TITLE

The ZooScan and the ZooCAM zooplankton imaging systems are intercomparable: A benchmark on the Bay of Biscay zooplankton.

Supplementary Materials

NB-SS slopes by stations

For both instruments, the steeper slopes were found along the coast meaning that the proportion of small individuals is higher than that of large individuals in coastal areas. On the other hand, the flatter slopes were seen in the middle of the continental shelf north to 45°N meaning that the large organisms dominated the zooplankton community in this area. (Fig. S1a and b).

Abnormal differences between instruments test procedure

The procedure was implemented as follow: (i) for each variable we computed the pair-wise ratios $R1 = V_{ZooScan} / V_{ZooCAM}$, and $R2 = V_{ZooCAM} / V_{ZooScan}$, by station. R1 or R2 being close to 1 mean that both instruments values are similar. On the contrary, $R_1 >>> 1$ or $R_2 <<< 1$ mean that the ZooScan value for the considered variable at the considered station or taxa was much larger than the ZooCAM's, or vice versa. Then, (ii), we assembled a vector combining all the $R_1 > 1$ and $R_2 > 1$ (i.e. $[R_1^+R_2^+]$ with n = 61 for stations), and plotted the distributions of those ratios as boxplots, for each above-mentioned variables. Boxplots enable the visualization of outliers within a distribution. And (iii) the outlier thresholds were set as values below $T_{inf} = q1/(q3-q1)$ and above $T_{sup} = q3/(q3-q1)$, q1 and q3 being the 25^{th} and 75^{th} percentile values of the distribution, respectively, and T_{inf} and T_{sup} corresponded to the maximum extent of the whiskers in each boxplot. T_{inf} and T_{sup} are calculated to take into account the interquartile range of the assessed distributions. It follows from this method that outliers can be identified only above T_{sup}, as T_{inf} minimum theoretical value is one (the values of $R1^+$ and $R2^+$ equal one when ZooScan and ZooCAM values for the variable considered are equal). These outlier thresholds correspond to approximately ±2.7σ and 99.3 percent of the distribution not being outlier if the data are normally distributed. The outliers detected in the distributions of $[R_1R_2]$ finally enable the identification of the stations or taxa for which there was an abnormally large difference between ZooScan and ZooCAM values for the variable considered.

The boxplots enabling the identification of these stations and taxa are presented hereafter in Figs. S2 and S3.

Mean sizes and abundances spatial patterns

Fig. S4

Fig. S5

Analyses of total biovolumes

The analyses presented in the Assessment section were also applied to the biovolumes calculated by station and by taxa. The results are presented hereafter.

Biovolumes by stations

The total biovolumes calculated at each sampling station ranged from 180 $mm^3.m^{-3}$ to 1950 $mm^3.m^{-3}$ (mean ± sd: 735 ± 414 $mm^3.m^{-3}$) with the ZooScan data and from 83 $mm^3.m^{-3}$ to 2740 $mm^3.m^{-3}$ (mean ± sd: 890 ± 640 $mm^3.m^{-3}$) with the ZooCAM data. The total biovolumes showed six stations with significant differences between instruments values. Four out of the six stations had higher biovolumes with the ZooCAM (stations U0197 at 1.4°W – 44°N, U0203 at 1.7°W – 44.5°N, U0256 at 2.3°W 46.2°N and U0286 at 2.7°W – 47.2°N) and the two other had higher biovolumes with the ZooScan (station U0315, 4.9°W – 47.2°N and station U0320, 3.8°W – 46.5°N, Fig. S6). Those stations were excluded from the linear regression and the correlation test. Therefore, the linear regression was fitted to the data of 55 stations and had an estimated slope of 0.77 (p-value = 4.22×10^{-15}, R^2 = 0.68, Fig. S6a, Table S1). Furthermore, both datasets were significantly correlated with a coefficient of 0.83 (Table S1). Both spatial patterns exhibited higher biovolumes and higher differences in biovolumes computations at coastal stations (Fig. S6b and c).

Table S1: Linear regression parameters and Spearman correlation coefficient and p-value for the biovolumes calculated by sampling stations. Linear regressions were fitted to the data considering n stations, outliers excluded.

Variables by station	**n**	**Linear regression equation**	**R^2**	**p-value**	**Cor. coefficient**	**p-value**
Total biovolumes	55	y = 0.77x + 0.07	0.69	4.22×10^{-15}	0.83	4.22×10^{-15}

Biovolumes by taxa

The biovolume proportions (in %) of each taxa (n = 27) were compared over the whole Bay of Biscay continental shelf, between the two instruments. Only the taxa contributing at least 1% to the total biovolume were considered (Table 1). The Calanoida and the Calanidae were the two taxa contributing the most to the total biovolume for both instruments, followed by the Acartiidae. The Siphonophorae displayed a higher biovolume proportion within the ZooCAM data, contrary to the shrimp-like organisms for which the proportion was higher within the ZooScan data. Finally, the Actinopterygii and the Harosa contributed more than 1% to the total biovolume estimated with the ZooCAM which was not the case in the ZooScan data. Pairwise Wilcoxon tests run on the taxa biovolumes proportions calculated at each station showed no significant differences at any stations, indicating that the community composition was highly similar between both instruments.

The total biovolumes calculated by taxa were coherent between the ZooScan and the ZooCAM (Fig. S7). Three taxa for which the biovolumes were significantly different between the ZooScan and ZooCAM data were identified as the Annelida larvae which showed higher biovolumes with the ZooScan, and the Harosa and the Siphonophorae having higher biovolumes with the ZooCAM. A fitted linear regression excluding these taxa had an estimated slope of 1.01 (p-value = 1.19×10^{-14}, R^2 = 0.93) and the correlation test revealed a correlation coefficient of 0.96 (Table S2).

Table S2: Linear regression parameters and Spearman correlation coefficient and p-value for the biovolumes calculated by taxa. Linear regressions were fitted to the data considering n stations, without taking into account the taxa having biovolumes significantly different between both instruments.

Variables by taxa	**n**	**Linear regression equation**	**R^2**	**p-value**	**Cor. coefficient**	**p-value**
Biovolumes	24	y = 1.01x + 0.06	0.93	1.19×10^{-14}	0.97	1.19×10^{-14}

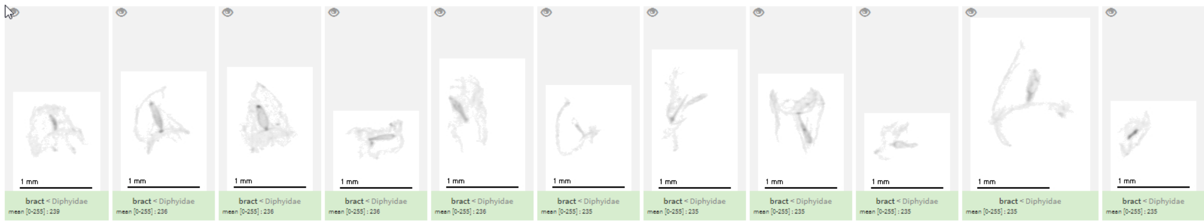
Comparison of the community spatial structure

The first three principal components of the biovolumes PCAs represented 80% and 75.6% of the biovolumes total variance in the ZooScan and ZooCAM datasets, respectively. The hierarchical clustering of grid cells' coordinates on the three first principal components revealed four clusters for both instruments which are presented with its most characteristic taxa in Fig. S8. The spatial patterns were highly similar between the ZooScan and the ZooCAM, exhibiting a coastal-offshore and a North-South gradients (Fig. S8a and c). For both instruments, the Appendicularia, the fish larvae, the Cladocera and the small copepods Temoridae, Harpacticoida and Poecilostomatoida were characteristic of coastal areas, while the large copepods Metridinidae marked the offshore clusters (Fig. S8b and d). Differences were noted for the gelatinous organisms, e.g. the Siphonophorae and the Hydrozoans, and the Harosa which indicated the southern coastal cluster of the ZooCAM data (Fig. S8d) but not that of the ZooScan data. On the contrary, the Bivalvia, the Annelida larvae and the

Chaetognatha indicated the southern coastal cluster only for the ZooScan data. The Thecosomata, the Cirripedia larvae and the Acartiidae marked the northern coastal cluster and the Euchaetidae characterised offshore clusters only for the ZooScan data (Fig. S8b).

Siphonophorae images

ZooScan Diphyidae Siphonophores bracts:



ZooCAM Diphyidae Siphonophores bracts:

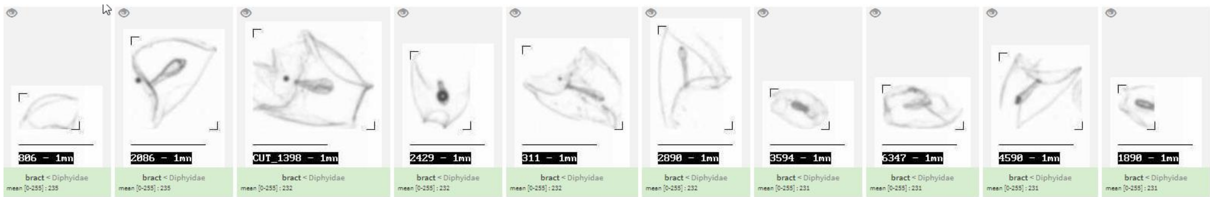


Fig. S9: Siphonophorae individual images captured by the ZooScan and the ZooCAM