Supplementary materials of the paper titled "The bottom mixed layer depth (BMLD) as an indicator of subsurface chlorophyll-a" on the method used to identify MLD and BMLD. The details describe the function "abmld.R".

MLD and BMLD have been identified developing an algorithm based on Chu and Fan (2011) framework to produce a method able to cope with various density profiles exhibiting a pycnocline (examples in Fig. A1 in Appendix A). The algorithm's sequence identifies the depth with the largest density difference between a mixed and a stratified layer using i) an adaptation of the maximum angle method (Chu and Fan, 2011) and ii) a cluster analysis on the density difference ($\Delta\rho_z = |\rho_z - \rho_{z+1}|$) (diagram of the algorithm in Fig. S1a). The method is designed to work with equal, high-resolution, intervals of density values (z) in the profiles. In order to distinguish MLD from BMLD, their selection is achieved by splitting the observations throughout the profile into two distinct groups, *Split1* and *Split2* (Fig. S1b and Fig. S1c), each one respectively used to identify MLD and BMLD. *Split1* includes the density values within the first observation close to the surface (z_1) and two measurement intervals δ (here 1 m) above BMLD ($z_{BMLD} - 2\delta$), while *Split2* extends from 2δ above the depth halfway through the ρ range ($0.5\Delta\rho = ((\rho_{max} - \rho_{min})/2) - 2$) up to the depth at which the total number of points from the surface to the bottom amounts up to 90% of the entire profile ($z_{0.9\Delta\rho} = 90\%$ of $_1^n z$). Since *Split1* is based on BMLD, the algorithm identifies MLD after BMLD.

For all depths between z_1 and $z_{0.9\Delta\rho}$, the angle φ has been measured at $z(x, y)$ (where x is the density and y is depth) between two vectors (V1, V2) fitting a linear regression ($y \sim x$) each. The two vectors have been calculated using 2δ before and after each observation (z) (V1 = from $[z - 2]$ to z, and V2 = from z to $[z + 2]$) (Fig. S1b and Fig. S1c). Although Chu and Fan (2011) suggested to measure the tangent of the angle between V1 and V2 (φ), we encountered some issues identifying BMLD in those profiles that decreased in density below the BMLD (Fig. A1d, Appendix A). Therefore, the algorithm has been improved by calculating the angle φ. Since the slope (or angular coefficient, β) of a linear regression is the tangent of the angle between the line and the x-axis, the angle φ was obtained from two angles extracted from the coefficients measured by V1 and V2 according to the sign of β: i) positive β (see example in Fig. S1d, angle τ and the orange vector) refers to the angle between the vector and the horizontal plane with y equal to the intercept (α), or ii) negative β (see example in Fig. S1d, angle ω and the blue vector) refers to the angle between the vector and the vertical plane with $x = 0$. The angle φ at each observation (φ_z) is measured by summing up, or subtracting, the angles derived from the coefficients, β_1 and β_2 for V1 and V2, according to their partial contribution to φ, which can be summarized under four different conditions:

$$\varphi_z = \begin{cases} \text{atan}(|\beta_1|) + \text{atan}(|\beta_2|), & \beta_1 > 0 \text{ and } \beta_2 > 0 \\ \text{atan}(|\beta_2|) - \left(\frac{\pi}{2} - \text{atan}(|\beta_1|)\right), & \beta_1 > 0 \text{ and } \beta_2 < 0 \\ \text{atan}(|\beta_1|) + \left(\frac{\pi}{2} - \text{atan}(|\beta_2|)\right), & \beta_1 < 0 \text{ and } \beta_2 > 0 \\ |\text{atan}(|\beta_1|) - \text{atan}(|\beta_2|)|, & \beta_1 < 0 \text{ and } \beta_2 < 0 \end{cases} \quad (1)$$

where atan() refers to the arctangent of the coefficients β_1 and β_2.

Up to this stage, the algorithm selects MLD and BMLD on the adapted maximum angle method (Chu and Fan, 2011). However, the exclusive use of the maximum angle method would have biased the selection due to local variation and instability conditions of the water column (Fig. A1b, c, e, f in Appendix A). Therefore, a K-Mean cluster analysis (Lloyd, 1982) was adopted in the algorithm to improve the selection of the pycnocline limits by adding a further step of selection on the 3 and 5 largest φ for MLD and BMLD, respectively. Since the transition from surface mixing layer to the

pycnocline is sharper than that one from the pycnocline to the below mixed layer, the number of φ candidates is higher for BMLD than MLD. The cluster analysis classifies the density difference at depth ($\Delta\rho_z = |\rho_z - \rho_{z+1}|$) into groups (see below), assuming that $\Delta\rho_z$ values within a mixed layer would belong to a unique cluster.

MLD's selection is made amongst the 3 largest φ, and the first φ_z amongst the descendent ordered candidates meeting the following conditions was selected: i) the observations (*z*) within the mixed water column belong to the same cluster classification (CC), the candidate φ_z must have $CC_z = CC_{z+1}$ and $CC_z \neq CC_{z1}$ (CC at surface z_1), ii) and $\Delta\rho_{z-1} < \Delta\rho_z$. In MLD's selection, the $\Delta\rho_z$ is grouped in two clusters since we would expect two main variations of $\Delta\rho$ in *Split1*: a small gradient on the surface mixed section and a bigger one at the pycnocline due to stratification. The same approach has been adopted for BMLD's identification amongst the 5 largest φ, although the inclusion of three clusters instead of two improved the performance of the algorithm since the region of the water column transiting from the pycnocline to the bottom mixed layer is smoother than in MLDs (e.g. Fig. A1b in Appendix A). The first φ_z amongst the descendent ordered candidates meeting the following conditions was selected as BMLD: i) $CC_z = CC_{z-1}$ and $CC_z \neq CC_{z0.9\Delta\rho}$ (CC at the z=0.9Δρ), and ii) $\Delta\rho_z < \Delta\rho_{z-1}$. Adding the conditions controlling for a similar classification of $\Delta\rho_z$ at depths above MLD and below BMLD resulted in decisive outcomes, correctly identifying the mixed layers within those density profiles having a pycnocline fractured in chunks with different or similar gradients. However, when the conditions associated with clustering were not found among the candidates φ, the algorithm was not necessary and therefore the simplest methods were adopted to select i) MLD with a threshold gradient $\Delta\rho_z > 0.02$ mg m^{-3}, and ii) BMLD as the largest φ (Fig. S1a). The algorithm was developed in R v3.6.3 (R Core Team, 2018), and the K-mean density was calculated using the *kmeans* function using Lloyd (1982) algorithm (*stats* package).
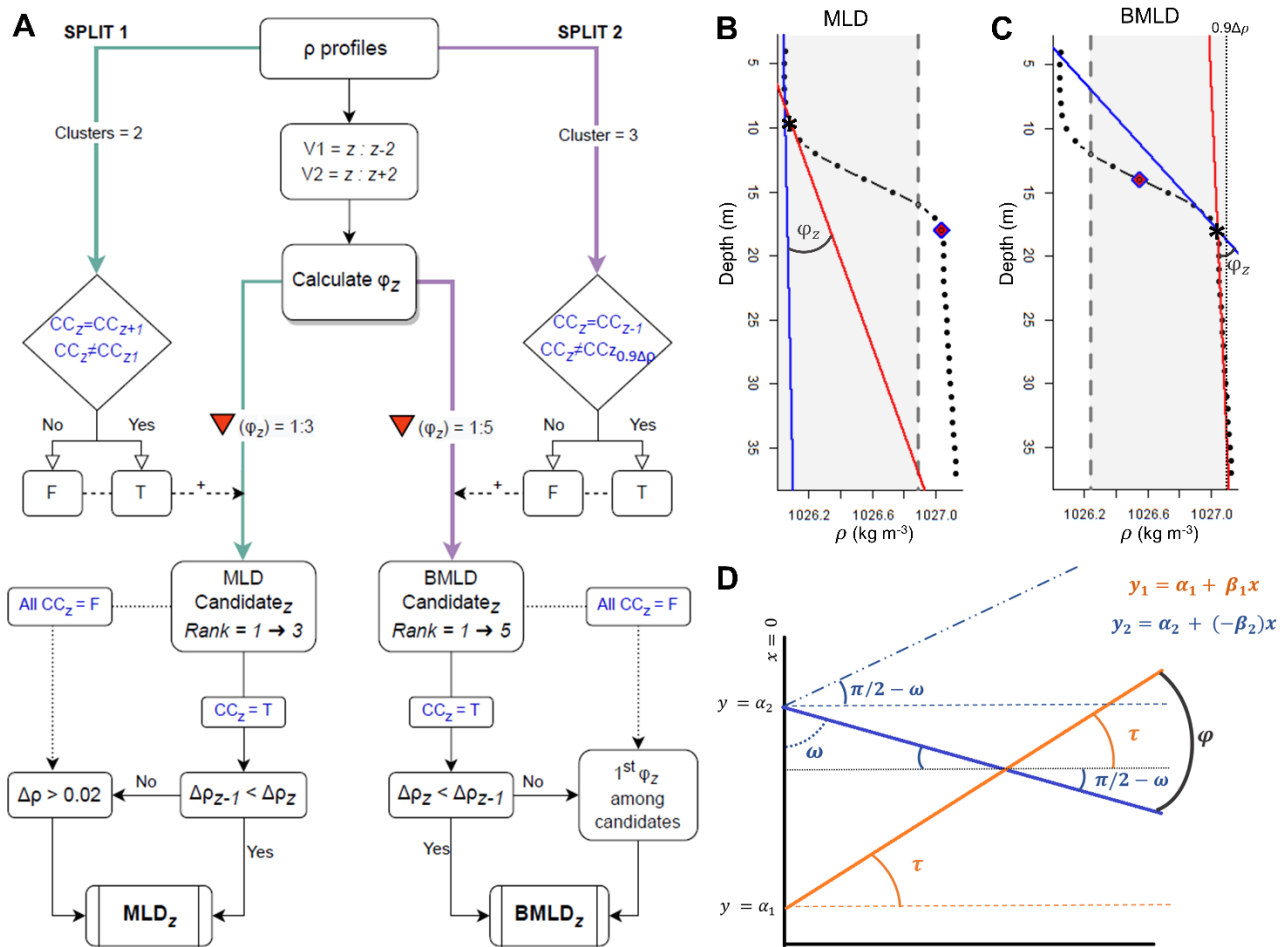
Figure S1: main steps of MLD and BMLD selection: (a) diagram of the algorithms, where green arrows belong to Split1 and purple arrows to Split2, text in blue is the portion of the algorithm relying on cluster analysis (K-mean), "F" and "T" are the results, false and true, of the conditions expressed in the rhombuses. The φ is measured for each observation (z), and the largest (3 for MLD and 5 for BMLD) φ are considered as candidates of MLD and BMLD. The candidates are descendent ordered (Rank 1 → 3 or Rank 1 → 5) and the first candidate meeting the other conditions will be identified as MLD or BMLD. If any candidate meets the conditions, the original methods are used (threshold method > 0.02 and maximum angle φ). (b) and (c) are plots of the same density profile representing the attributes used in the algorithm: grey region includes the observations (black dots) used to identify MLD and BMLD, which extends in (b) from the surface to two depths above BMLD (purple rhombus), and in (c) from two depths above the middle of the pycnocline (purple rhombus) to 0.9Δρ. MLD and BMLD are reported by a black star in (b) and (c) respectively. In (b) and (c), the vectors V1 (blue line) and V2 (red line) are drawn for each z (black star) and φ_z is reported. Plot (d) shows of one of the four conditions reported in Eq. (1) measuring φ: V1 (orange line) with a positive slope (β_1) and V2 (blue line) with a negative slope (β_2).

Bibliography

Chu, P. C. and Fan, C.: Maximum angle method for determining mixed layer depth from seaglider data, J. Oceanogr., 67, 219–230, https://doi.org/10.1007/s10872-011-0019-2, 2011.