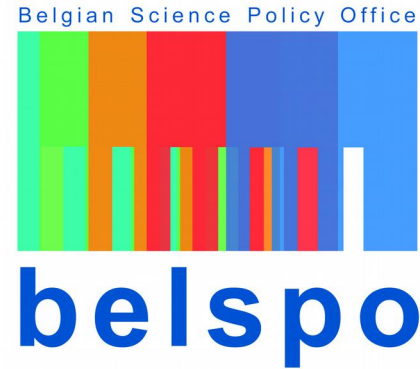
- Image analysis (FlowCAM) combined with supervised analysis usable in real-time for coarse classification of phytoplankton
- Deployment in routine survey network with higher discrimination needs manual validation to lower the error (e.g., IFREMER - Rephy)
- Validation of suspects combined with statistical error correction greatly reduces the number of items to validate for a given error rate



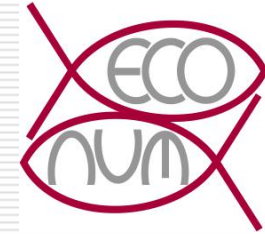
UMONS
Université de Mons



ULB



Belgian Science Policy Office
belspo



Study in collaboration with *IFREMER*, also funded
by the *Belgian Science Policy*



Ifremer