

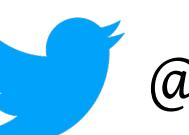
Complexaton Challenge 5

Trawling through the rubbish: Data mining of the scientific literature on marine plastic pollution

Erik van Sebille and Mikael Kaandorp
e.vansebille@uu.nl | m.l.a.kaandorp@uu.nl



Institute for
Marine and Atmospheric
research Utrecht



@UFollowtheOcean

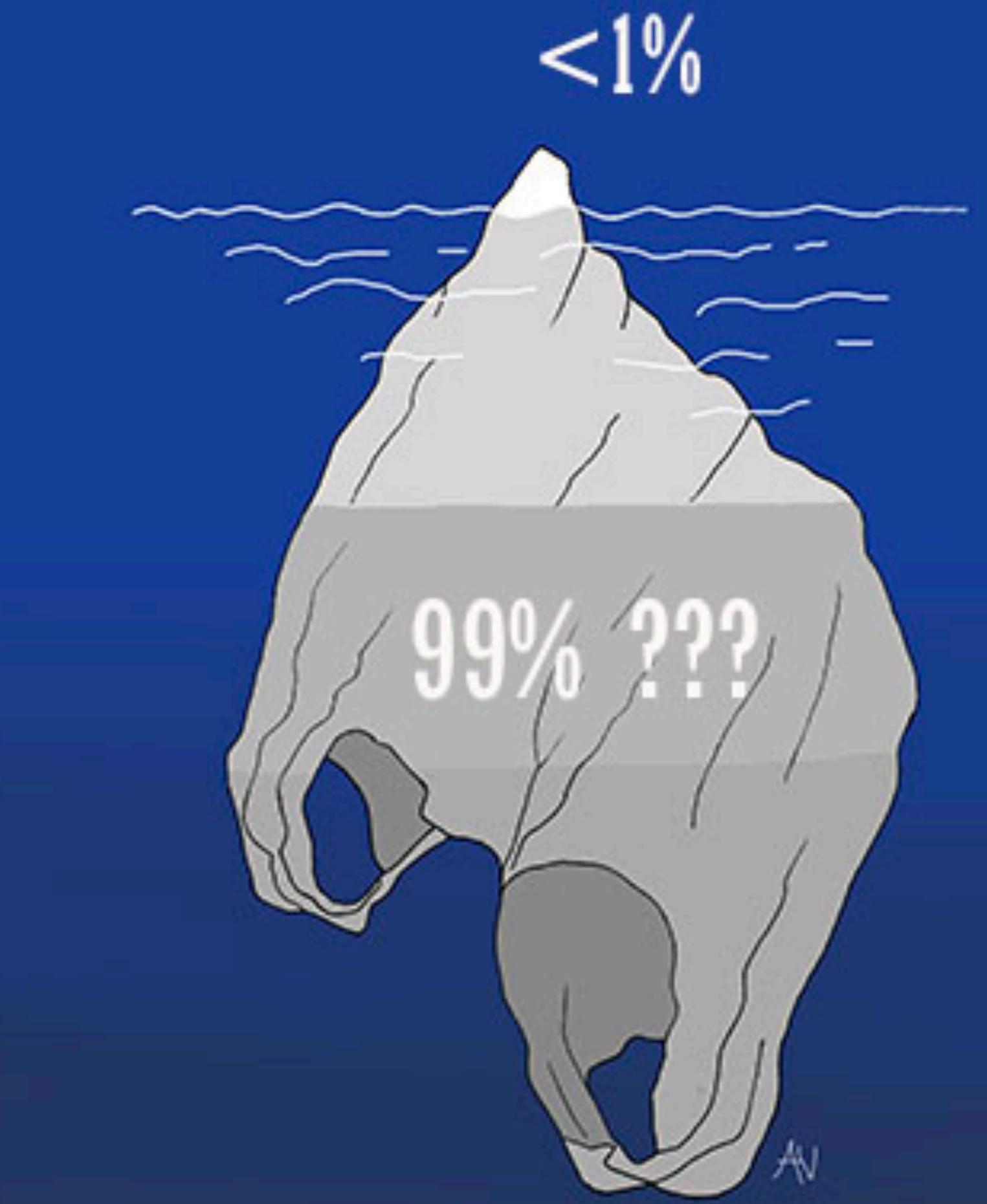
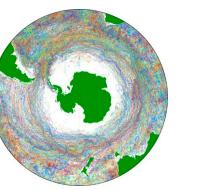
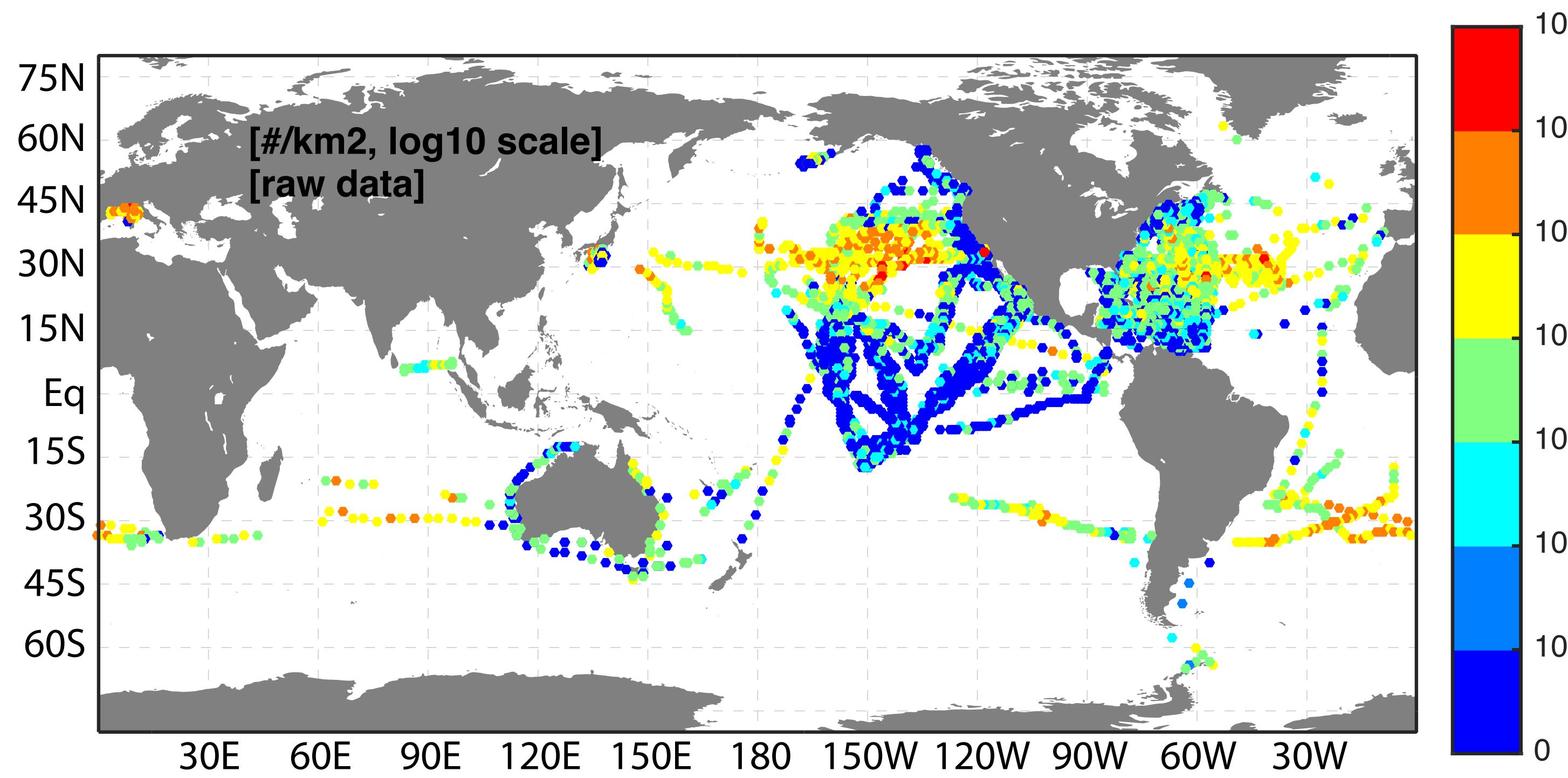
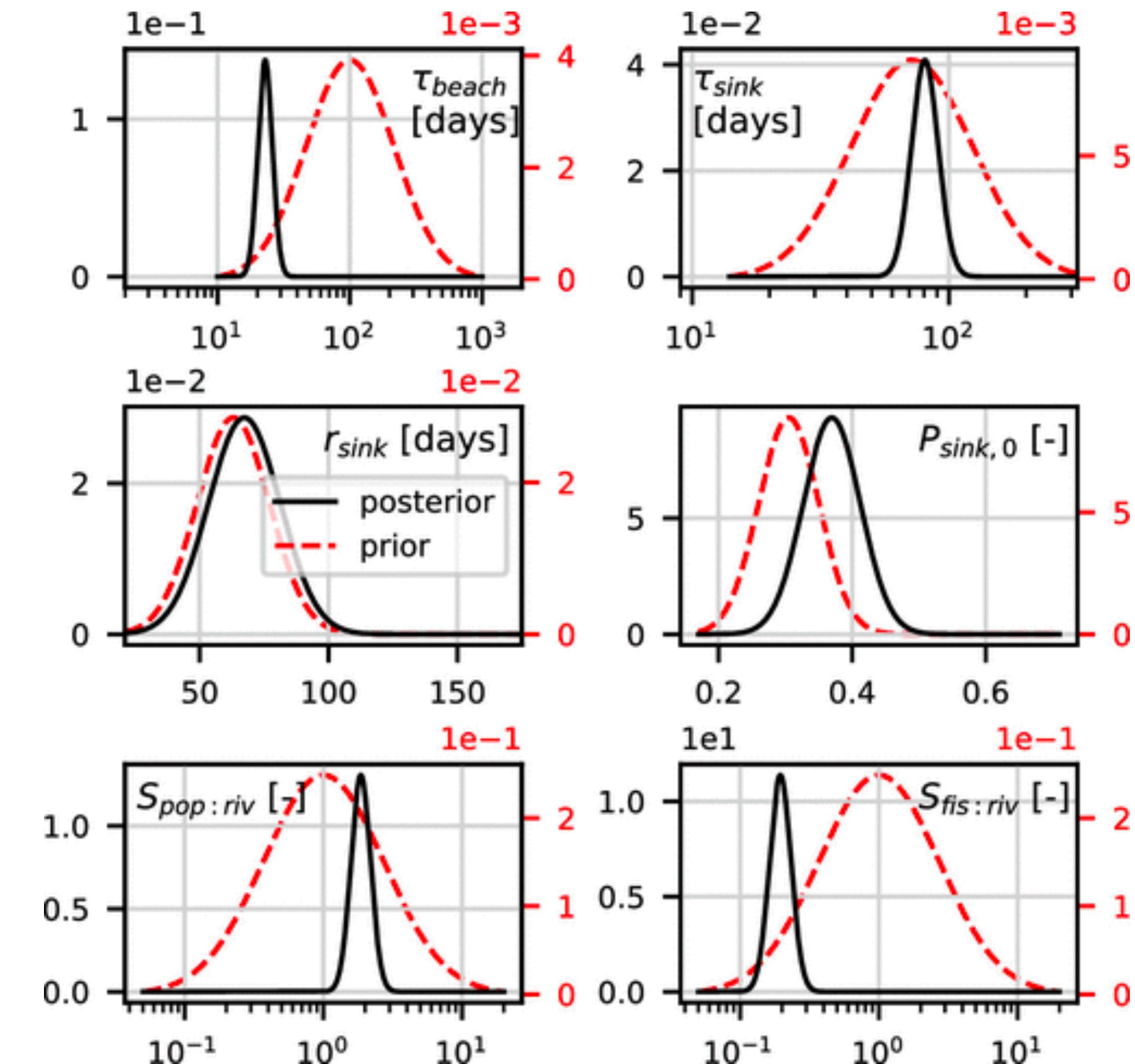
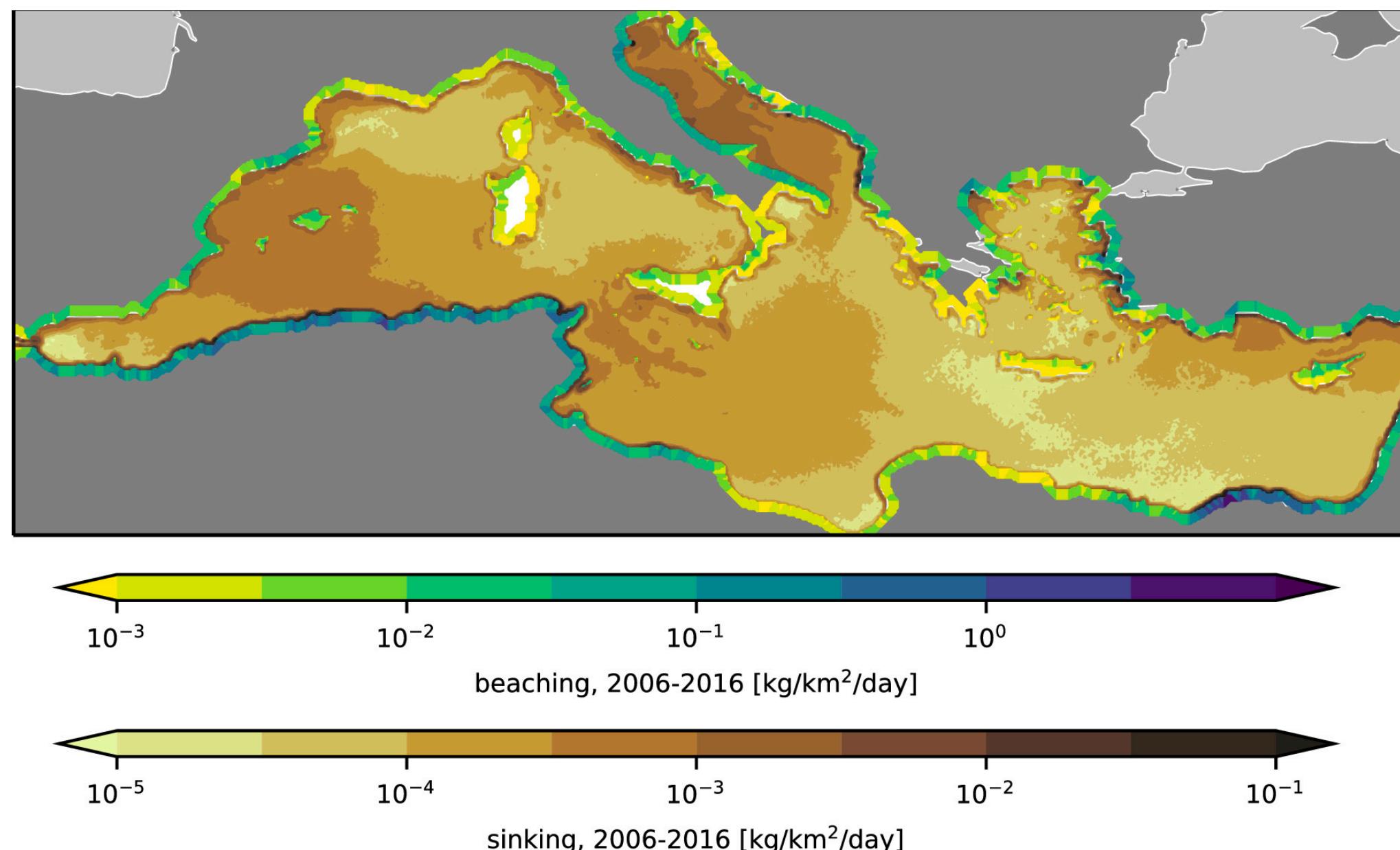
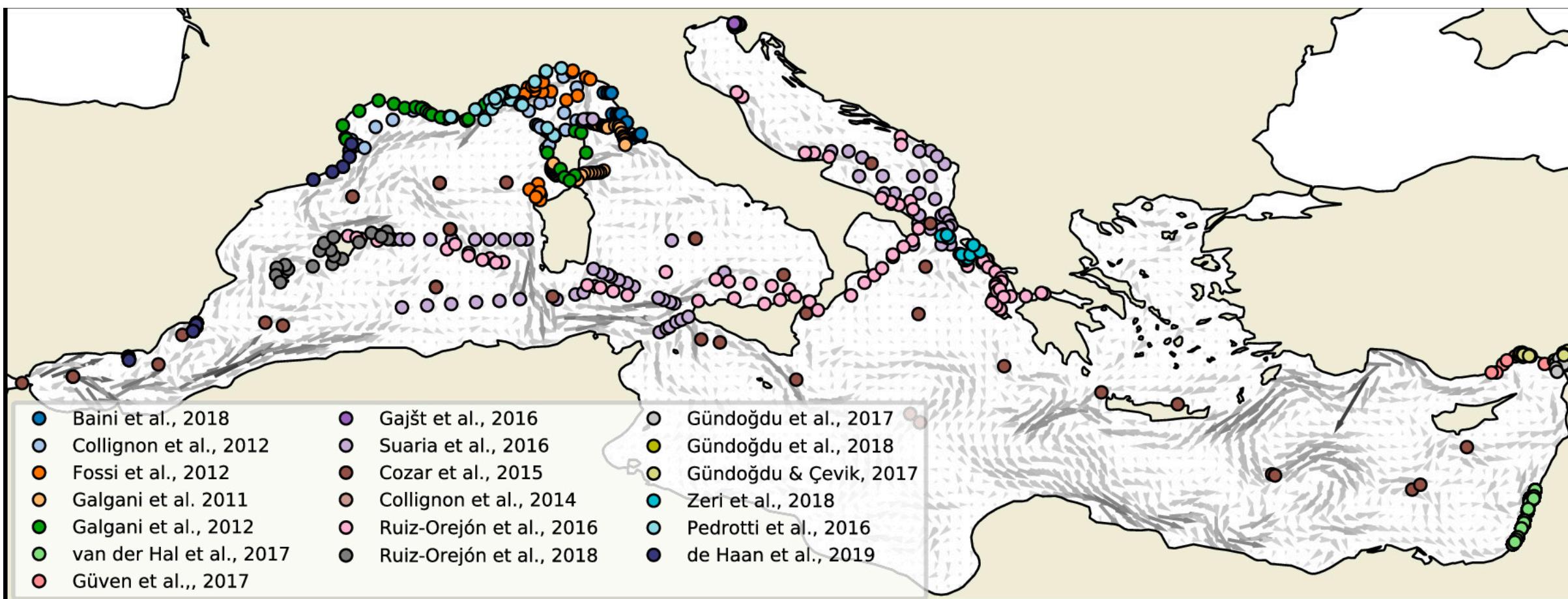


Illustration: Aike Vonk - uu.nl/plasticsoep

Gathering data from the ocean

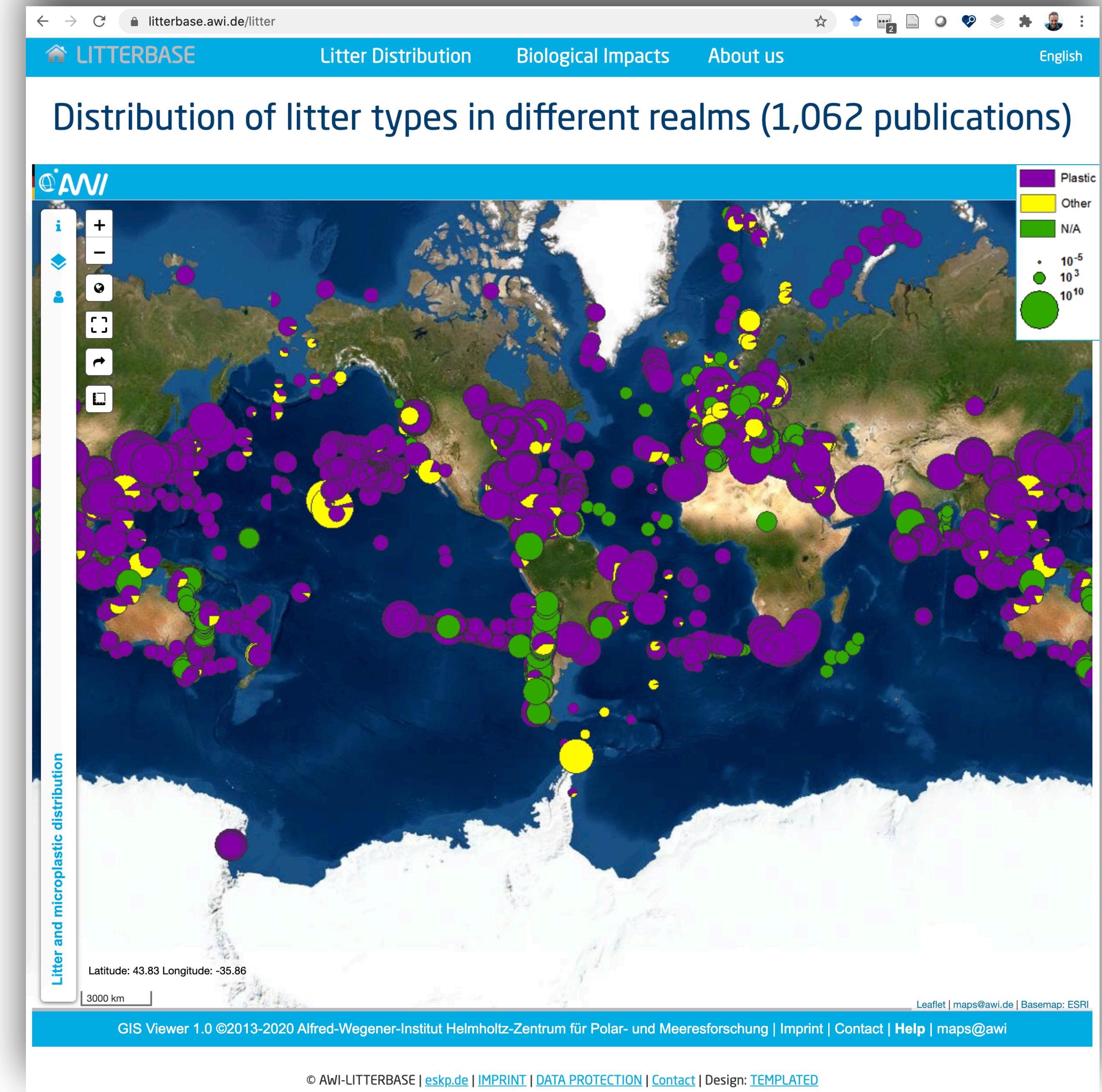


Closing the budget in the Mediterranean using data assimilation



~50% beached | ~40% sunk | ~10% still afloat

The LITTERBASE as a start



Can we automatically search the literature?

(ocean OR sea) AND
(plastic OR microplastic
OR debris OR litter)

Start from Richard Thompson's seminal
'Lost at Sea: Where is all the Plastic?'

Returned **87,360** articles,
including many on plastic
shipping containers,
ocean liners etc

Richard's article is cited **785** times; citation
network at depth=4 yields **96,999** articles(!)

Select only those with 'Environmental
Studies' or 'Oceanography' ASJC code
and search for tables that have caption
containing "plastic"

Returns **759** tables, from **444** articles

But parsing these tables is not trivial

Challenge is different units in table

- Not only measured quantities (#/km², g/m², #/l, #/m³, etc)
- Also different georeferencing systems

Table A4. Neuston station positions, plastic particle abundance and dry mass for Cruise CalCOFI-0607, off southern California, July 2006.

Line	Station	Latitude	Longitude	Particle abundance (no./m ³)	Plastic dry mass (mg/m ³)
76.7	49	35.09	-120.78	0.0150	0.0045
76.7	51	35.02	-120.92	0.0000	0.0000
76.7	55	34.89	-121.20	0.0000	0.0000
76.7	60	34.72	-121.55	0.0258	0.0245
76.7	70	34.38	-122.25	0.0000	0.0000
76.7	80	34.06	-122.94	0.0313	0.0047
76.7	90	33.72	-123.64	0.0759	0.4446

Table 1. Sampling stations, their distance from the coast, coordinates and microplastic concentrations.

Stations	Distance (nautical. mile)	Longitude (N)	Latitude (E)	MP concentration ($\times 10^3$ par. m ⁻³)	
				November	February
G2	2		41° 01' 51"	38° 38' 14"	0.31
T2	2		41° 10' 24"	39° 25' 23"	1.02
C2	2		40° 59' 44"	40° 14' 27"	1.24
P2	2		41° 14' 27"	40° 54' 32"	0.48
K2	2		41° 31' 48"	41° 30' 29"	0.39
G8	8		41° 06' 07"	38° 34' 39"	2.59
					1.01

Challenges

1. How to automatically identify peer-reviewed papers that contain data on observations of plastic in the ocean and on beaches?
2. How to automatically parse that data into a database?
3. How to geo-tag the plastic observations to location and time of sampling?