

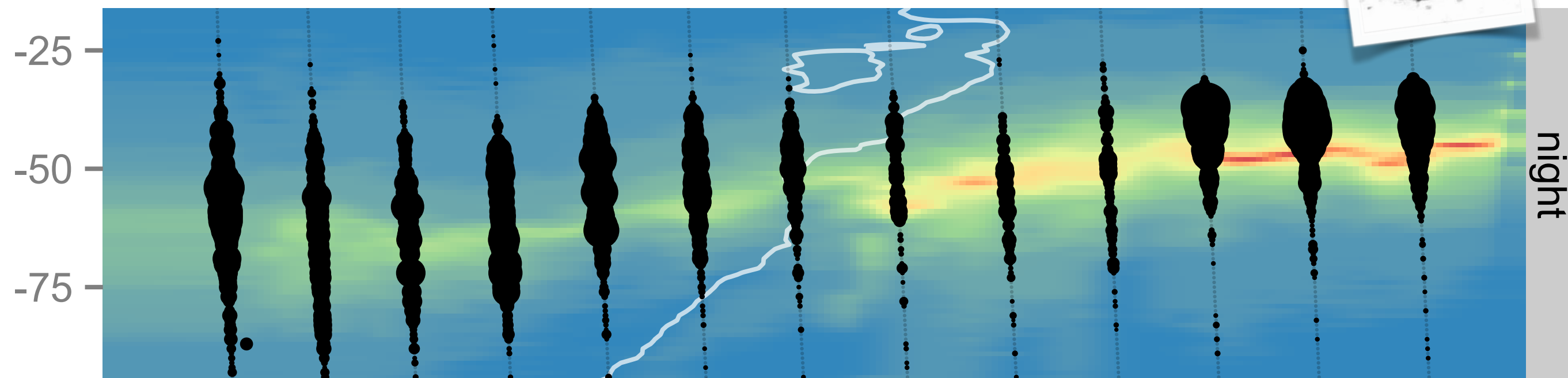
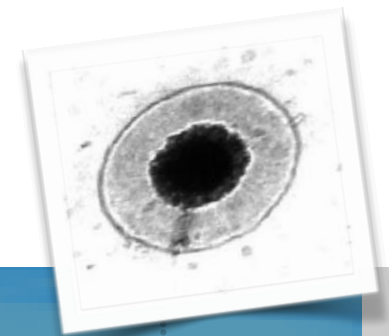
WKMLEARN, 2018-04-18

R Faillettaz, M Picheral, J Y Luo, C Guigand, R K Cowen, J-O Irisson



# Imperfect automatic image classification successfully describes plankton distribution patterns

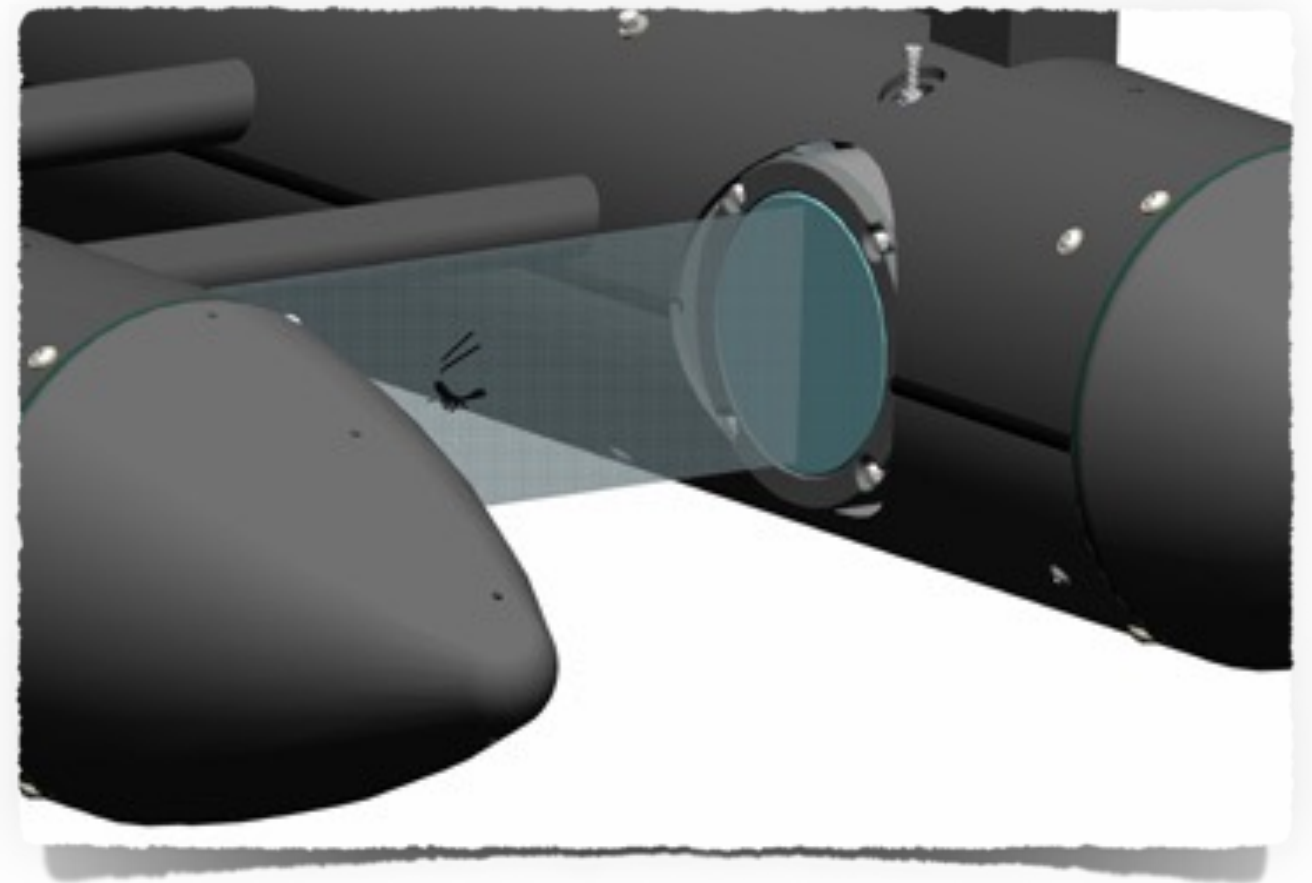
<http://hal.upmc.fr/hal-01324904>



# In Situ Ichthyoplankton Imaging System (ISIIS)



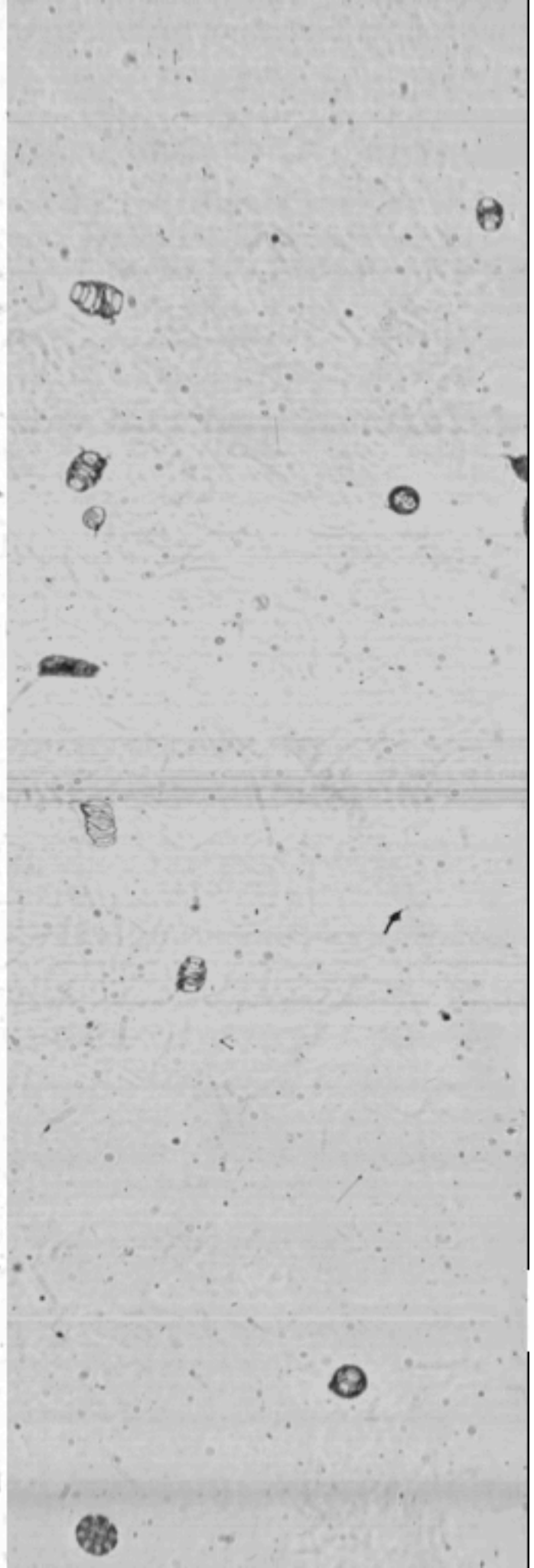
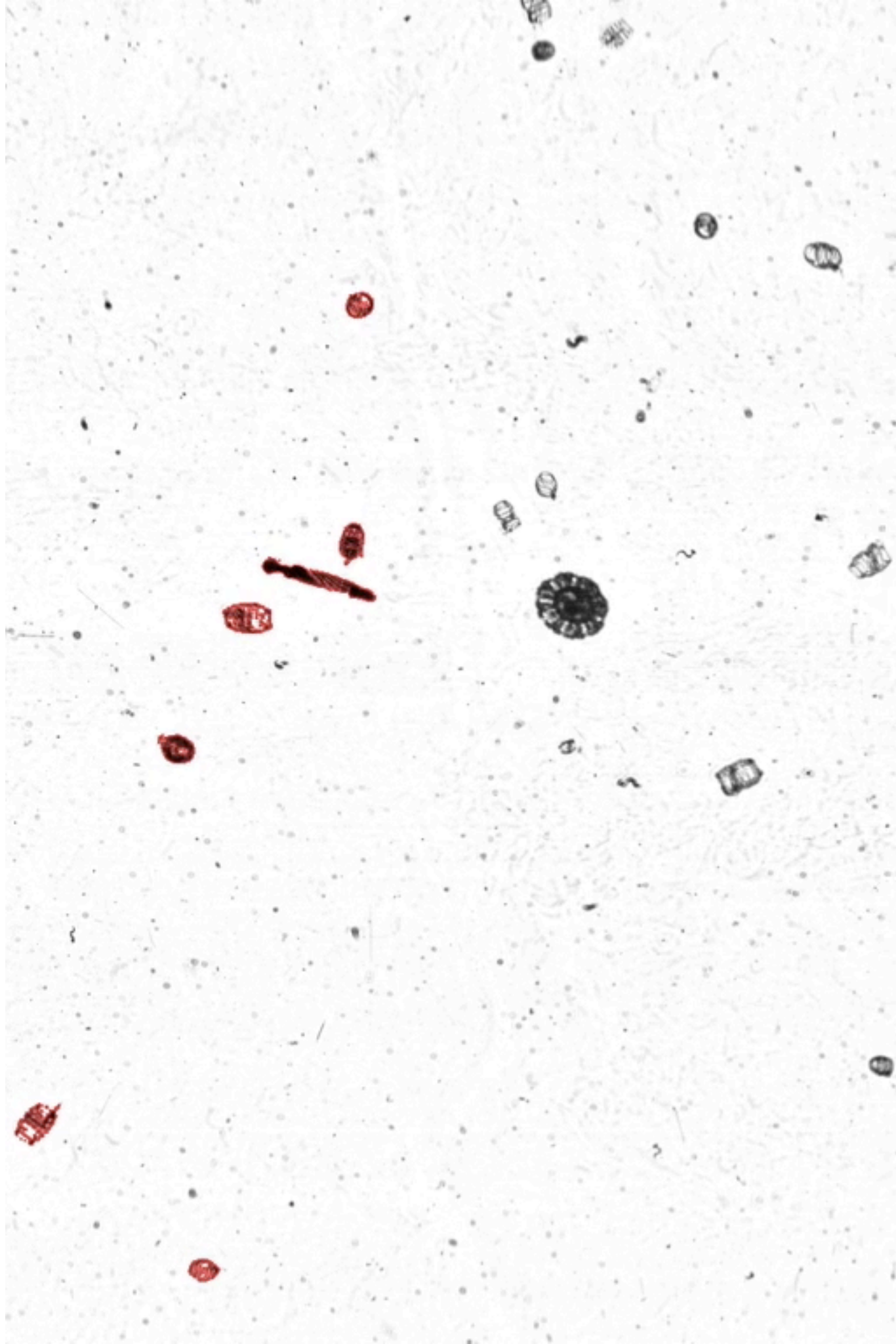
CTD  
Fluorometer  
Oxygen  
ADCP



**BellaMare**

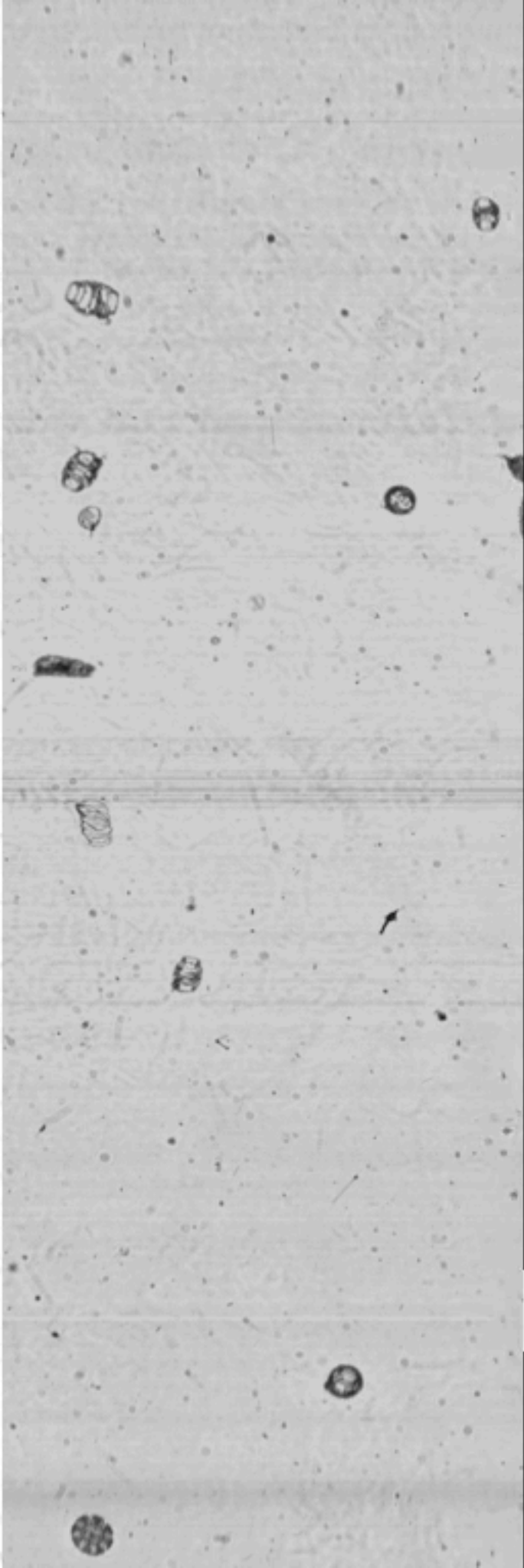
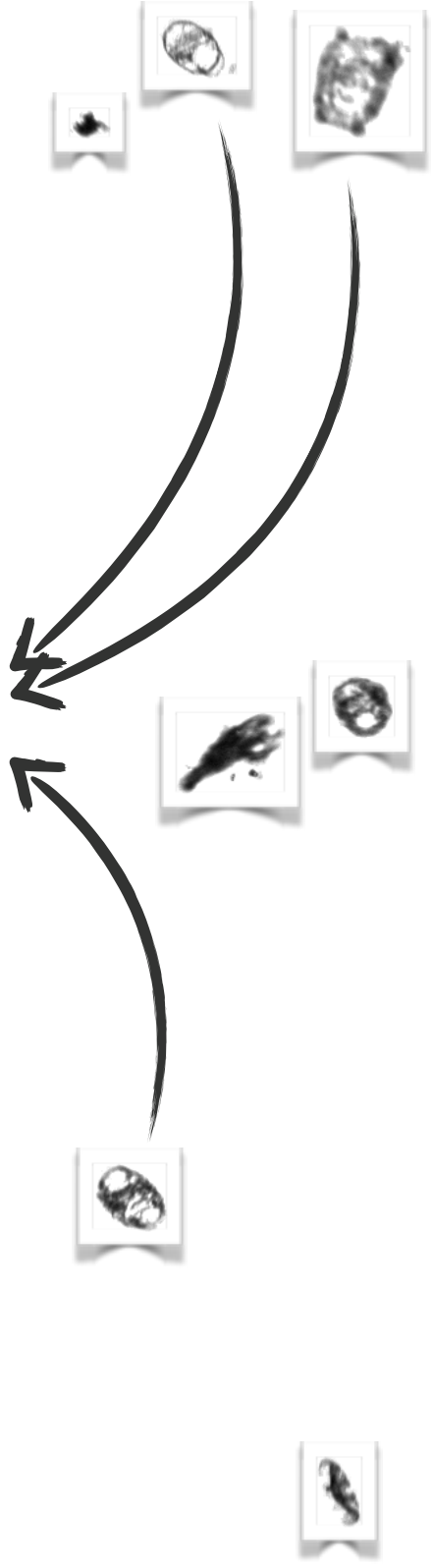
UNIVERSITY OF MIAMI  
ROSENSTIEL  
SCHOOL of MARINE &  
ATMOSPHERIC SCIENCE







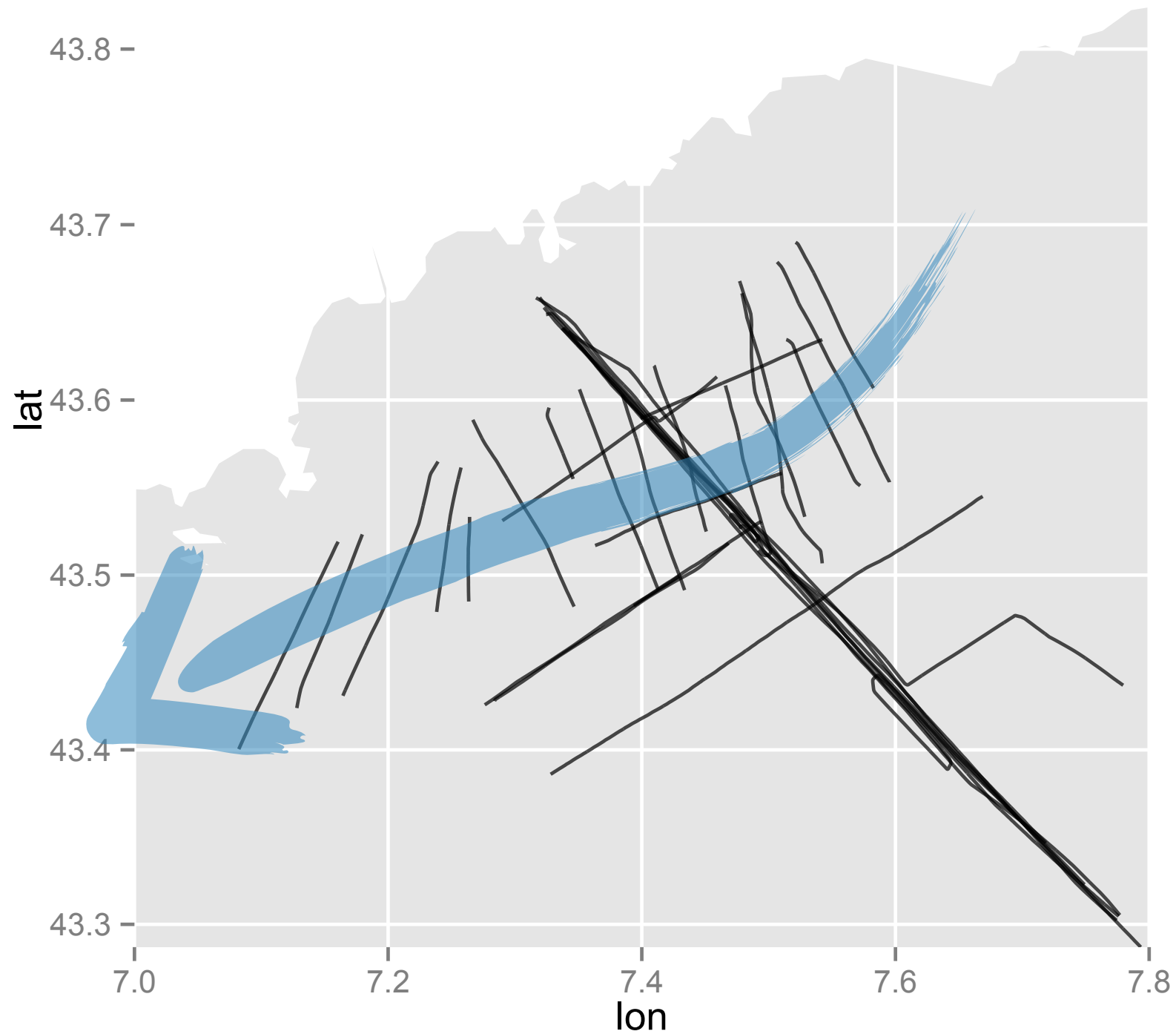
Computer-assisted identification





# Sampling trajectory

---



Describe a mesoscale front

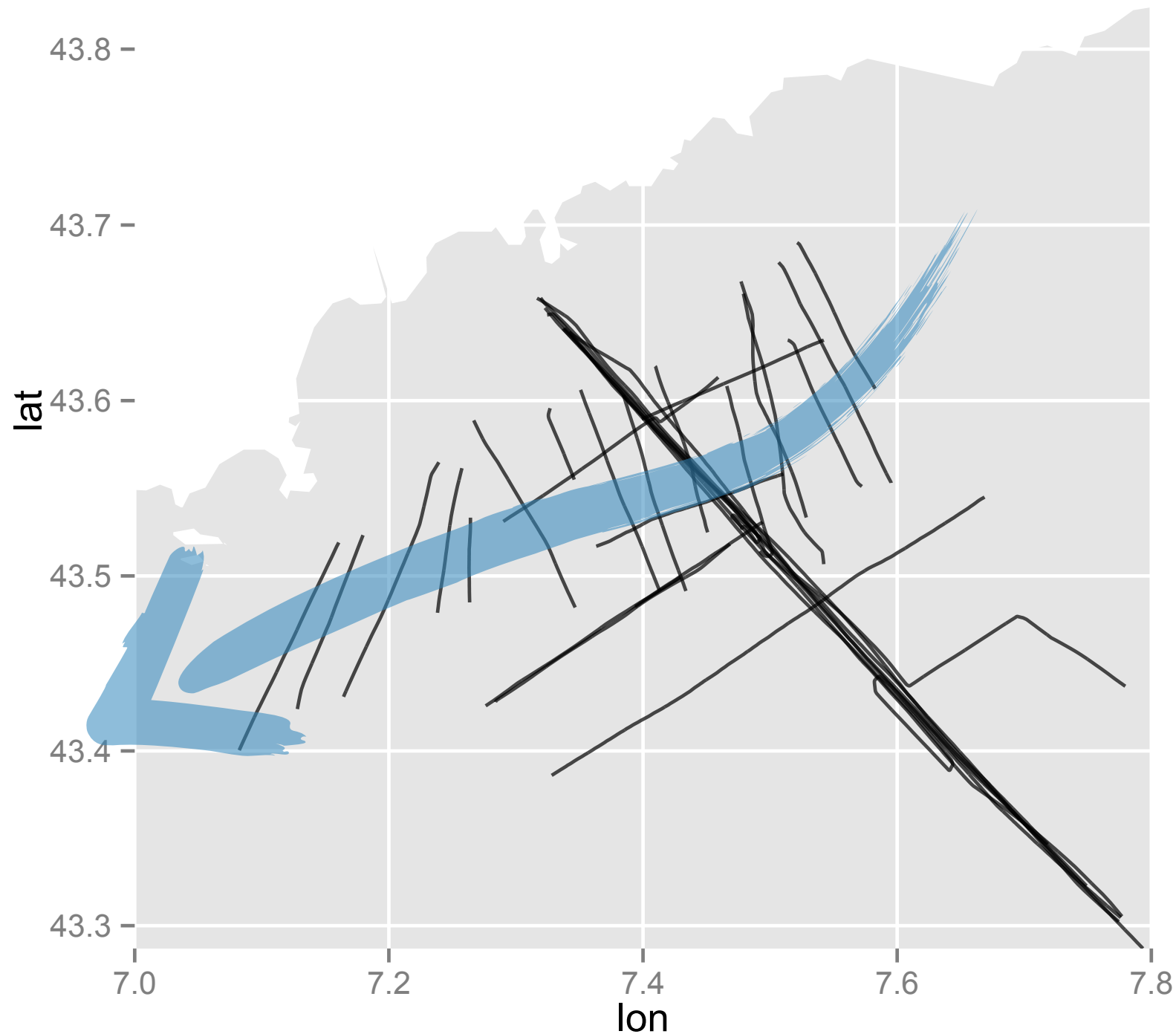
10 days at sea

25TB raw data

14M objects extracted

# Sampling trajectory

---



Describe a mesoscale front

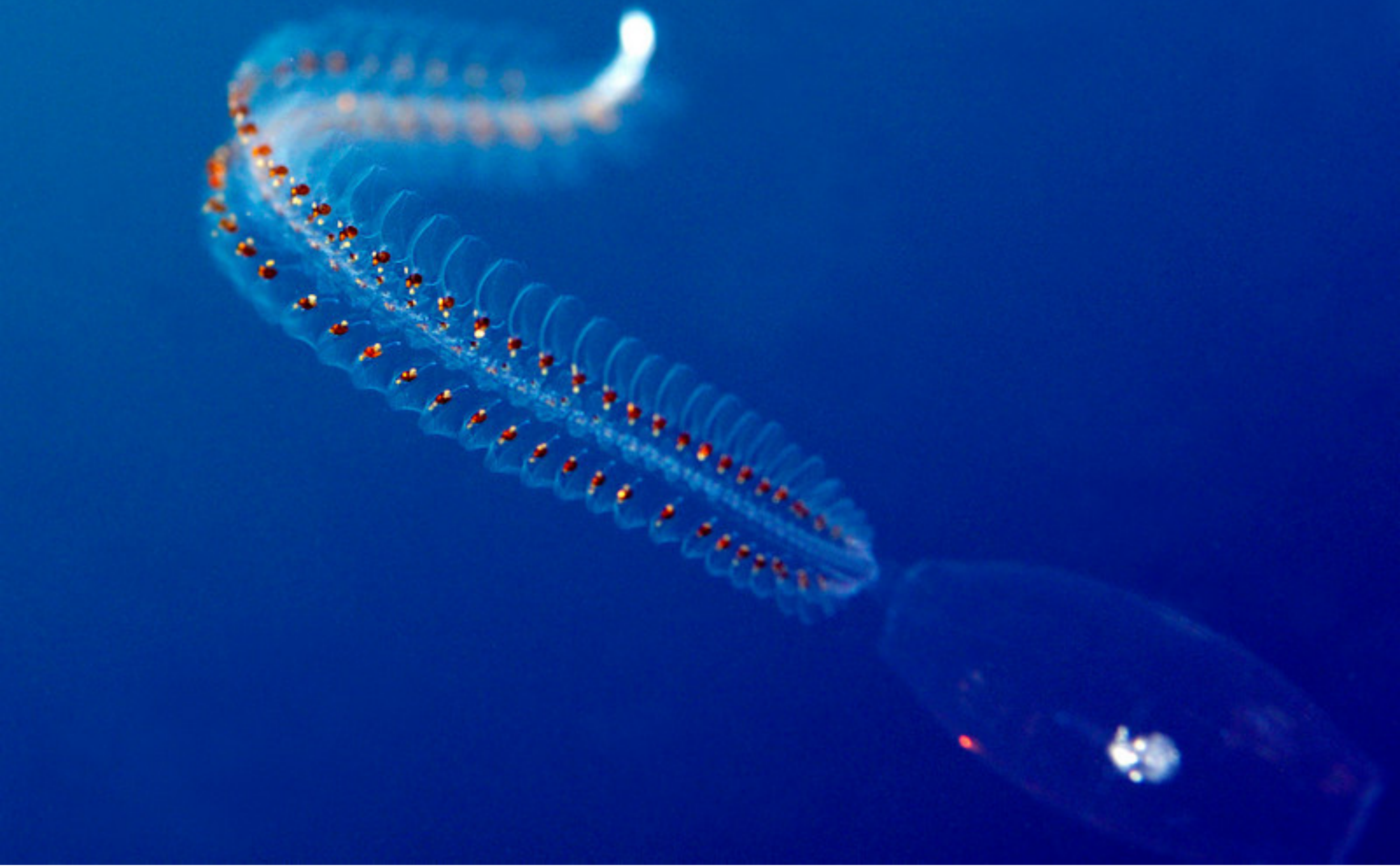
10 days at sea

25TB raw data

14M objects extracted

*Do we really have  
to check it all?*

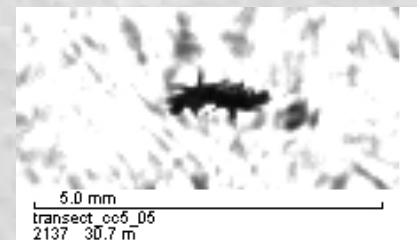
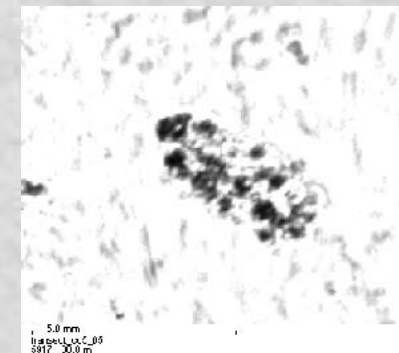
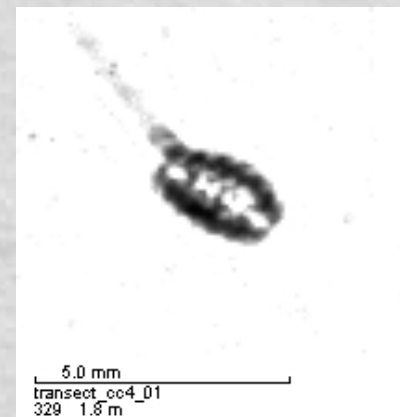
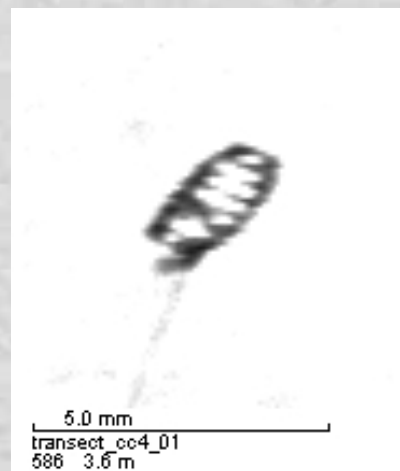
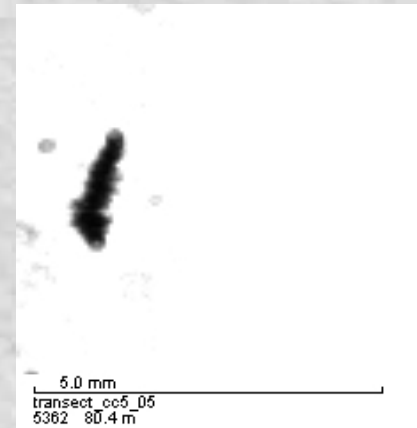
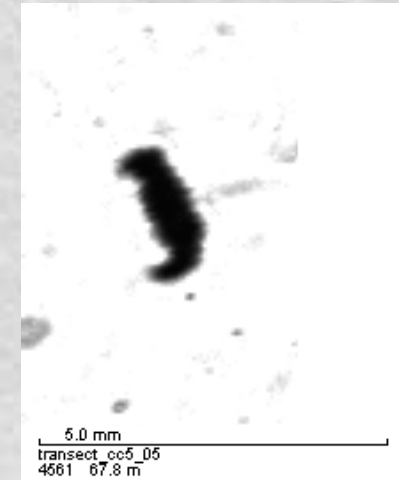
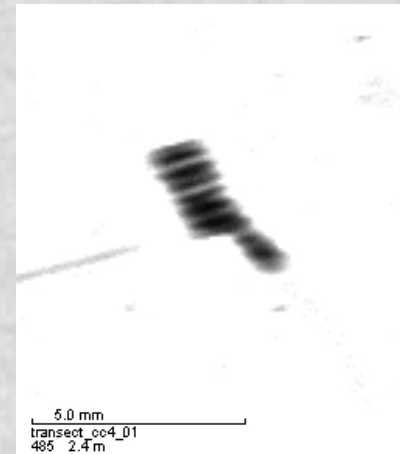
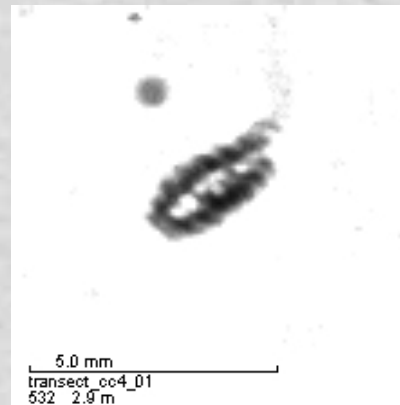
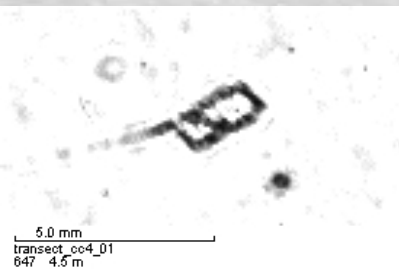
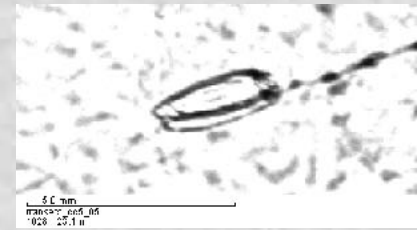
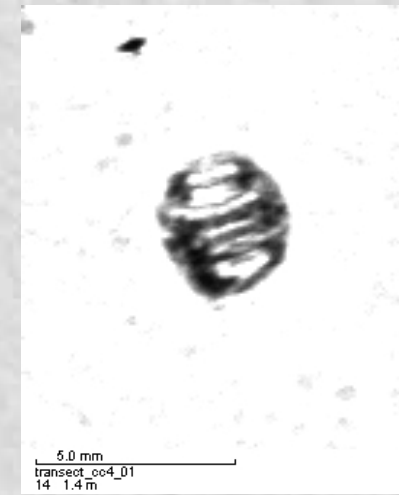
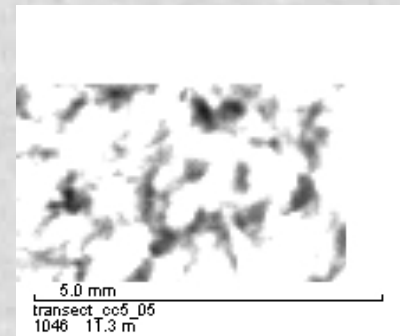
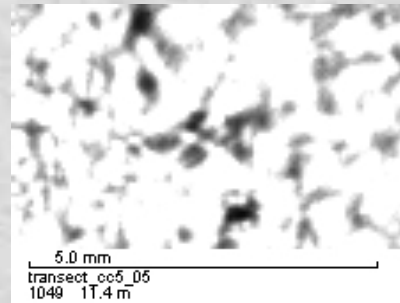
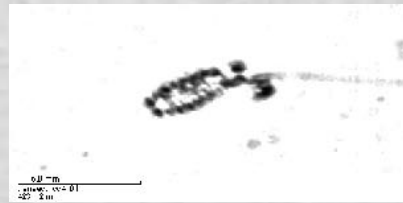




# **Classifying doliolids**

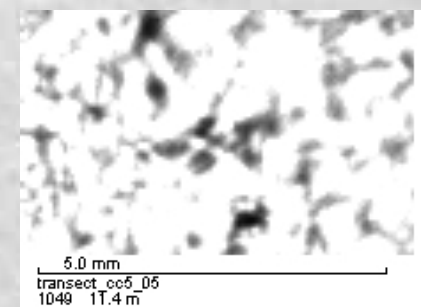
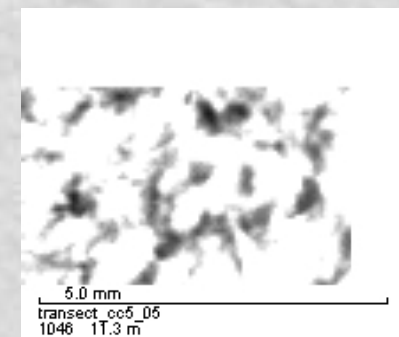
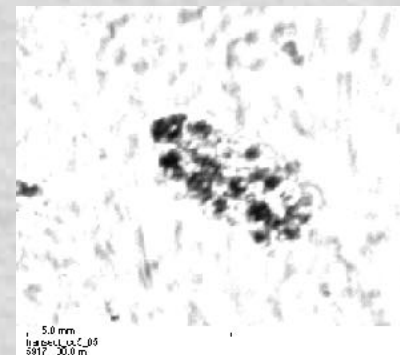
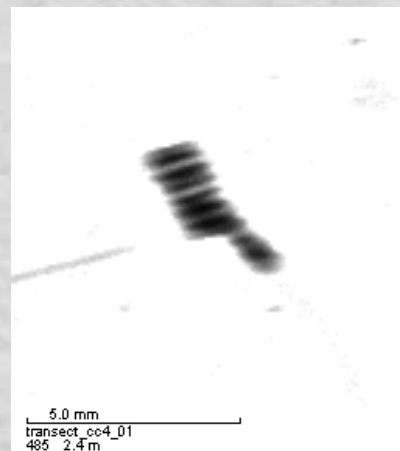
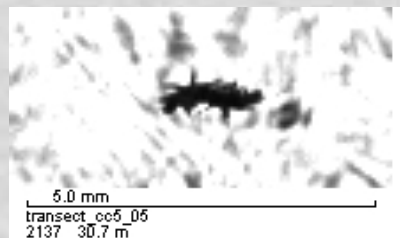
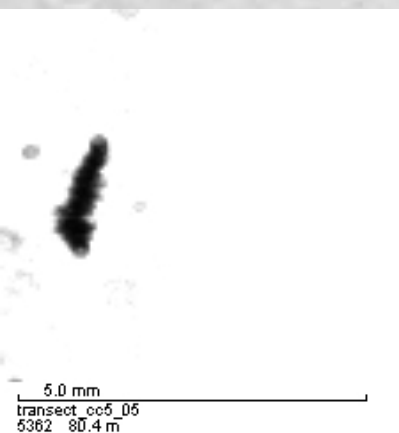
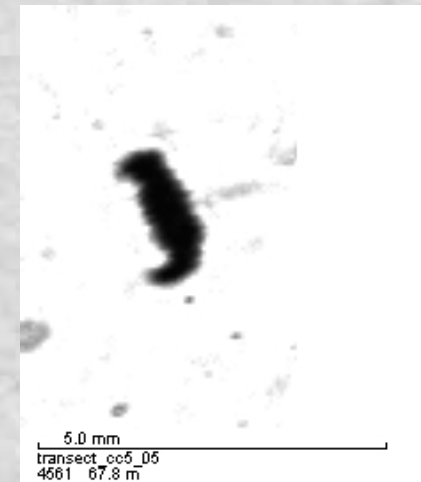
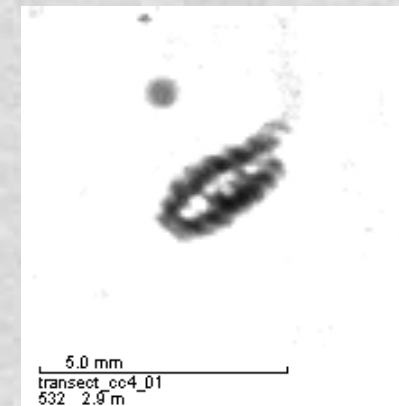
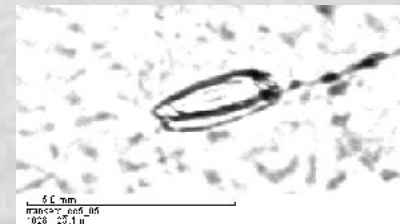
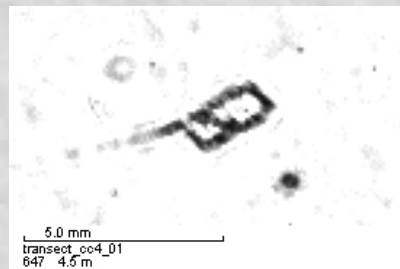
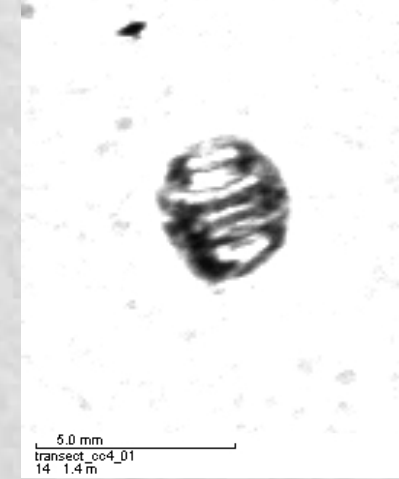
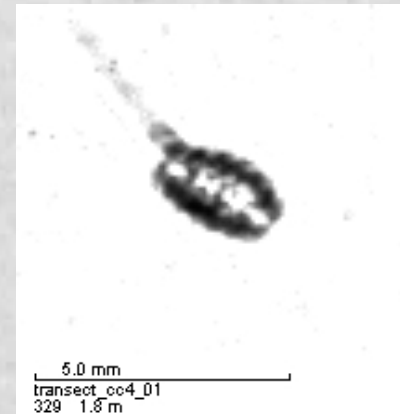
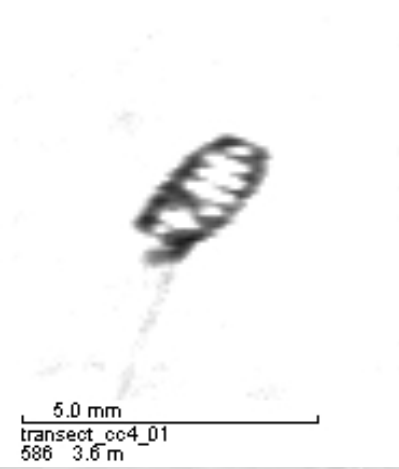
Raw prediction





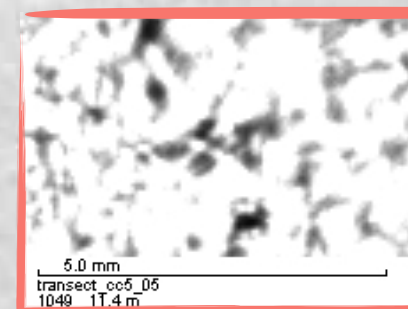
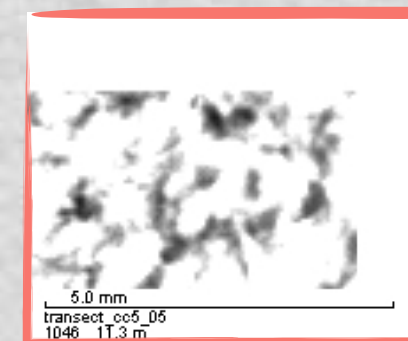
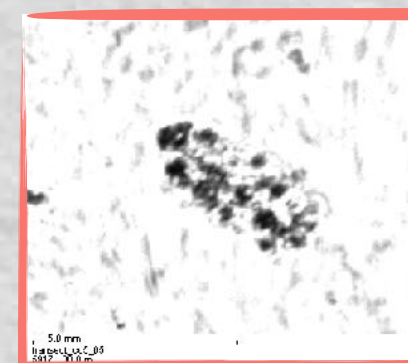
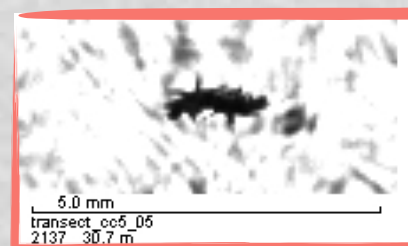
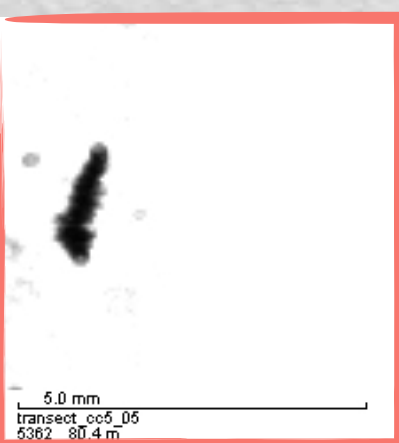
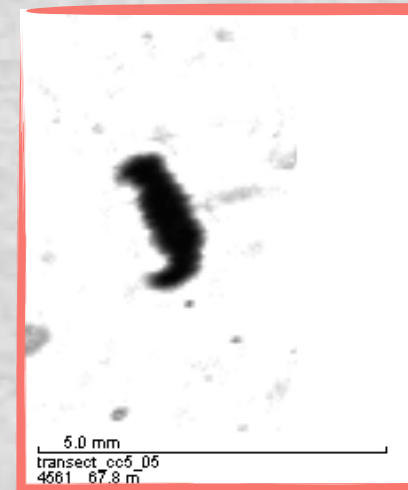
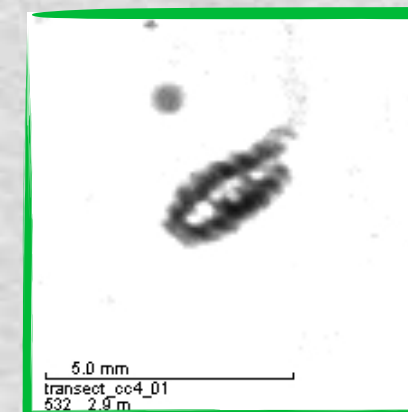
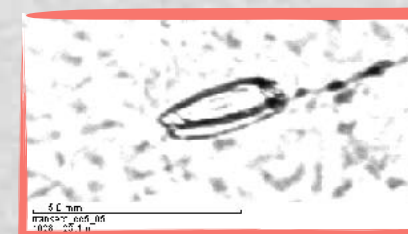
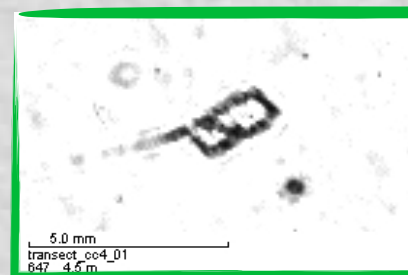
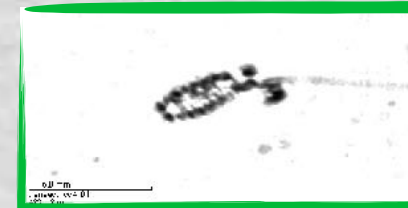
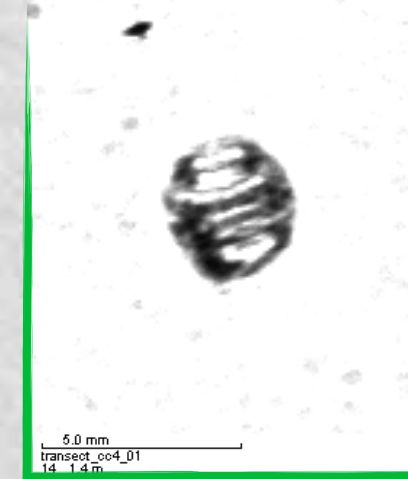
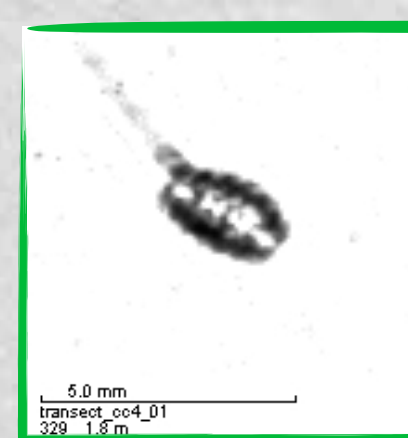
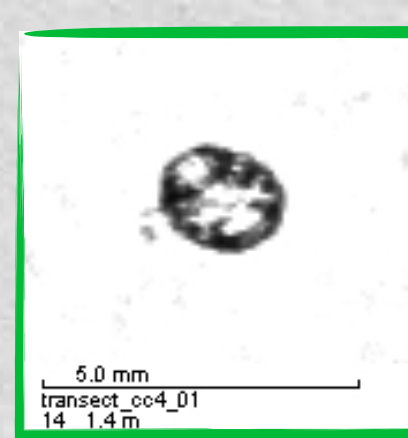
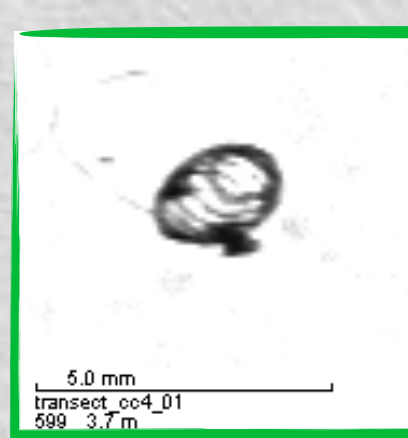
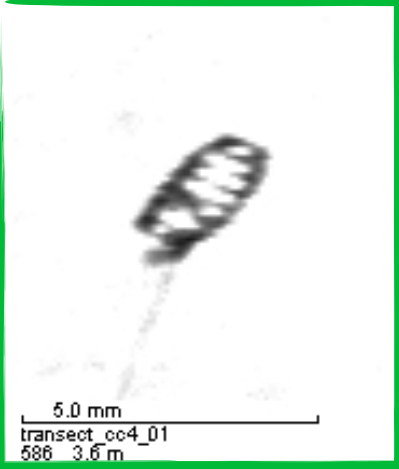
# Classifying doliolids

Raw prediction



# Classifying doliolids

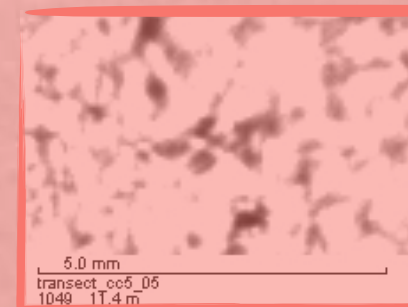
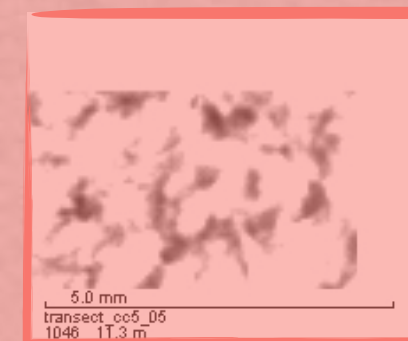
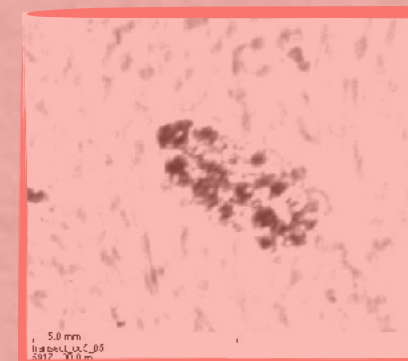
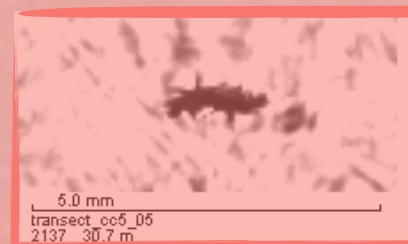
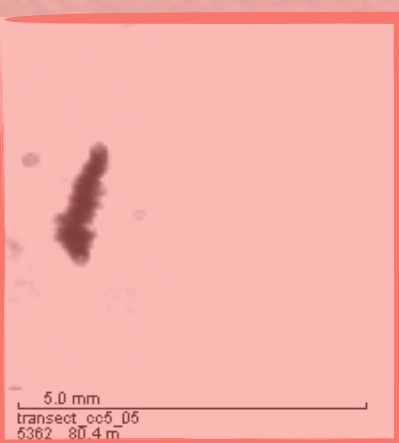
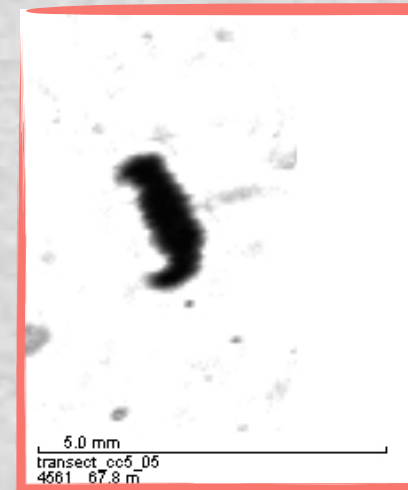
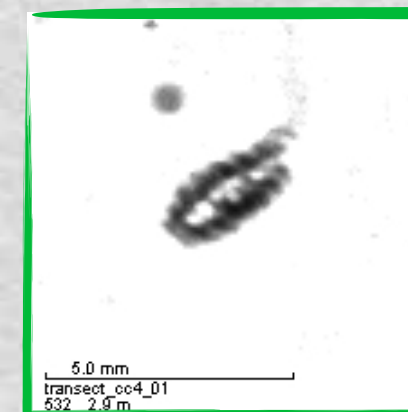
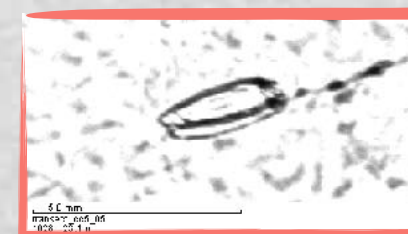
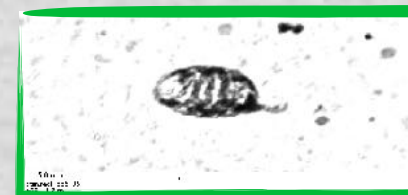
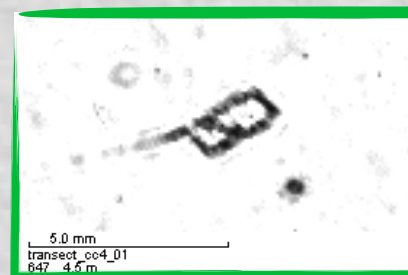
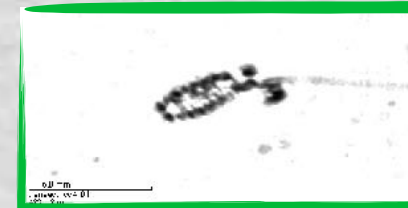
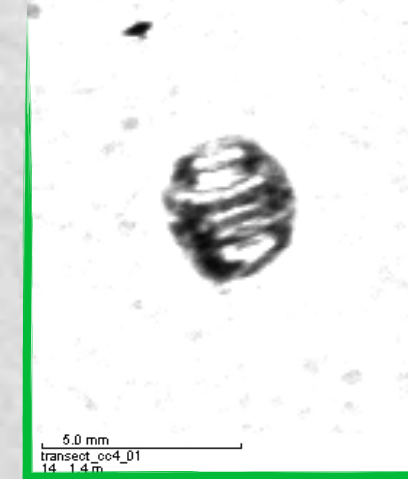
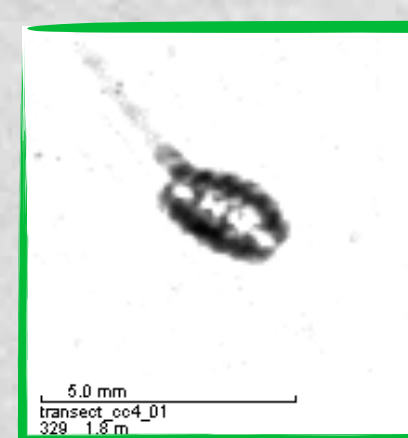
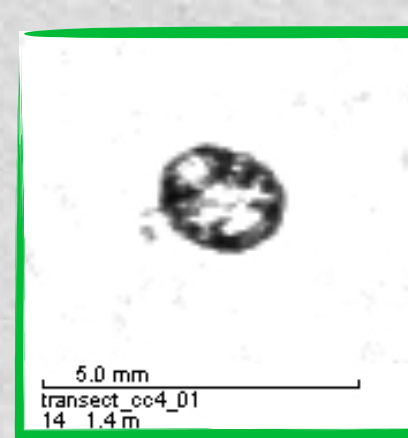
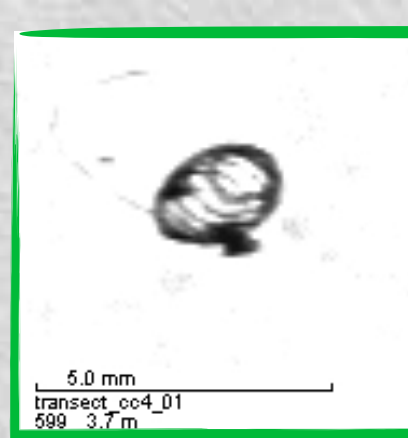
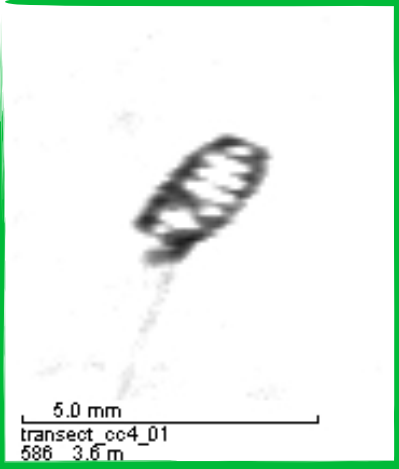
Sorted by prediction score



# Classifying doliolids

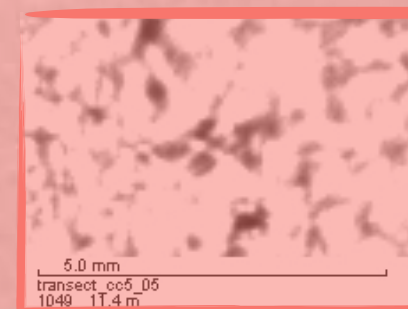
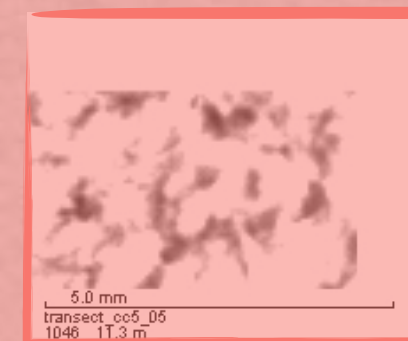
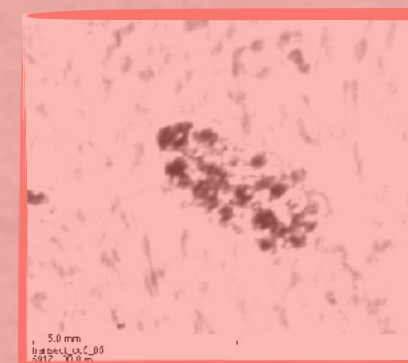
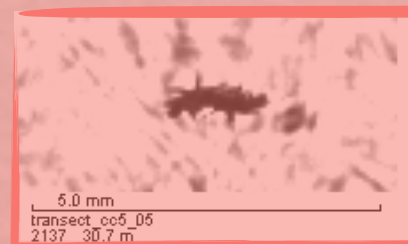
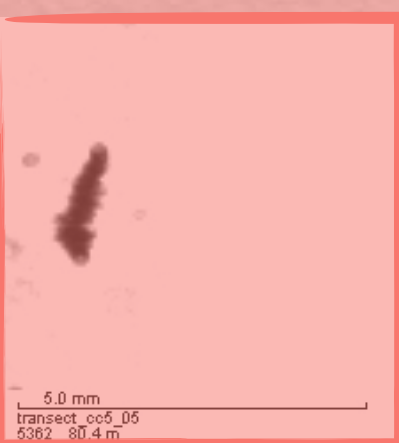
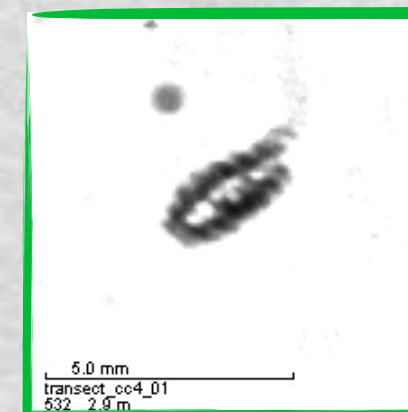
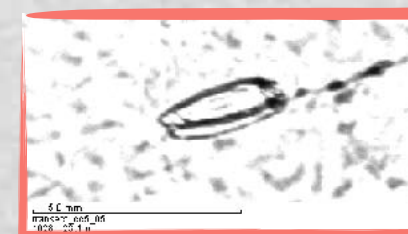
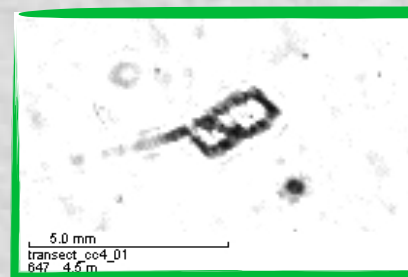
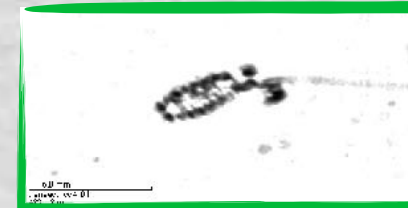
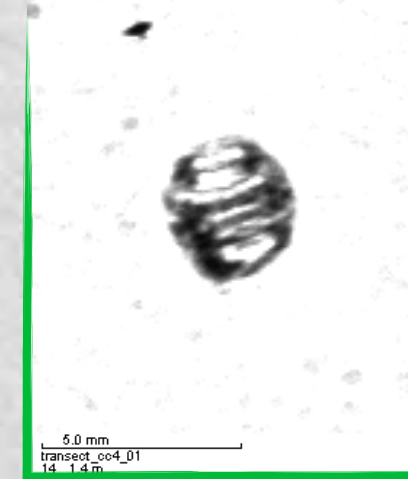
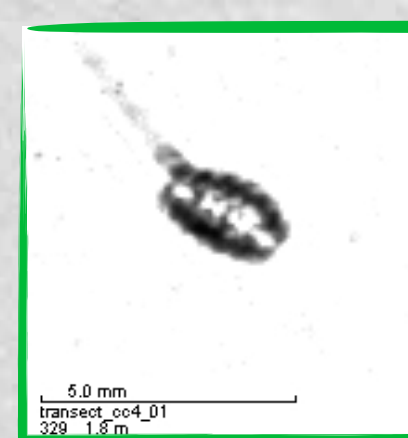
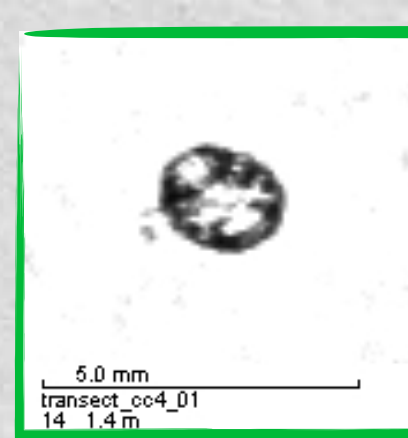
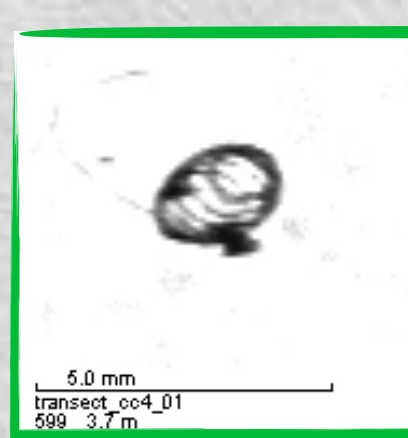
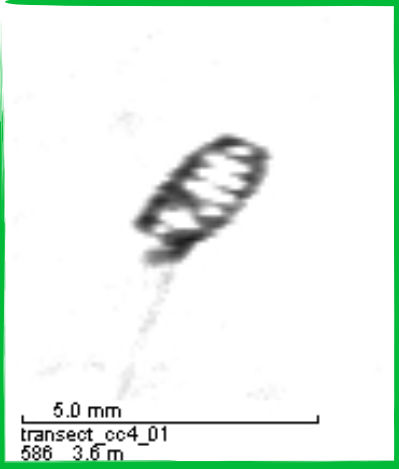
Discard “unsure” data





# Classifying doliolids

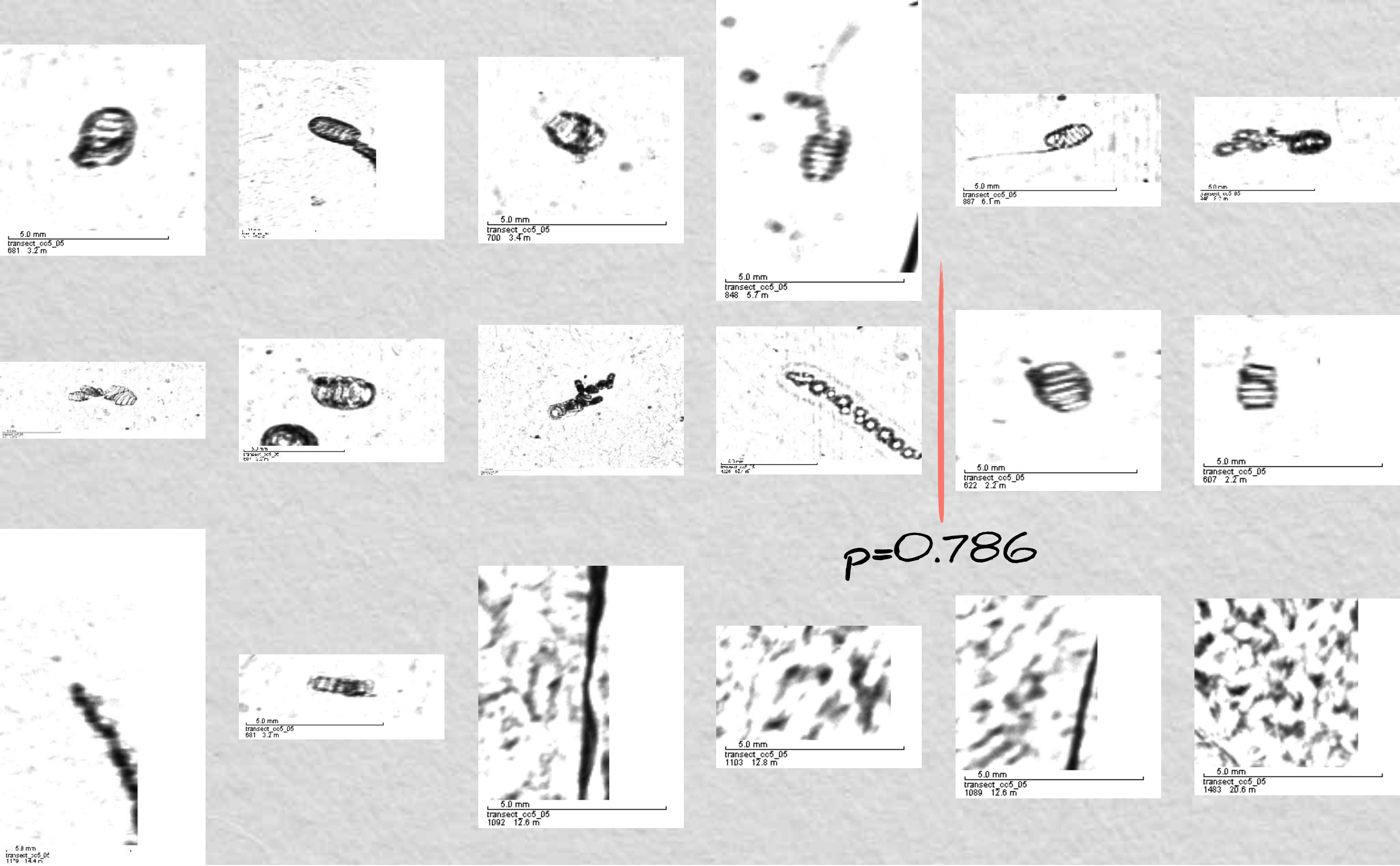
Discard “unsure” data



# Classifying doliolids

Discard “unsure” data

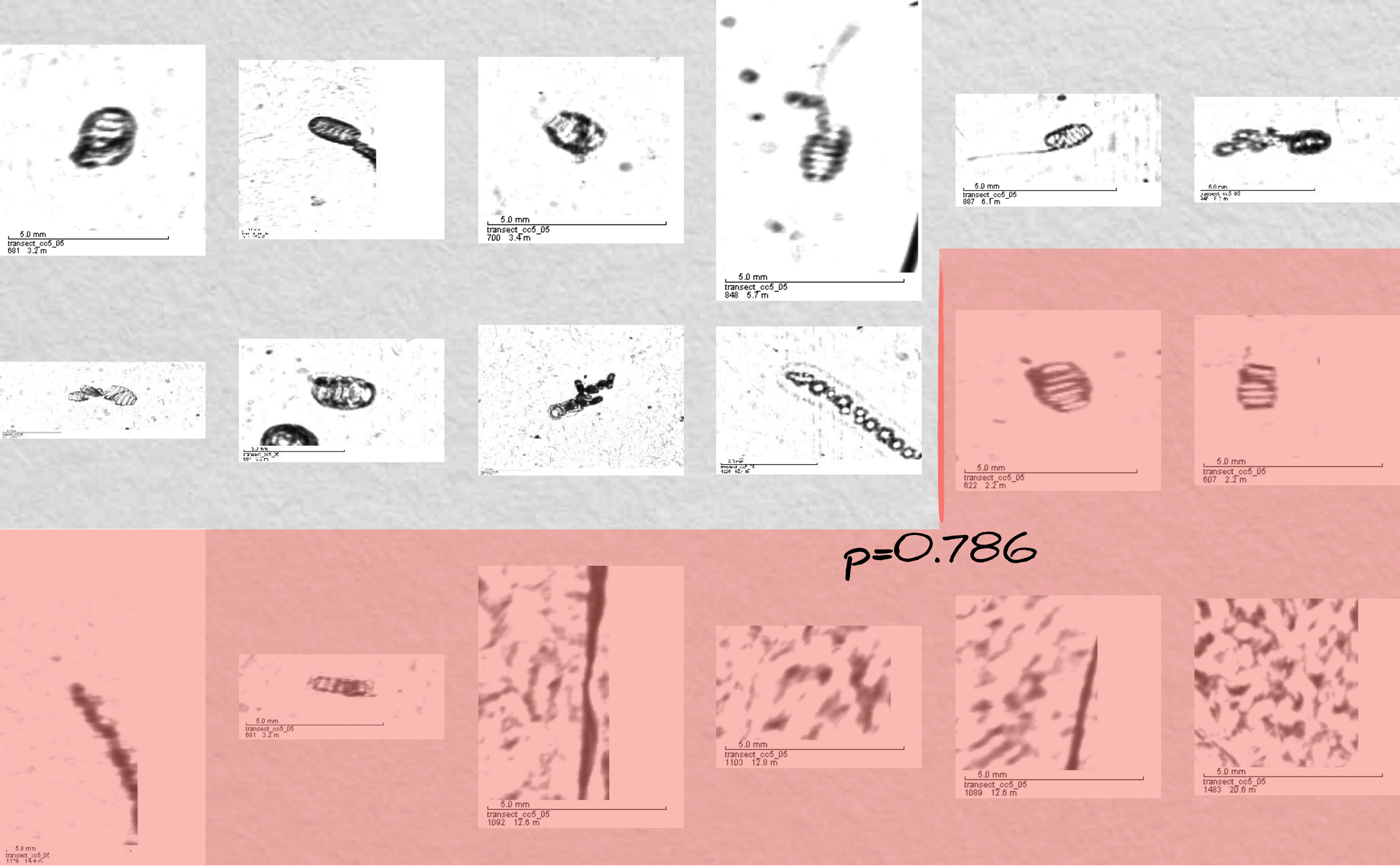




# Classifying doliolids

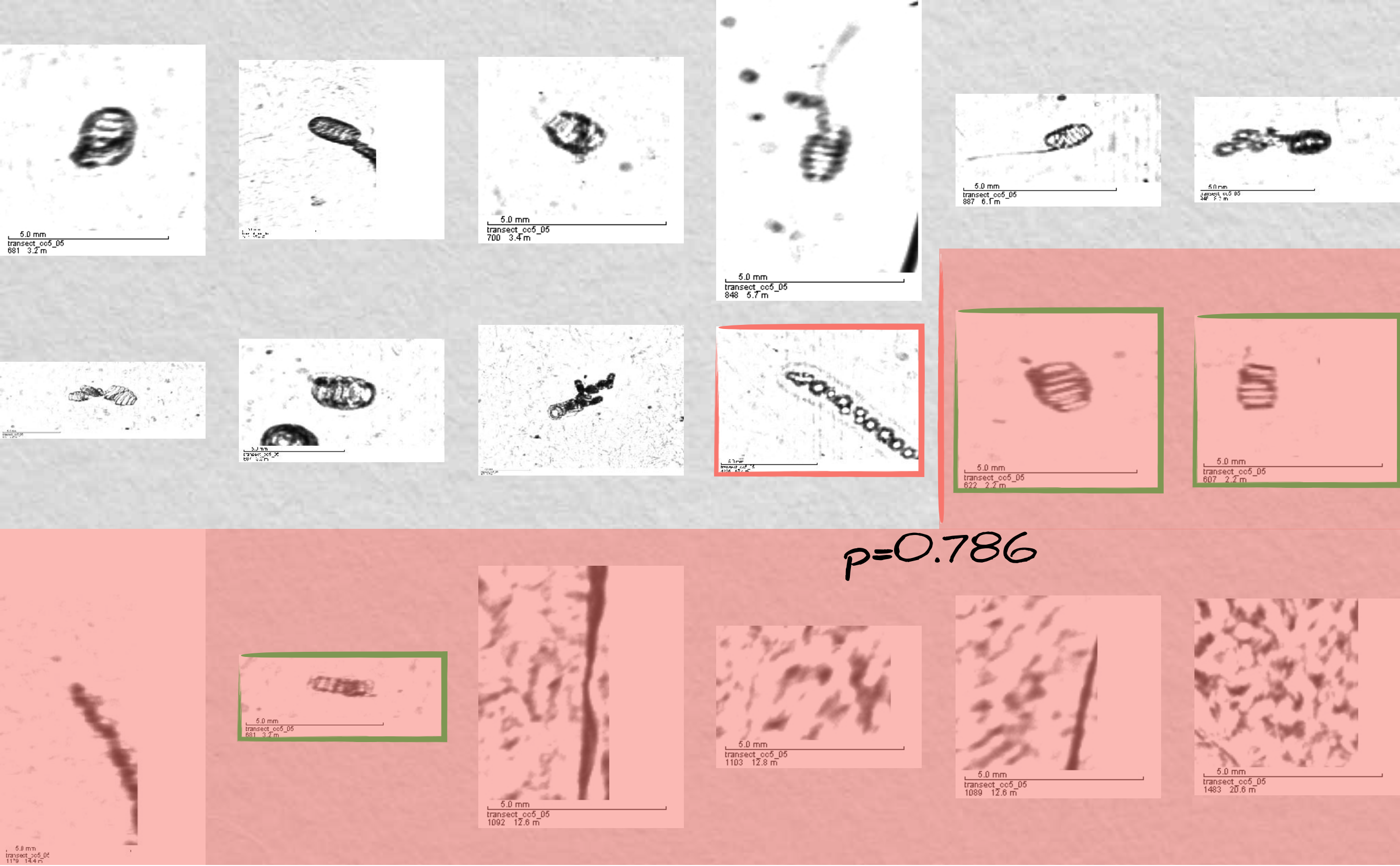
New data set sorted by classification score





# Classifying doliolids

New data set sorted by classification score



# Classifying doliolids

New data set sorted by classification score

# It works but how much is discarded?

**Table 2**

Classification metrics before and after filtering out objects with low prediction confidence: number of particles before filtering ( $n$ ); percentage of data kept after filtering; precision, recall, and F1 score before and after filtering, and difference (after–before). Improvements (positive differences) are bolded. Non-living groups are presented first, groups of biological interest second.

Class	$n$	%kept	Precision			Recall			F1		
			Before	After	Diff	Before	After	Diff	Before	After	Diff
Dark aggregates	60 164	6.5	77	95	<b>19</b>	50	7	–43	60	7	–54
Light aggregates	4 209	4.2	8	17	<b>9</b>	53	4	–49	14	4	–10
Fibres	8 055	6.9	46	85	<b>38</b>	56	7	–49	51	7	–44
Copepods	17 459	22.4	54	88	<b>34</b>	72	22	–49	62	22	–39
Doliolids	30 478	40.2	80	95	<b>16</b>	64	40	–24	71	40	–31
Fish larvae	802	23.2	12	80	<b>67</b>	62	23	–39	21	23	<b>3</b>
Trachymedusae	524	50.6	9	62	<b>53</b>	79	51	–29	16	51	<b>35</b>
Diatom chains	11 015	28.6	75	97	<b>22</b>	72	29	–43	73	29	–45
Acantharian radiolarians	1 021	18.9	7	65	<b>58</b>	74	19	–55	14	19	<b>5</b>
Radiolarian colonies	4 367	16.7	24	94	<b>70</b>	62	17	–45	35	17	–18
Solitary radiolarians	13 049	65.7	68	88	<b>19</b>	89	66	–23	77	66	–12
Shrimps	213	52.6	51	89	<b>38</b>	74	53	–21	60	53	–7



# It works but how much is discarded?

**Table 2**

Classification metrics before and after filtering out objects with low prediction confidence: number of particles before filtering ( $n$ ); percentage of data kept after filtering; precision, recall, and F1 score before and after filtering, and difference (after–before). Improvements (positive differences) are bolded. Non-living groups are presented first, groups of biological interest second.

Class	$n$	%kept	Precision			Recall			F1		
			Before	After	Diff	Before	After	Diff	Before	After	Diff
Dark aggregates	60 164	6.5	77	95	<b>19</b>	50	7	–43	60	7	–54
Light aggregates	4 209	4.2	8	17	<b>9</b>	53	4	–49	14	4	–10
Fibres	8 055	6.9	46	85	<b>38</b>	56	7	–49	51	7	–44
Copepods	17 459	22.4	54	88	<b>34</b>	72	22	–49	62	22	–39
Doliolids	30 478	40.2	80	95	<b>16</b>	64	40	–24	71	40	–31
Fish larvae	802	23.2	12	80	<b>67</b>	62	23	–39	21	23	<b>3</b>
Trachymedusae	524	50.6	9	62	<b>53</b>	79	51	–29	16	51	<b>35</b>
Diatom chains	11 015	28.6	75	97	<b>22</b>	72	29	–43	73	29	–45
Acantharian radiolarians	1 021	18.9	7	65	<b>58</b>	74	19	–55	14	19	<b>5</b>
Radiolarian colonies	4 367	16.7	24	94	<b>70</b>	62	17	–45	35	17	–18
Solitary radiolarians	13 049	65.7	68	88	<b>19</b>	89	66	–23	77	66	–12
Shrimps	213	52.6	51	89	<b>38</b>	74	53	–21	60	53	–7

# It works but how much is discarded?

**Table 2**

Classification metrics before and after filtering out objects with low prediction confidence: number of particles before filtering ( $n$ ); percentage of data kept after filtering; precision, recall, and F1 score before and after filtering, and difference (after–before). Improvements (positive differences) are bolded. Non-living groups are presented first, groups of biological interest second.

Class	$n$	%kept	Precision			Recall			F1		
			Before	After	Diff	Before	After	Diff	Before	After	Diff
Dark aggregates	60 164	6.5	77	95	<b>19</b>	50	7	–43	60	7	–54
Light aggregates	4 209	4.2	8	17	<b>9</b>	53	4	–49	14	4	–10
Fibres	8 055	6.9	46	85	<b>38</b>	56	7	–49	51	7	–44
Copepods	17 459	22.4	54	88	<b>34</b>	72	22	–49	62	22	–39
Doliolids	30 478	40.2	80	95	<b>16</b>	64	40	–24	71	40	–31
Fish larvae	802	23.2	12	80	<b>67</b>	62	23	–39	21	23	<b>3</b>
Trachymedusae	524	50.6	9	62	<b>53</b>	79	51	–29	16	51	<b>35</b>
Diatom chains	11 015	28.6	75	97	<b>22</b>	72	29	–43	73	29	–45
Acantharian radiolarians	1 021	18.9	7	65	<b>58</b>	74	19	–55	14	19	<b>5</b>
Radiolarian colonies	4 367	16.7	24	94	<b>70</b>	62	17	–45	35	17	–18
Solitary radiolarians	13 049	65.7	68	88	<b>19</b>	89	66	–23	77	66	–12
Shrimps	213	52.6	51	89	<b>38</b>	74	53	–21	60	53	–7

# It works but how much is discarded?

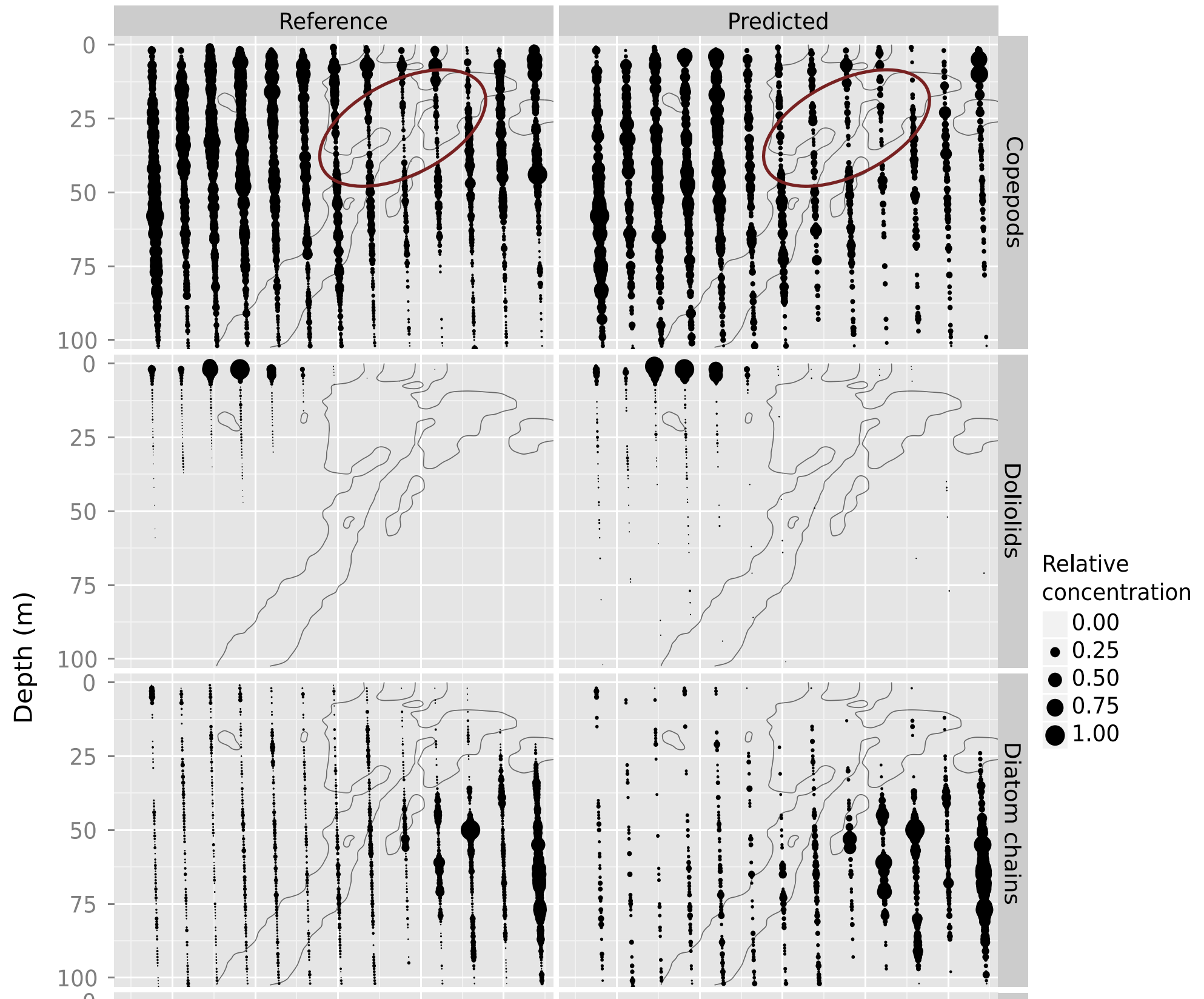
**Table 2**

Classification metrics before and after filtering out objects with low prediction confidence: number of particles before filtering ( $n$ ); percentage of data kept after filtering; precision, recall, and F1 score before and after filtering, and difference (after–before). Improvements (positive differences) are bolded. Non-living groups are presented first, groups of biological interest second.

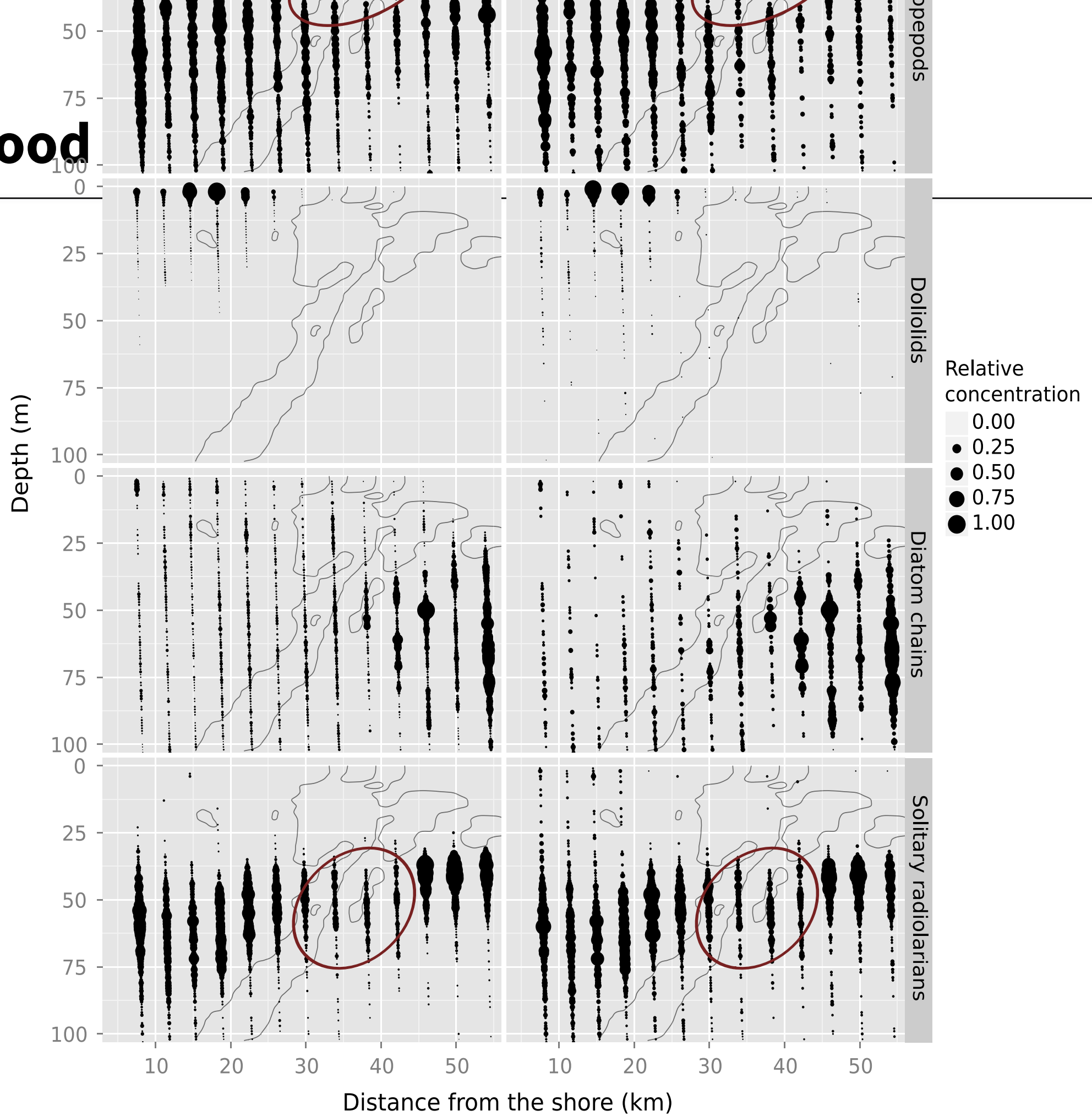
Class	$n$	%kept	Precision			Recall			F1		
			Before	After	Diff	Before	After	Diff	Before	After	Diff
Dark aggregates	60 164	6.5	77	95	<b>19</b>	50	7	–43	60	7	–54
Light aggregates	4 209	4.2	8	17	<b>9</b>	53	4	–49	14	4	–10
Fibres	8 055	6.9	46	85	<b>38</b>	56	7	–49	51	7	–44
Copepods	17 459	22.4	54	88	<b>34</b>	72	22	–49	62	22	–39
Doliolids	30 478	40.2	80	95	<b>16</b>	64	40	–24	71	40	–31
Fish larvae	802	23.2	12	80	<b>67</b>	62	23	–39	21	23	<b>3</b>
Trachymedusae	524	50.6	9	62	<b>53</b>	79	51	–29	16	51	<b>35</b>
Diatom chains	11 015	28.6	75	97	<b>22</b>	72	29	–43	73	29	–45
Acantharian radiolarians	1 021	18.9	7	65	<b>58</b>	74	19	–55	14	19	<b>5</b>
Radiolarian colonies	4 367	16.7	24	94	<b>70</b>	62	17	–45	35	17	–18
Solitary radiolarians	13 049	65.7	68	88	<b>19</b>	89	66	–23	77	66	–12
Shrimps	213	52.6	51	89	<b>38</b>	74	53	–21	60	53	–7



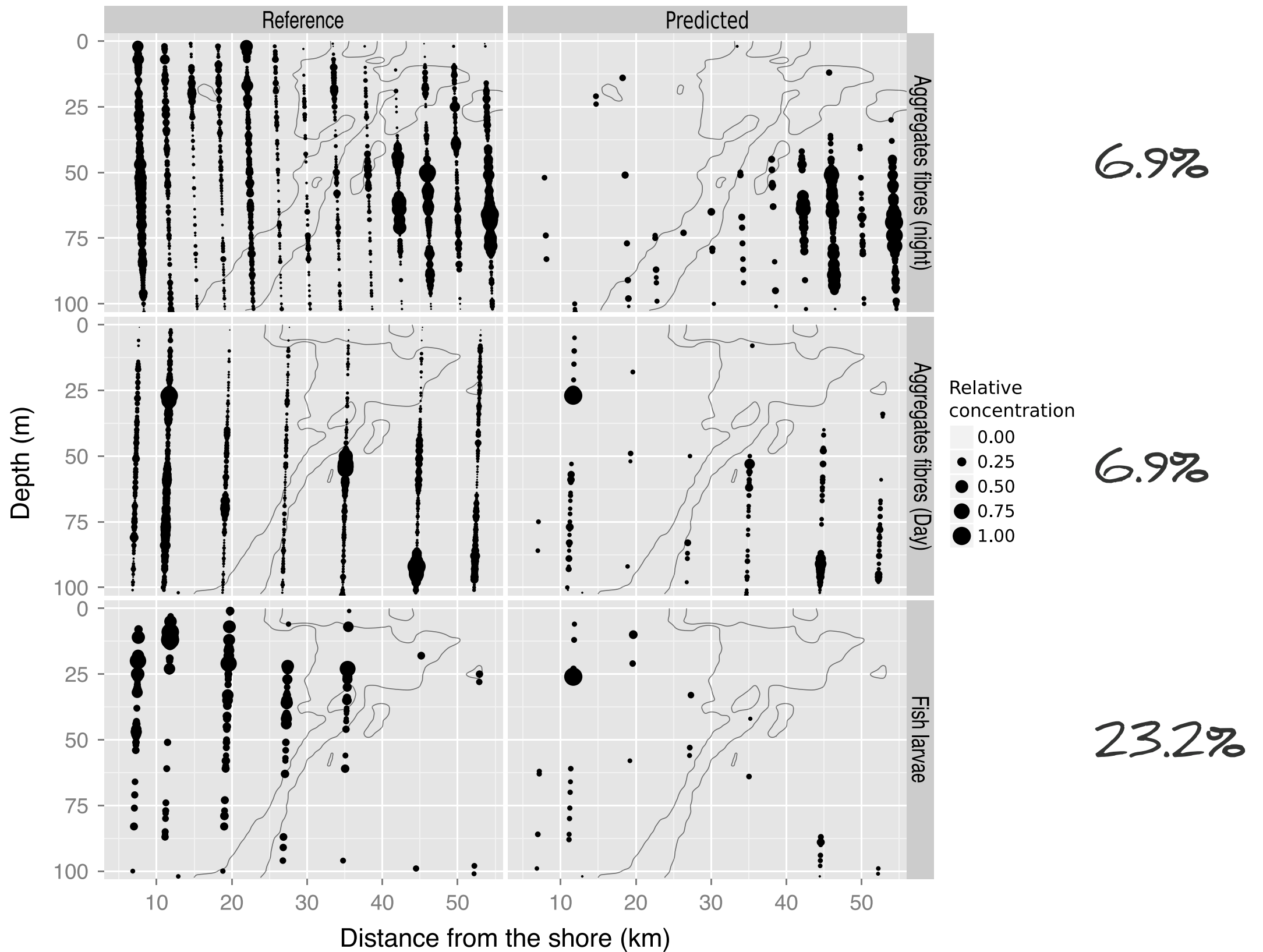
# The good



# The good

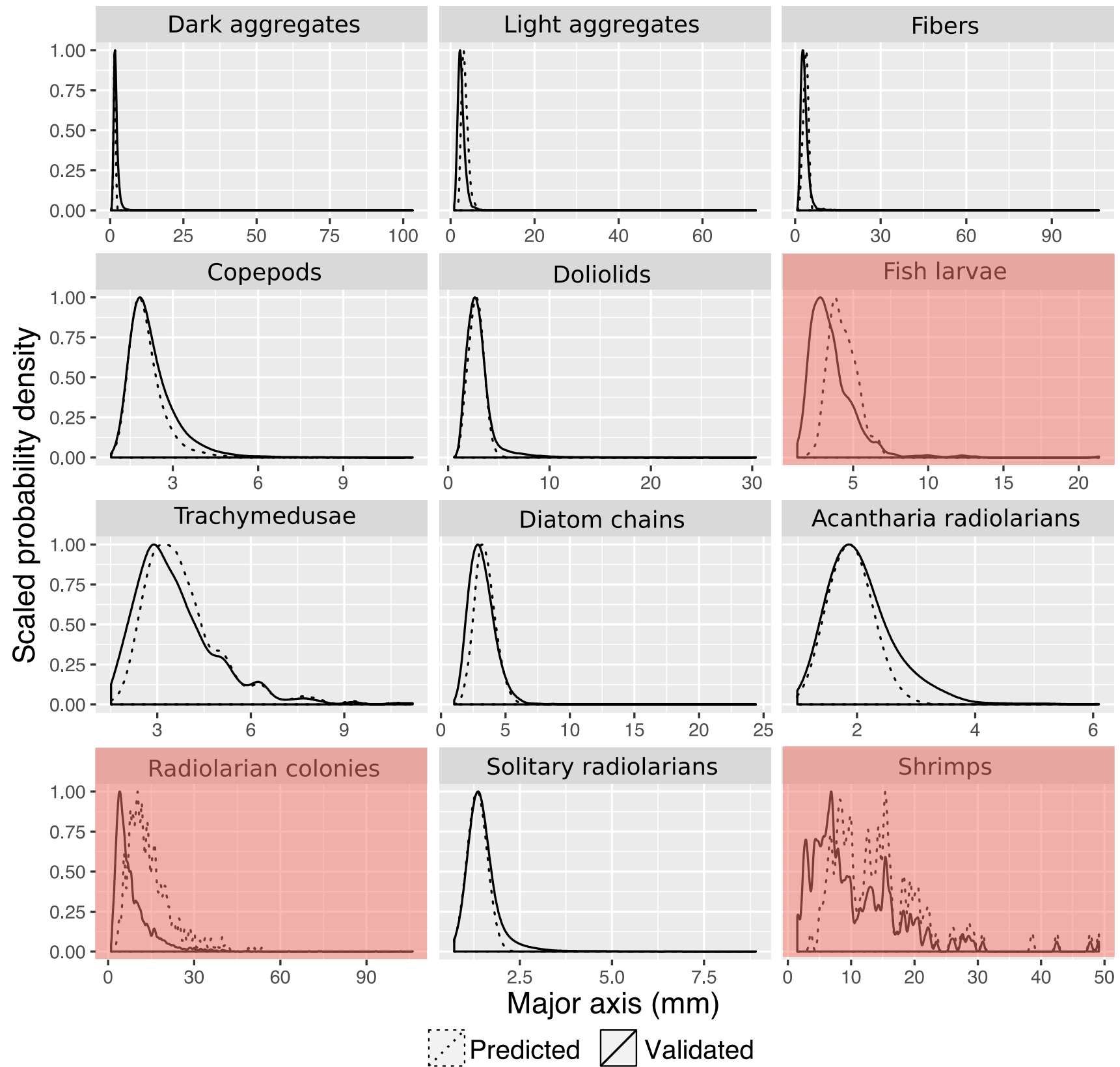


# The bad

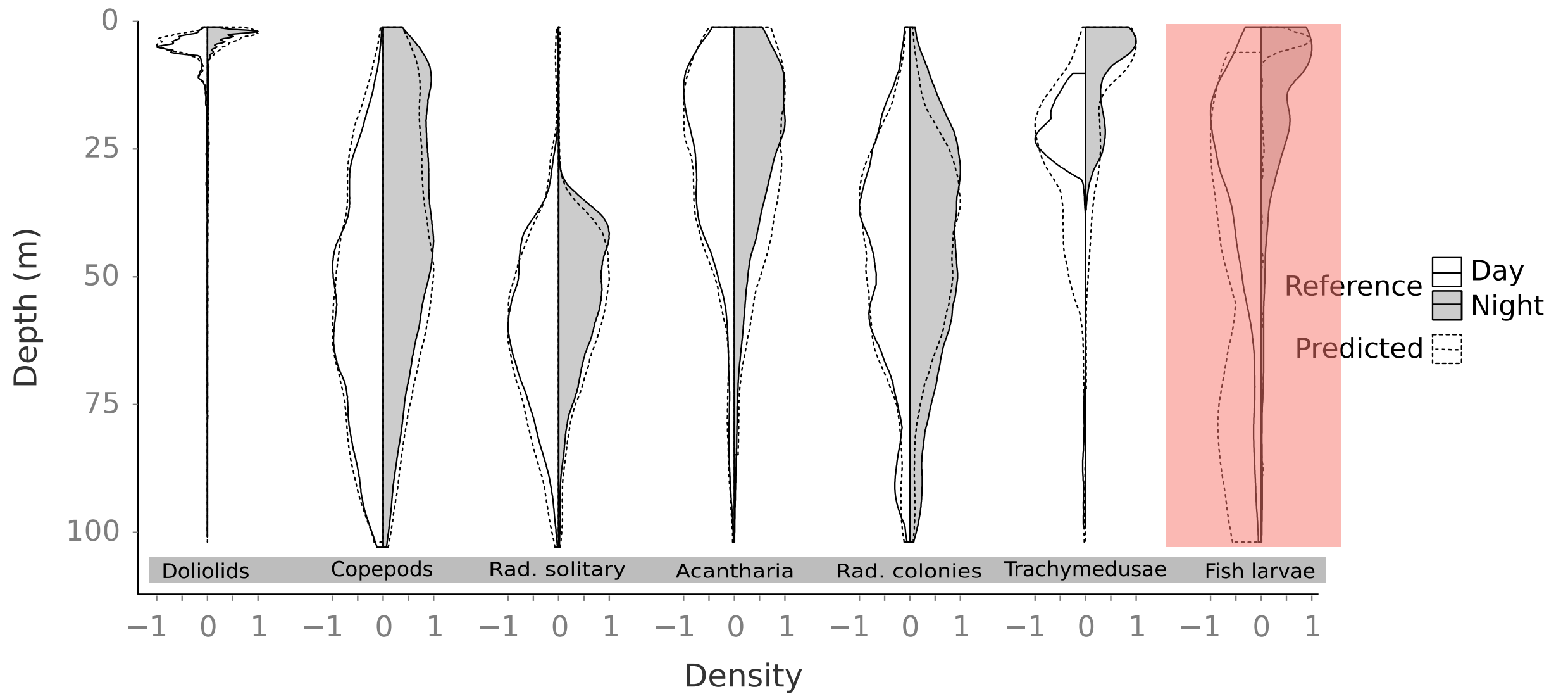




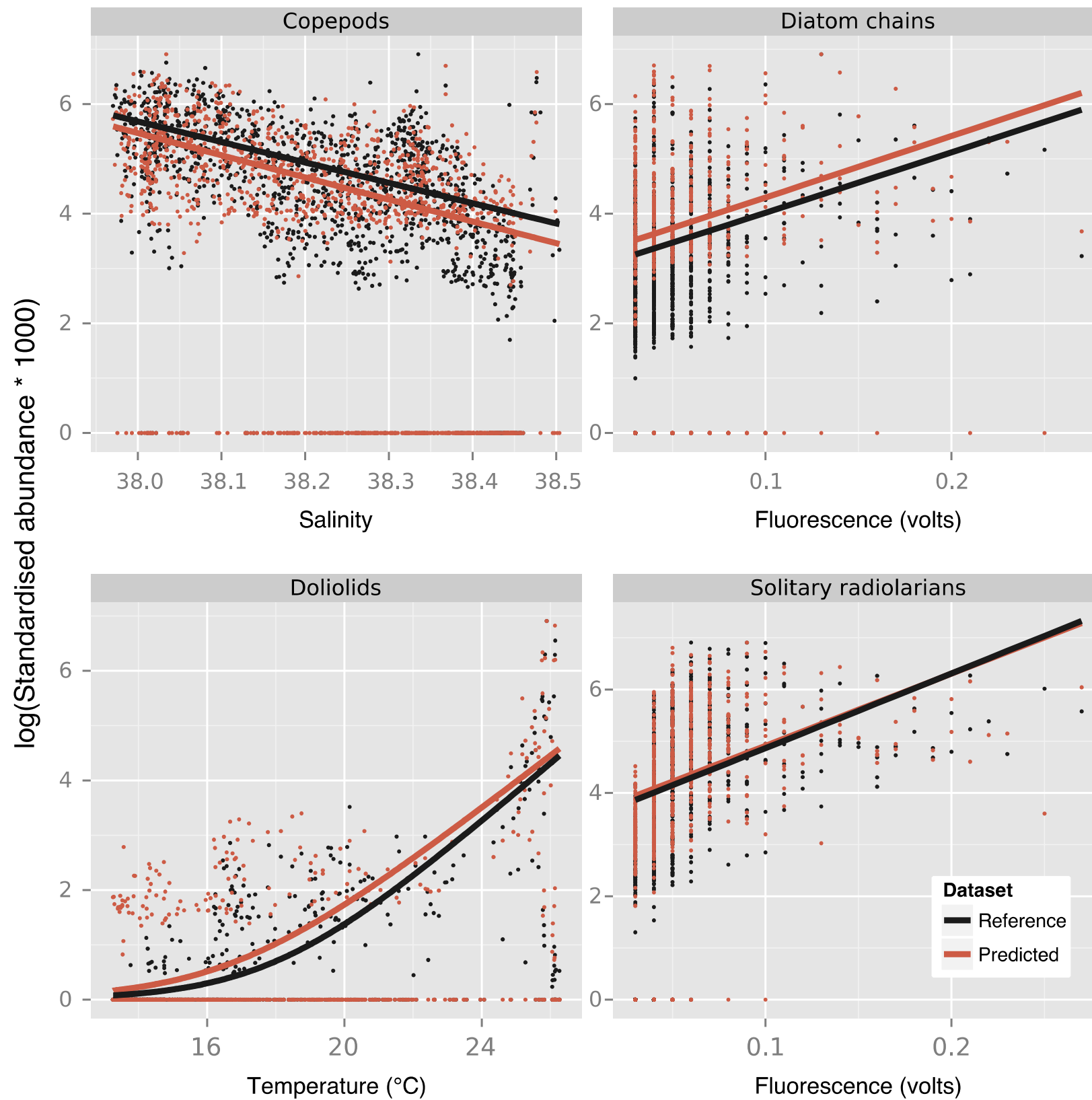
# Size distribution



# Diel migration



# Environmental relationships





# Take home message

---

Use P

Thank you for your attention



**Co-authors:**

*R Faillettaz*

*M Picheral*

*J Y Luo*

*C M Guigand*

*R K Cowen*

*J-O Irisson*

**Contributors**

*A Maupetit*

*F ferrando*

