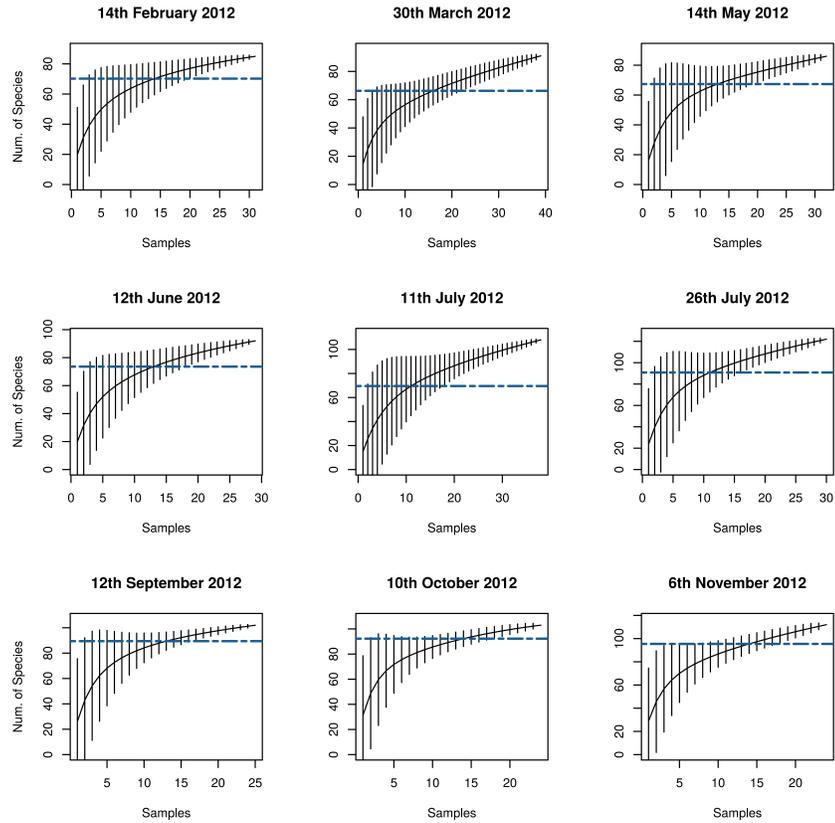
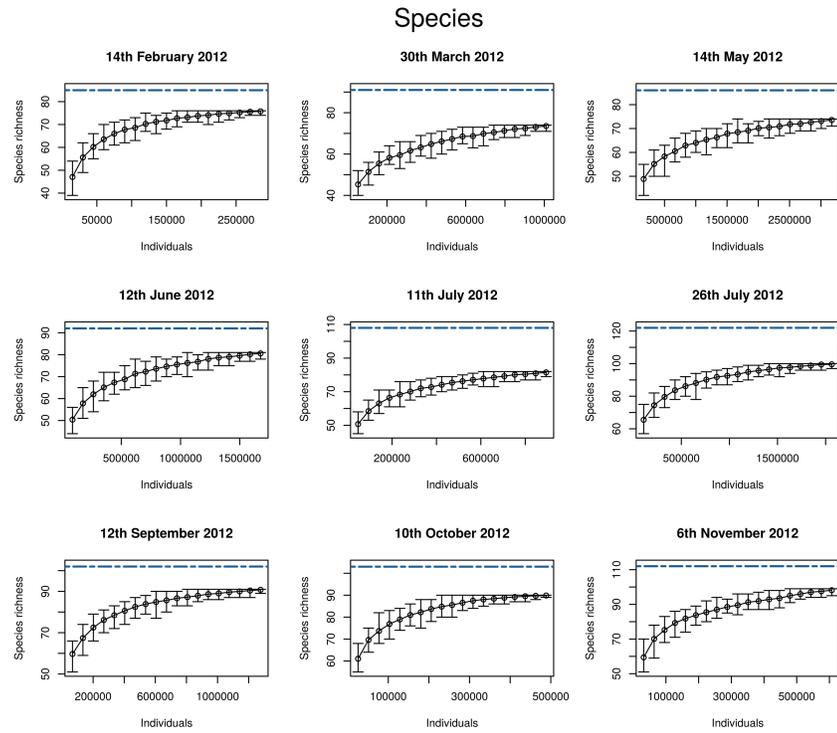


## Appendix S1. Species accumulation curves



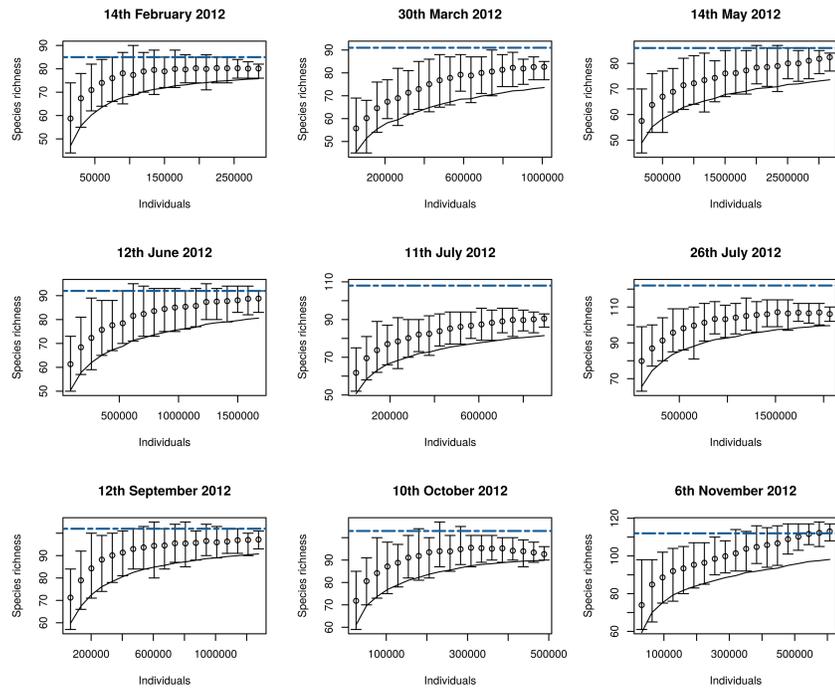
**Figure S1 Sample-based species accumulation curves.** Species accumulation curves (SAC) created using the number of samples collected at the same sampling station for the nine sampling cruises. SACs represent the number of species observed against the sampling-effort, in this case number of samples. The total number of samples is the sum of bottle and net samples used for collecting microplankton. The horizontal blue line limits the species richness observed counting bottle samples, where the number of individual of each species was noted down (to an abundance-level). This value was used as number of species observed,  $S_{obs}$ . The rest of the species were collected using net samples, which were noted as present or absent (to an incidence-level). The total amount of species (bottle and net) is referred to as the maximum of species richness observed,  $S_{mobs}$ .

## Appendix S2. Estimators performance in relation to the individuals counted



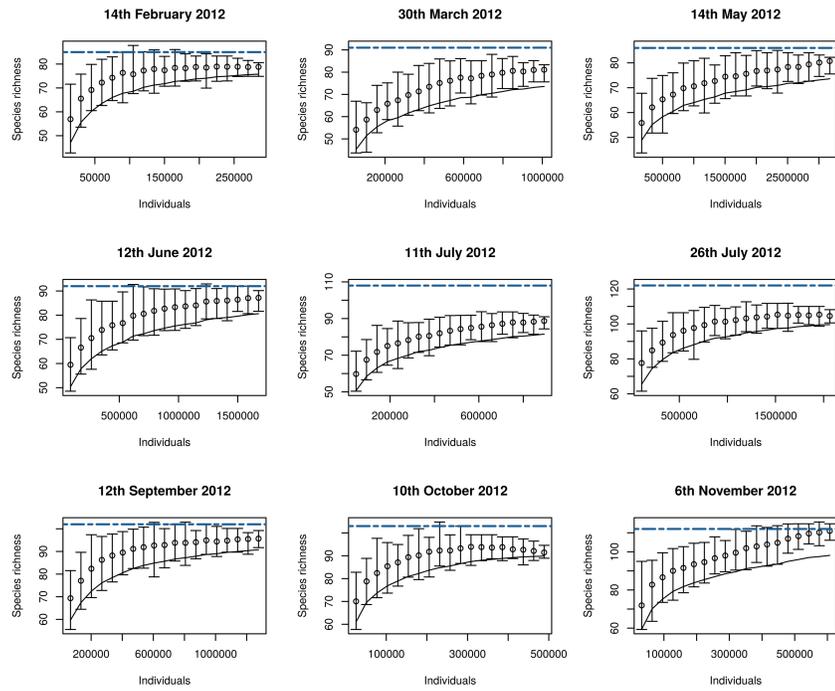
**Figure S2 Species observed.** Mean of species observed in relation to the individuals counted. Error bars show the variation in the 100 random subsamples over which non-parametric estimators were calculated.

### Jack1 (Ab.)



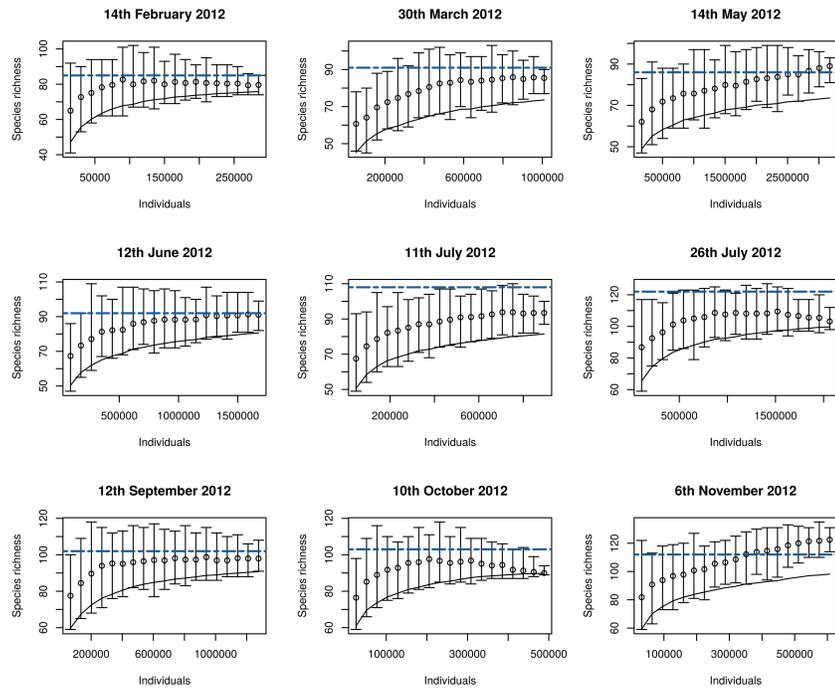
**Figure S3 Jackknife 1.** Abundance-based Jackknife 1 mean of 100 replicates of non-parametric estimates with error bars in relation to the individuals counted. The species accumulation curve is represented by a solid black line and the horizontal blue dashed line indicates  $S_{obs}$ .

### Jack1 (In.)



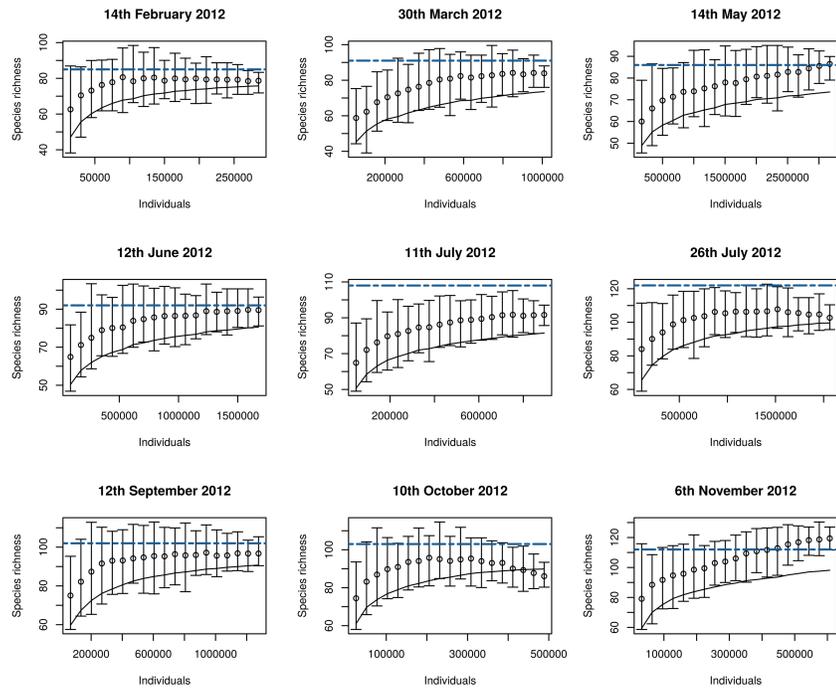
**Figure S4 Jackknife 1.** Incidence-based Jackknife 1 mean of 100 replicas of non-parametric estimates with error bars in relation to the individuals counted. The species accumulation curve is represented by a solid black line and the horizontal blue dashed line indicates  $S_{mobs}$ .

## Jack2 (Ab.)

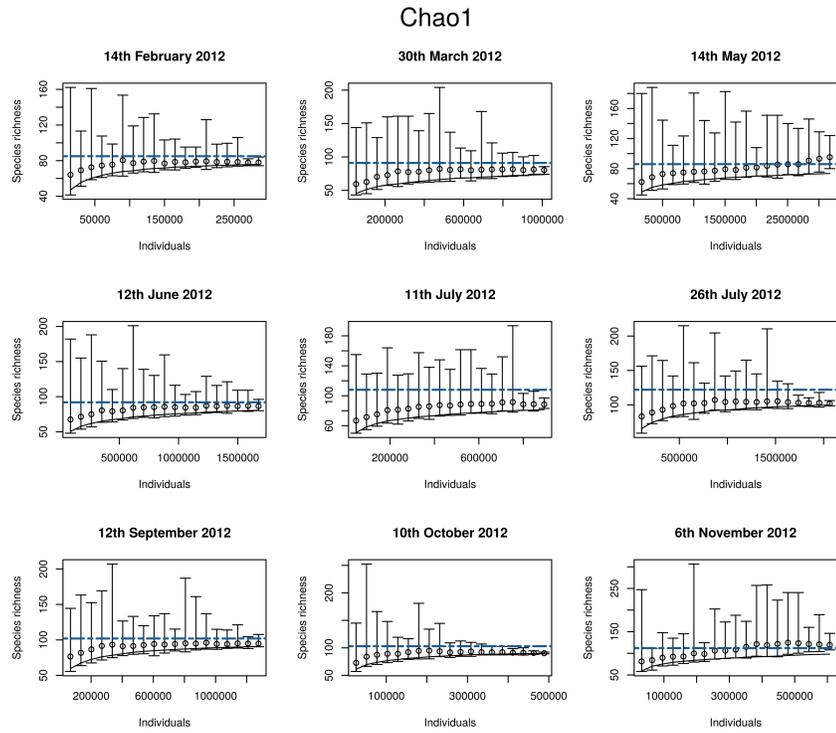


**Figure S5 Jackknife 2.** Abundance-based Jackknife 2 mean of 100 replicas of non-parametric estimates with error bars in relation to the individuals counted. The species accumulation curve is represented by a solid black line and the horizontal blue dashed line indicates  $S_{mobs}$ .

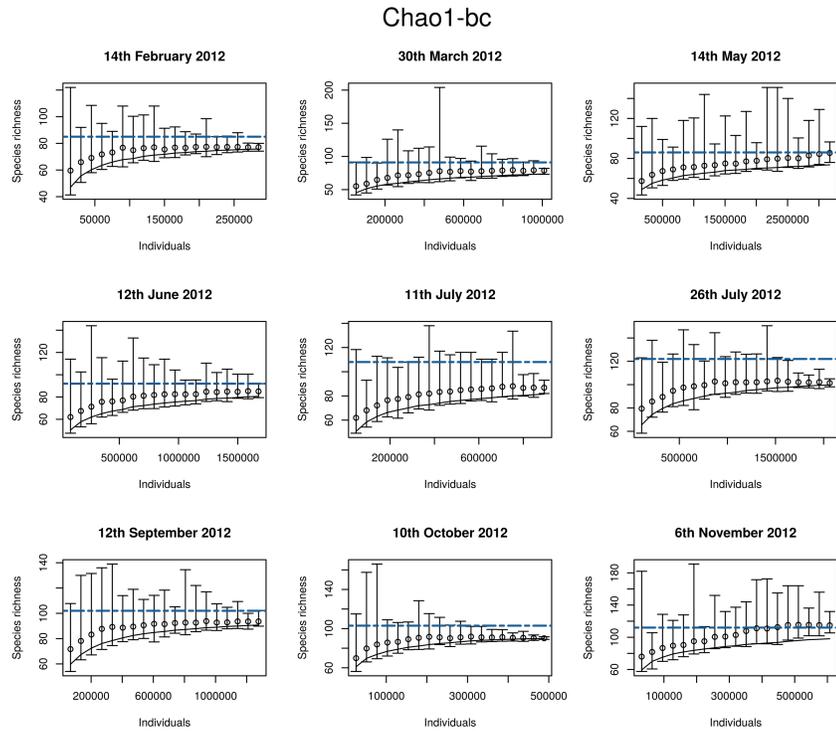
### Jack2 (In.)



**Figure S6 Jackknife 2.** Incidence-based Jackknife 2 mean of 100 replicas of non-parametric estimates with error bars in relation to the individuals counted. The species accumulation curve is represented by a solid black line and the horizontal blue dashed line indicates  $S_{mobs}$ .

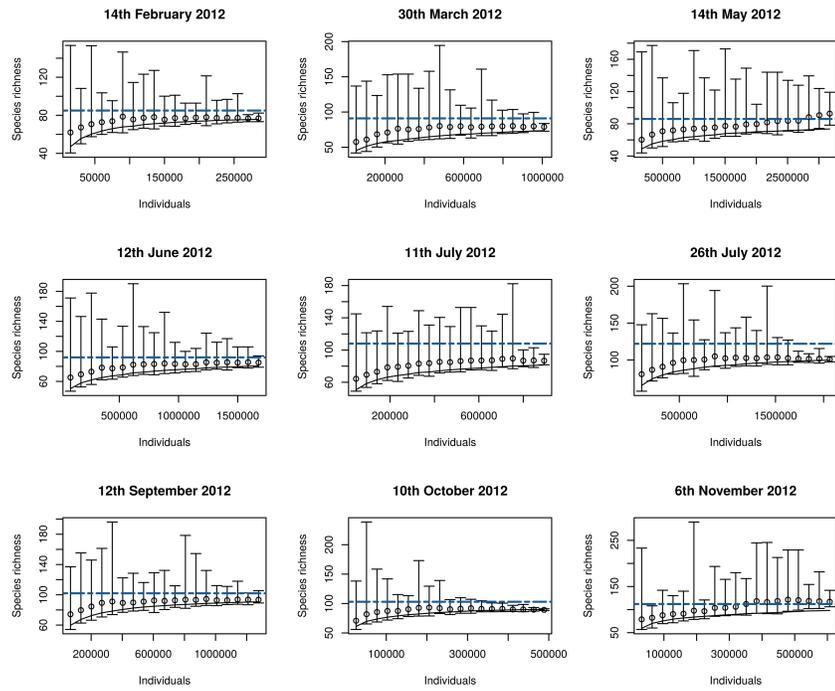


**Figure S7 Chao1.** Chao1 mean of 100 replicas of non-parametric estimates with error bars in relation to the individuals counted. The species accumulation curve is represented by a solid black line and the horizontal blue dashed line indicates  $S_{mobs}$ .

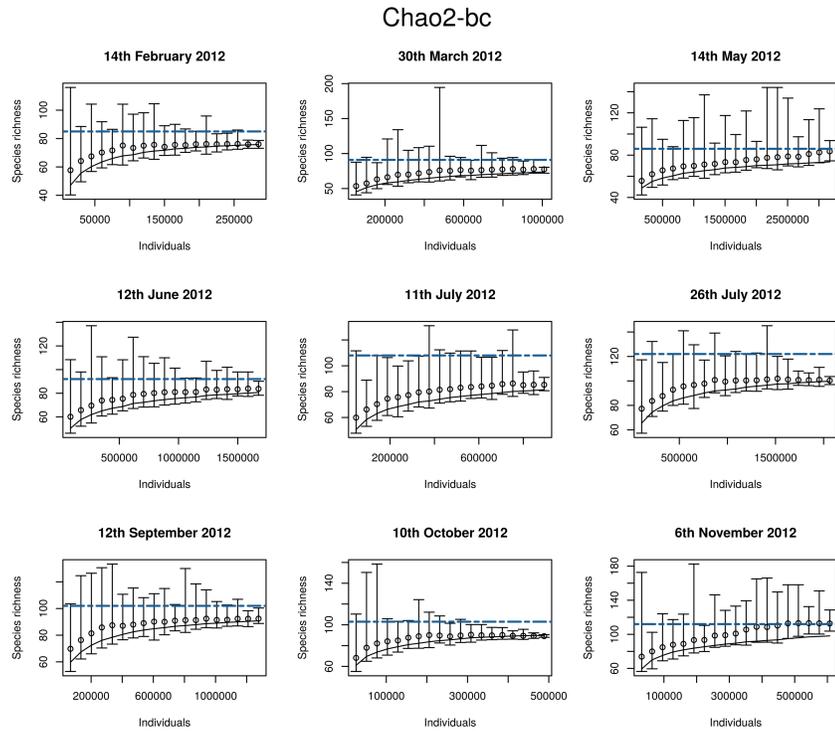


**Figure S8 Chao1-bc.** Chao1-bc mean of 100 replicas of non-parametric estimates with error bars in relation to the individuals counted. The species accumulation curve is represented by a solid black line and the horizontal blue dashed line indicates  $S_{obs}$ .

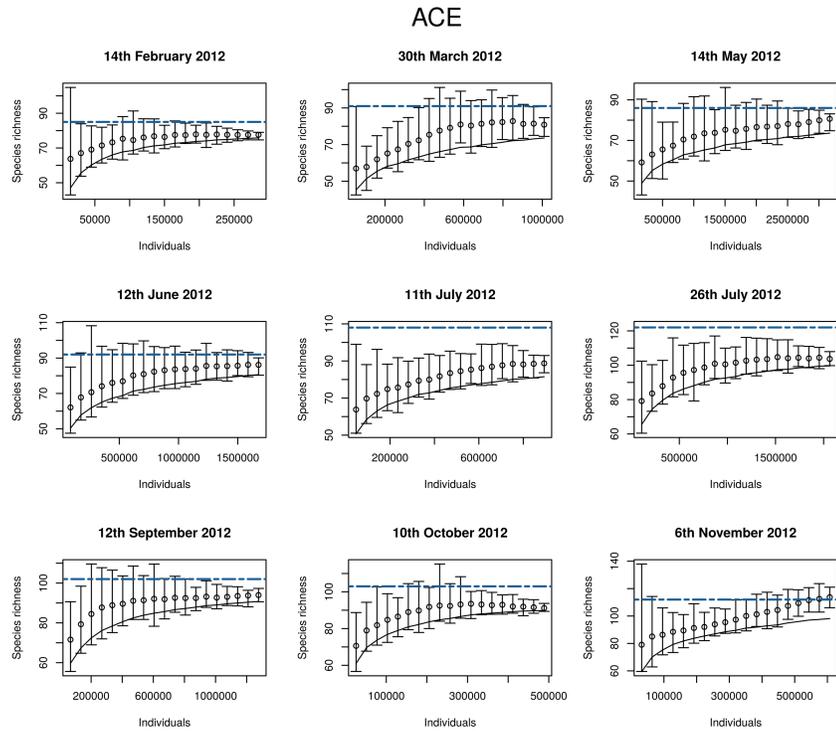
## Chao2



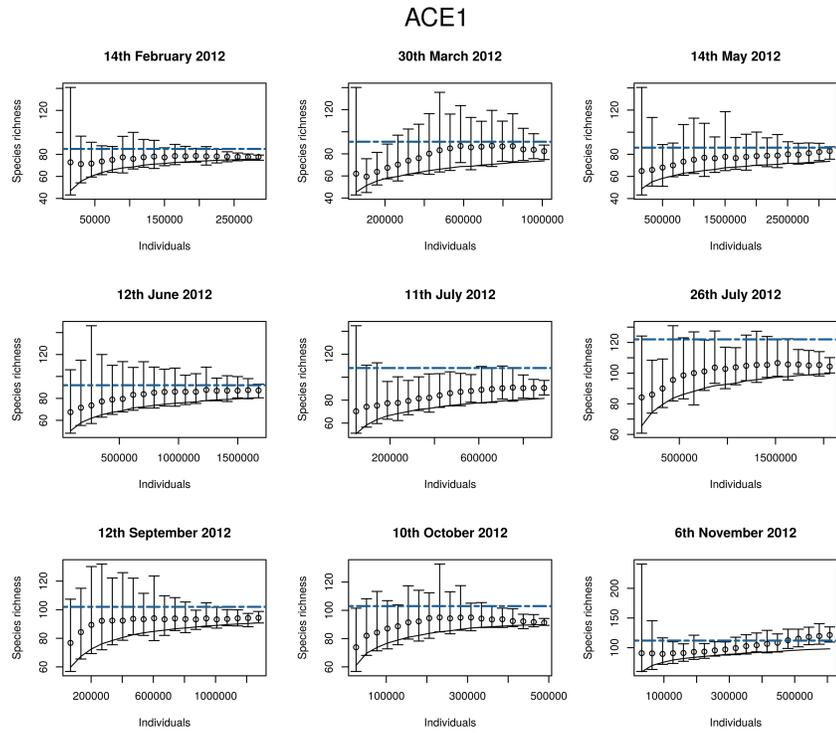
**Figure S9 Chao2.** Chao2 mean of 100 replicas of non-parametric estimates with error bars in relation to the individuals counted. The species accumulation curve is represented by a solid black line and the horizontal blue dashed line indicates  $S_{obs}$ .



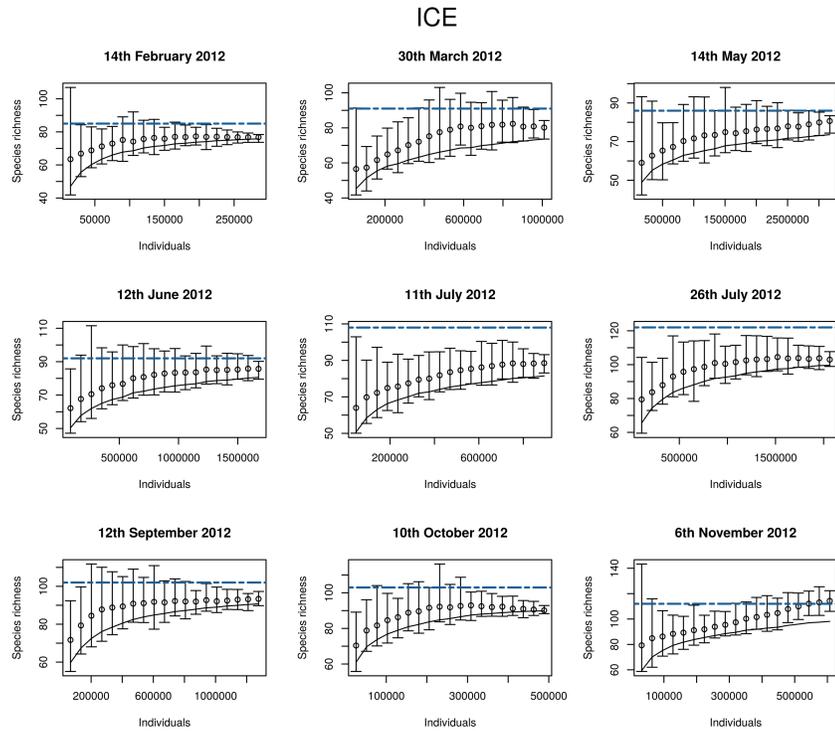
**Figure S10 hao2-bc.** Chao2-bc mean of 100 replicas of non-parametric estimates with error bars in relation to the individuals counted. The species accumulation curve is represented by a solid black line and the horizontal blue dashed line indicates  $S_{obs}$ .



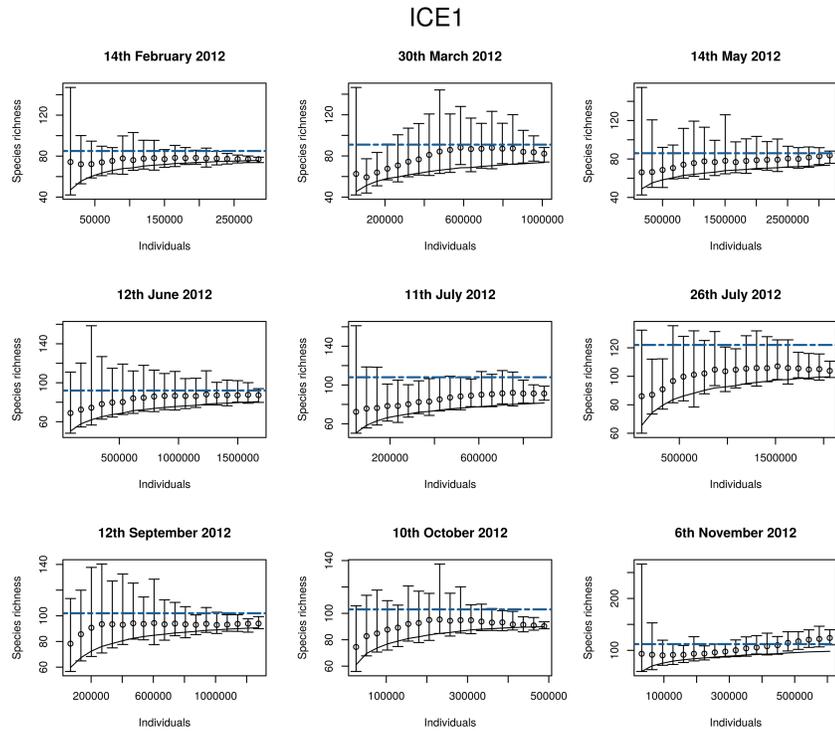
**Figure S11 ACE.** ACE mean of 100 replicas of non-parametric estimates with error bars in relation to the individuals counted. The species accumulation curve is represented by a solid black line and the horizontal blue dashed line indicates  $S_{obs}$ .



**Figure S12 ACE1.** ACE1 mean of 100 replicas of non-parametric estimates with error bars in relation to the individuals counted. The species accumulation curve is represented by a solid black line and the horizontal blue dashed line indicates  $S_{obs}$ .

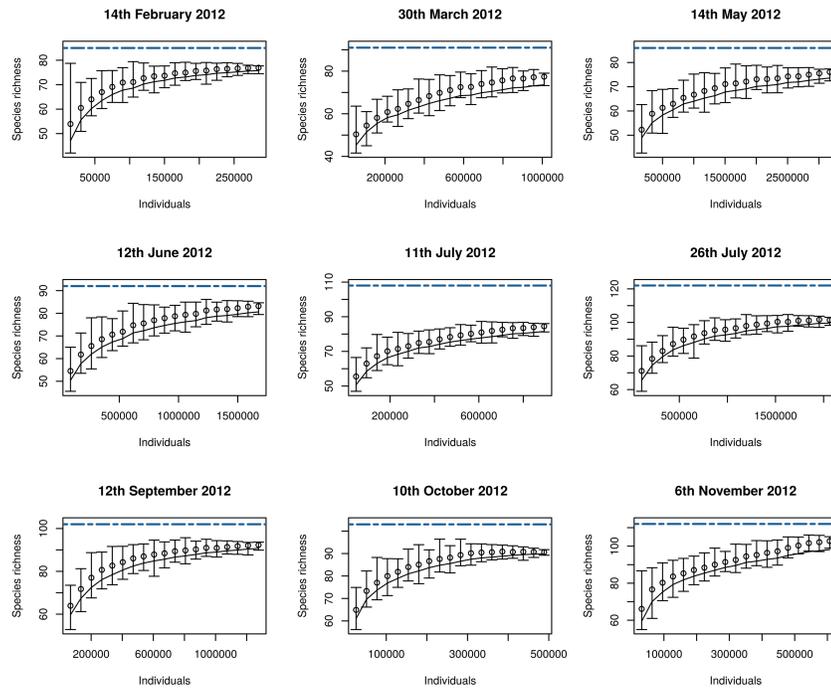


**Figure S13 ICE.** ICE mean of 100 replicas of non-parametric estimates with error bars in relation to the individuals counted. The species accumulation curve is represented by a solid black line and the horizontal blue dashed line indicates  $S_{mobs}$ .



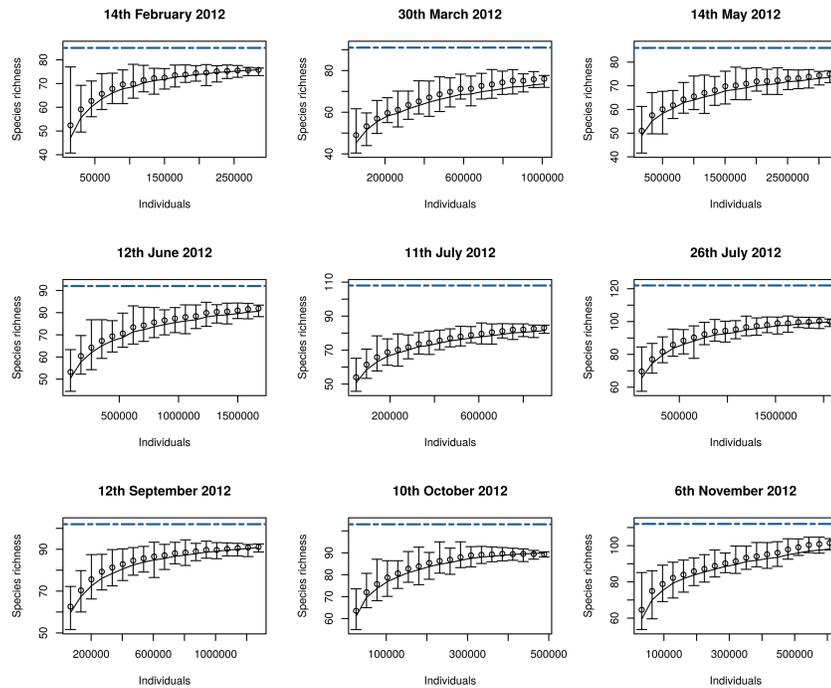
**Figure S14 ICE1.** ICE1 mean of 100 replicas of non-parametric estimates with error bars in relation to the individuals counted. The species accumulation curve is represented by a solid black line and the horizontal blue dashed line indicates  $S_{obs}$ .

### Homogeneous Model (Ab.)



**Figure S15 Homogeneous Model.** Abundance-based Homogeneous Model mean of 100 replicas of non-parametric estimates with error bars in relation to the individuals counted. The species accumulation curve is represented by a solid black line and the horizontal blue dashed line indicates  $S_{mobs}$ .

### Homogeneous Model (In.)



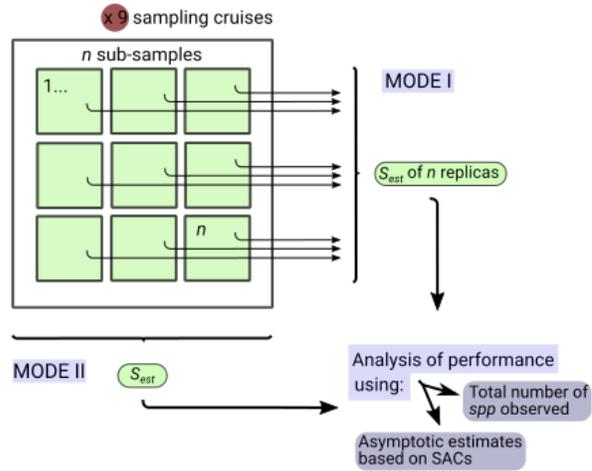
**Figure S16 Homogeneous Model.** Incidence-based Homogeneous Model mean of 100 replicas of non-parametric estimates with error bars in relation to the individuals counted. The species accumulation curve is represented by a solid black line and the horizontal blue dashed line indicates  $S_{mobs}$ .

## Appendix S3. An analysis of non-parametric asymptotic estimations performance using multiple replicates and the total abundance of species

In the article we described how classical non-parametric species-richness estimators (Jackknife and Chao's families and, coverage-based estimators) performed for a collection of microplankton of Ría de Vigo (NW Iberian Peninsula) (DOI: 10.6084/m9.figshare.4555456). Non-parametric asymptotic estimators were applied multiples times to samples that went on augmenting in individuals (See *Methods* section). Our intention was to test if these estimators are useful for common field effort (volume). We paid more attention to how estimates performed using a reduced fraction of individuals. Moreover, our objective was to fairly compare species-richness estimations for samples of the same sampling station but collected at different seasons. Thus, sample replicates were obtained rarefying the community to smaller sizes in each case, following what Heck *et al.* 1975 and Hulbert 1971 had described, using `vegan` R library. Then estimates' performance was evaluated as individuals went on increasing.

For comparison with the previous methodology, in this appendix we show a summary of non-parametric asymptotic estimates that were obtained treating samples another way. In the first case, we used non-parametric estimators using around 10 bottle sample replicates of equal volume, and that were collected during the same sampling cruise (from here on, *Mode I*). Replicates' volume was decided according the amount of photosynthetic biomass in the seawater in each cruise, as Lund *et al.* 1958 and Utermöhl 1958 indicated. Depending on when those samples were collected, the volume of replicas varied between 5 and 50 mL. In the other case we used the total species abundance of approximately 10 bottle samples replicas of similar volume, adding all of them up (from here on, *Mode II*). In both cases non-parametric asymptotic estimators performance was evaluated as was described in *Methodology* section in the article.

Evaluating estimator performance requires a value to compare with, an approximation of the true species richness value ( $S_{true}$ ). In the main text the maximum of species collected with either bottles and net,  $S_{mobs}$ , is accepted as a reference of the microplankton species richness provided that sampling effort (seawater volume usually inspected) was multiplied by ten. For comparison, an estimate of the maximum of species present in samples (bottle + net) obtained calculating the asymptote of the species accumulation curve (SAC). Mode I and Mode II results were compared with the maximum of species collected in samples ( $S_{mobs}$ ) and with those asymptotic estimates as well.



The procedure for calculating the asymptote, first, and analysis of estimates using Mode I and Mode II species abundances, second, are shown and discussed in the following sections.

### 0.1 Obtaining the asymptote of the species accumulation curve

Species-accumulation curves (SAC) were built describing how many species are observed as seawater volume inspected increases. We obtained the asymptotes of this SACs using `mmSAR` R library. The logistic regression of SACs was done considering the *Power Law*, *Exponential*, *Negative Exponential*, *Monod*, *Rational*, *Logistic*, *Lomolino* and *Weibull* functions. How distributions fitted the SAC was evaluated using the corrected Akaike's Information Criterion (AICc). Two of the distributions that fitted more consistently all the SACs were selected and used in following sections. Their maximum of species richness estimates (estimations of  $S_{true}$  at a reference point) are an alternative to  $S_{mobs}$  and were used for evaluating non-parametric estimators proficiency. Consider, though, that the *Exponential* function grows in such a way that is not to be expected in a real case.

Cruise	Num Samples	Volume (mL)	$S_{mobs}$	AICc Exponential	AICc Weibull
14 <sup>th</sup> Feb	15	35800	85	52.61	48.44
30 <sup>th</sup> Mar	17	22100	90	50.01	44.15
14 <sup>th</sup> May	14	24084	86	51.28	52.19
12 <sup>th</sup> June	14	25016	92	54.41	35.38
11 <sup>th</sup> Jul	12	23622	108	30.90	35.61
26 <sup>th</sup> Jul	13	26435	122	32.82	31.65
12 <sup>th</sup> Sep	14	27031	102	47.57	52.01
10 <sup>th</sup> Oct	15	26424	103	44.88	40.11
6 <sup>th</sup> Nov	15	25527	111	75.91	52.97

**Table S1 AICs weights for SACs.** AICs values of Exponential and Weibull functions estimations are listed provided that they performed better in general rather than the other functions considered. Aside of AICc values, the table includes a summary of number of samples inspected, number of subsamples or replicates, the total volume of seawater inspected (in mL) and the maximum number of species observed. Exponential and Weibull estimations were used for evaluation of non-parametric estimators for Mode I and Mode II data.

## 0.2 Mode I. Estimators performance based on multiple sample replicates

In the following tables non-parametric estimators performance is evaluated using the maximum of species observed in bottle and net samples ( $S_{mobs}$ ), and Exponential and Weibull functions estimates as a reference. Non-parametric estimates are studied at Mode I, that were obtained with multiple subsamples of each cruise.

	$S_{mobs}$				Exponential				Weibull				Sp SD
	SME		SRMSE		SME		SRMSE		SME		SRMSE		
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	
Homogenous Model	0.45	0.05	0.63	0.18	0.62	0.05	0.63	0.18	0.45	0.05	0.63	0.18	4.60
Chao1	0.35	0.10	1.41	0.73	0.55	0.06	1.41	0.73	0.35	0.10	1.41	0.73	11.54
Chao1-bc	0.40	0.07	0.95	0.34	0.59	0.05	0.95	0.34	0.40	0.07	0.95	0.34	7.41
ACE	0.37	0.07	1.00	0.25	0.56	0.06	1.00	0.25	0.37	0.07	1.00	0.25	7.95
ACE-1	0.30	0.09	1.63	0.46	0.51	0.07	1.63	0.46	0.30	0.09	1.63	0.46	13.75
1st order jackknife	0.40	0.06	0.63	0.15	0.59	0.05	0.63	0.15	0.40	0.05	0.63	0.15	4.86
2nd order jackknife	0.34	0.07	0.86	0.17	0.54	0.05	0.86	0.17	0.34	0.07	0.86	0.17	6.96

**Table S2 Abundance-based non-parametric estimators.** Mean and standard deviation of bias and precision measures of abundance-based estimators. Estimators were used for Mode I data, which are around ten sample replicates of equal volume. Estimates were compared with the maximum of species richness observed in bottle and net samples ( $S_{mobs}$ ), and species estimates of Weibull and Power Law regression.

	Max Sp Obs				Power Law				Weibull				sp SD
	SME		SRMSE		SME		SRMSE		SME		SRMSE		
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	
Homogenous Model	0.50	0.05	0.54	0.18	0.65	0.04	0.54	0.18	0.50	0.05	0.54	0.18	3.80
Chao2	0.46	0.06	0.82	0.38	0.62	0.04	0.82	0.38	0.46	0.06	0.82	0.38	6.04
Chao2-bc	0.48	0.06	0.65	0.22	0.64	0.04	0.65	0.22	0.48	0.06	0.65	0.22	4.67
ICE	0.30	0.09	1.35	0.32	0.51	0.06	1.35	0.32	0.30	0.09	1.35	0.32	11.42
ICE-1	0.01	0.18	3.15	0.78	0.31	0.11	3.15	0.78	0.01	0.18	3.15	0.78	31.77
1st order jackknife	0.48	0.05	0.55	0.16	0.64	0.04	0.55	0.16	0.48	0.05	0.55	0.16	3.97
2nd order jackknife	0.48	0.05	0.55	0.16	0.64	0.04	0.55	0.16	0.48	0.05	0.55	0.16	3.97

**Table S3 Incidence-based non-parametric estimators.** Mean and standard deviation of bias and precision of abundance-based estimators. Estimators were used for Mode I data, which are around ten sample replicates of equal volume for each sample. Estimates were compared with the maximum of species richness observed in bottle and net samples ( $S_{mobs}$ ), and estimations obtained with a Weibull and Power Law regression.

### 0.3 Mode II. Estimates performance based on total abundances of species

In the following tables non-parametric estimators performance is evaluated using the maximum of species observed in bottle and net samples ( $S_{mobs}$ ), and Exponential and Weibull functions estimates as a reference. Non-parametric estimates are studied at Mode II so there is just one single estimate for each sample.

	$S_{mobs}$		Exponential		Weibull	
	SME	SRMSE	SME	SRMSE	SME	SRMSE
Homogenous Model	1.31	1.09	5.04	1.09	1.31	1.09
Chao1	0.79	1.25	4.62	1.25	0.80	1.25
Chao1-bc	1.01	1.21	4.80	1.21	1.02	1.21
ACE	0.94	1.25	4.74	1.25	0.95	1.25
ACE-1	0.75	1.42	4.59	1.42	0.76	1.42
1st order jackknife	0.75	1.17	4.59	1.17	0.76	1.17
2nd order jackknife	0.55	1.28	4.42	1.28	0.56	1.28

**Table S4 Abundance-based non-parametric estimators performance.** Bias and precision of each abundance-based estimate in Mode II.

	$S_{mobs}$		Exponential		Weibull	
	SME	SRMSE	SME	SRMSE	SME	SRMSE
Homogenous Model	1.44	1.10	5.15	1.10	1.45	1.10
Chao2	0.96	1.23	4.76	1.23	0.97	1.23
Chao2-bc	1.16	1.20	4.92	1.20	1.17	1.20
ICE	0.99	1.28	4.78	1.28	0.99	1.28
ICE-1	0.74	1.49	4.58	1.49	0.75	1.49
1st order jackknife	0.92	1.17	4.72	1.17	0.93	1.17
2nd order jackknife	0.76	1.29	4.60	1.29	0.77	1.29

**Table S5 Incidence-based non-parametric estimators performance.** Bias and precision of each incidence-based estimate in Mode II.

## Appendix S4. Rarefaction and extrapolation of species richness

Here we detail how the species richness of the microplankton community varied over time employing a rarefaction and extrapolation (R/E) method. We used the methodology described by Chao & Jost 2012, Colwell *et al.* 2012 and Chao *et al.* 2014 and making use of iNEXT R package (version 2.0.12). See Hsieh, T. C. *et al. Methods Ecol Evol* 7, 1451–1456 (2016) for more a detailed summary of this method and for more information about that package. We paid attention to the species richness, understood as the Hill number of order  $q = 0$ , and used the microplankton species abundances of bottle samples as a reference.

In one case, we considered samples separately and applied the R/E to extrapolate each sample to the double of individuals (See Fig. S17 and S18). In the other case, we compared the species richness of samples considering all of them at the same time (See Fig. 4 and Fig. S19). R/E curve was obtained using the maximum number of individuals of the more individual-rich sample as a reference or standard sample. All microplankton samples were interpolated or extrapolated to that standard. Estimates were obtained always with a 95% confidence interval.

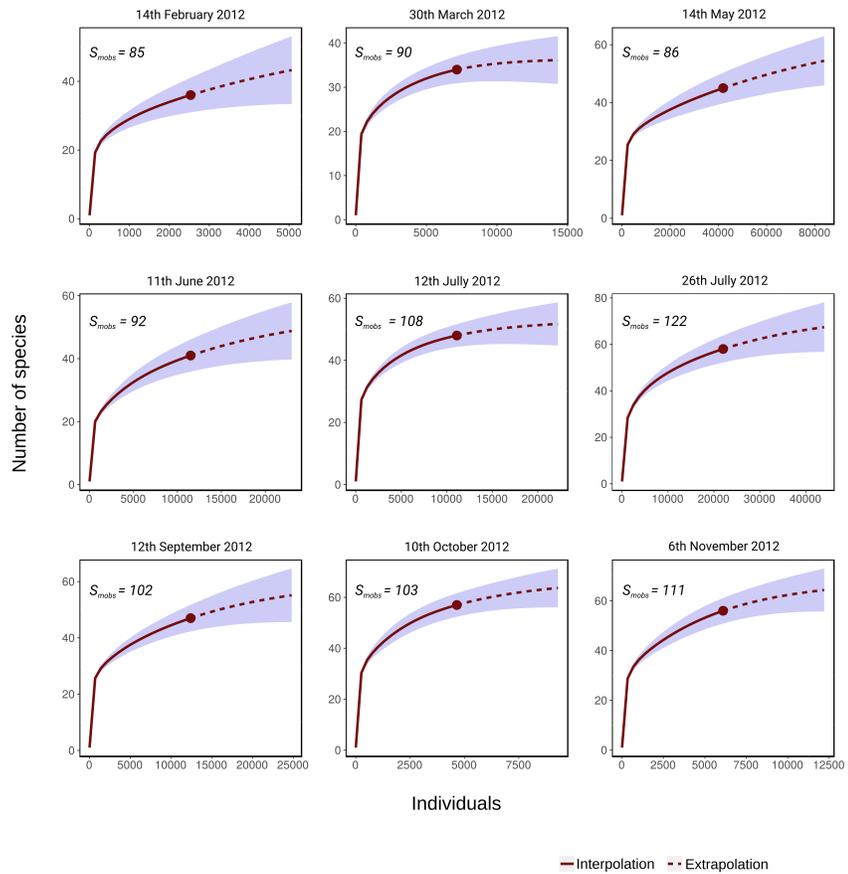
In both cases we did the R/E at two levels of sample completeness: i) using the species abundance observed in around ten replicas of the same volume (Mode I, Fig. S17 and Fig 4) and ii) with the total abundance of species observed in bottle samples (Mode II, Fig. S18 and Fig. S19). The R/E curve where Mode I samples are considered together is included in the main text (See Fig. 4).

Date	n	$S_{obs}$	$f_1$	$f_2$	SC	Estimator	Estimator SE	95% Lower	95% Upper	$S_{mobs}$
14 <sup>th</sup> Feb	2532	36	9	2	0.9999	56.24	20.18	39.97	139.12	85
30 <sup>th</sup> Mar	7160	34	5	5	0.9999	36.50	2.96	34.40	49.64	90
14 <sup>th</sup> May	41923	45	12	3	1.0000	69.00	20.20	50.72	145.72	86
12 <sup>th</sup> Jun	11496	41	11	4	1.0000	56.12	12.47	44.69	103.04	92
11 <sup>th</sup> Jul	11109	48	7	5	0.9999	52.90	4.84	48.97	72.69	108
26 <sup>th</sup> Jul	22000	58	14	6	1.0000	74.33	11.71	62.62	115.69	122
12 <sup>th</sup> Sep	12404	47	12	5	1.0000	61.40	11.18	50.75	102.33	102
10 <sup>th</sup> Oct	4651	57	11	6	1.0000	67.08	8.00	59.56	96.66	103
6 <sup>th</sup> Nov	6100	56	14	8	0.9998	68.25	8.59	59.54	98.33	111

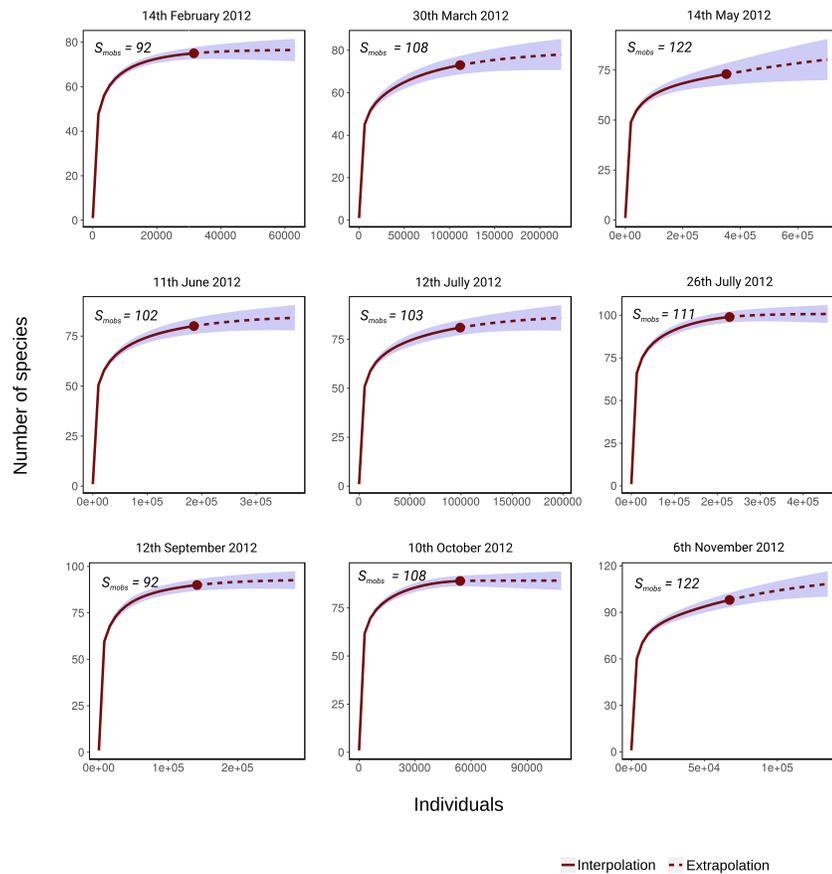
**Table S6 Summary of rarefaction and extrapolation for Mode I.** Rarefaction and extrapolation was applied to reference samples of Mode I. This table summarises the species observed in subsamples ( $S_{obs}$ ), number of singletons ( $f_1$ ) and doubletons ( $f_2$ ), the estimated sample coverage ( $SC$ ), the extrapolation to double size of individuals with a confidence interval of 95% and the maximum number of species observed in bottles and nets ( $S_{mobs}$ ).

Date	n	$S_{obs}$	$f_1$	$f_2$	SC	Estimator	Estimator SE	95% Lower	95% Upper	$S_{mobs}$
14 <sup>th</sup> Feb	31560	75	4	5	0.9964	76.60	2.16	75.22	86.79	85
30 <sup>th</sup> Mar	111805	73	9	6	0.9993	79.75	5.88	74.55	102.45	90
14 <sup>th</sup> May	351612	73	9	2	0.9997	93.25	20.19	76.97	176.16	86
12 <sup>th</sup> Jun	185593	80	8	6	0.9990	85.33	4.93	81.14	104.88	92
11 <sup>th</sup> Jul	99028	81	9	6	0.9994	87.75	5.88	82.55	110.45	108
26 <sup>th</sup> Jul	228630	99	6	10	0.9994	100.80	2.07	99.30	109.88	122
12 <sup>th</sup> Sep	141405	90	6	6	0.9990	93.00	3.24	90.53	106.81	102
10 <sup>th</sup> Oct	54103	89	2	10	0.9976	89.20	0.53	89.01	92.41	103
6 <sup>th</sup> Nov	67244	98	15	6	0.9977	116.75	13.08	103.46	162.42	111

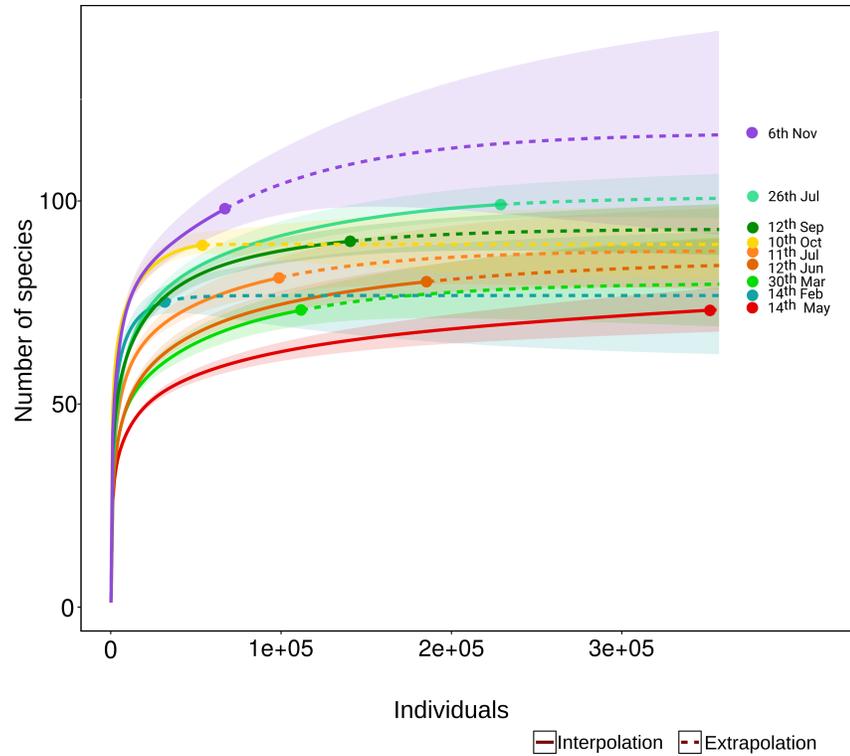
**Table S7 Summary of rarefaction and extrapolation for Mode II.** Rarefaction and extrapolation was applied to the single reference sample of Mode II. This table summarises the total number of species observed in the sample ( $S_{obs}$ ), the number of singletons ( $f_1$ ) and doubletons ( $f_2$ ), the estimated sample coverage ( $SC$ ), the extrapolation to double size of individuals with a confidence interval of 95% and the maximum number of species observed in bottles and nets ( $S_{mobs}$ ).



**Figure S17 Rarefaction and extrapolation curves.** R/E curves for each sample, where the solid line is rarefaction and the dashed line is extrapolation. Each sample was composed of subsamples replicas of equal size (Mode I, e.g. always replicas of 5 mL, or another value). The reference point for extrapolation was the double of individuals of each sample. 95% confidence intervals were represented. For comparison, the maximum number of species collected in bottle and net samples,  $S_{mobs}$ , is indicated at the top of each plot.



**Figure S18 Rarefaction and extrapolation curves.** R/E curves for each sample, where the solid line is rarefaction and the dashed line is extrapolation. No replica was discarded. In each sample every sample replica was considered independently of its size (Mode II). The extrapolation reference point was the double of individuals of each sample. 95% confidence intervals were represented. For comparison, the maximum number of species collected in bottle and net samples,  $S_{mobs}$ , is indicated at the top of each plot.



**Figure S19 Rarefaction and extrapolation curves.** R/E curve, where the solid line is rarefaction and the dashed line is extrapolation, that was obtained using the totality of replicates of each sample (See Mode II in Appendix S3). R/E was calculated using as an endpoint reference the most individual-rich sample, which was 14<sup>th</sup> May sample. 95% confidence intervals were represented. See Fig. 4 for a comparison with Mode I data.