**Supplementary information for “A RUpture-Based detection method for the Active mesopeLagIc Zone (RUBALIZ): a crucial step towards rigorous carbon budget assessments” by Fuchs, Baumas et al.**

Figure S1:

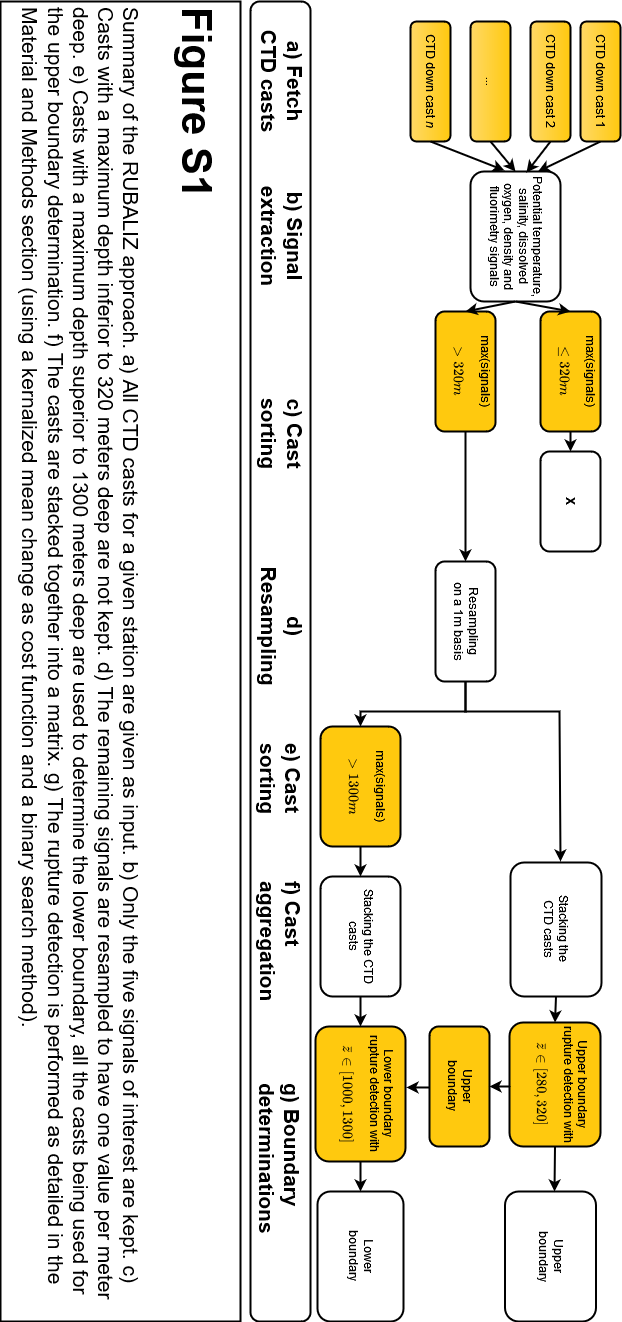
****

Table S1: Depths of the active mesopelagic zone boundaries determined by RUBALIZ

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **cruise** | **station** | **Upper boundary** | **std** | **Upper boundary**  **CTD number** | **Lower boundary** | **std** | **Lower boundary**  **CTD number** |
| D341 | PAP | 135 | 5 | 16 | 726 | 9 | 3 |
| DY032 | PAP | 127 | 31 | 16 | 751 | 26 | 3 |
| KN207-01 | QL-1 | 151 | 0 | 2 | 490 | 19 | 2 |
| KN207-01 | QL-2 | 191 | 0 | 1 | 783 | 198 | 1 |
| KN207-03 | PS-1 | 104 | 0 | 1 | 490 | 5 | 1 |
| KN207-03 | PS-3&4 | 110 | 1 | 1 | 684 | 4 | 1 |
| MALINA | 430 | 79 | 0 | 1 | 544 | 40 | 1 |
| MALINA | 540 | 84 | 1 | 1 | 558 | 75 | 1 |
| MALINA | 620 | 95 | 31 | 1 | 609 | 86 | 1 |
| PEACETIME | FAST | 109 | 1 | 21 | 628 | 28 | 8 |
| PEACETIME | ION | 118 | 1 | 11 | 498 | 25 | 6 |
| PEACETIME | TYR | 111 | 3 | 10 | 606 | 43 | 6 |
| TONGA | STATION 8 | 153 | 2 | 5 | 702 | 135 | 3 |

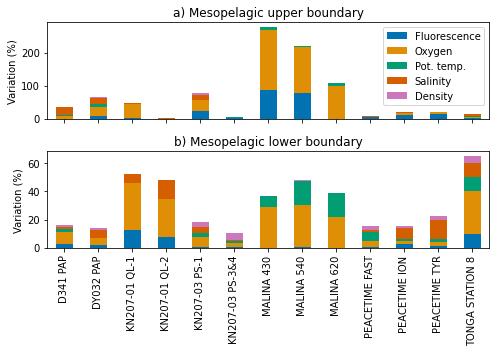


Figure S2: Variation of the boundary estimates due to the withdrawal of one variable from the CTD signal for the upper boundary (a) and lower boundary (b).

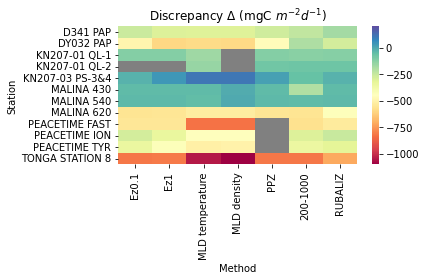


Figure S3: Discrepancies derived from the assessment of C budgets integrated between vertical boundaries defined from all different approaches including RUBALIZ. The discrepancy shown here is the difference between gravitational POC fluxes and prokaryotic C demand derived from H3-leucine measurements (CF leu/C of 1.55ng C pmol-1 Leu and PGE of 7%).

The gray cells correspond to stations for which a given method could not determine an upper boundary.

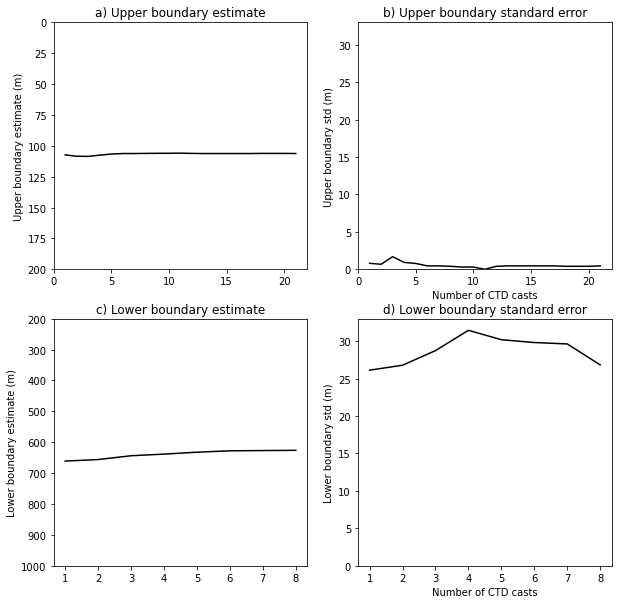


Figure S4: Example of the evolution of the boundaries estimate and associated standard errors (meters deep) when the number of CTDs available grows at the PEACETIME FAST station.

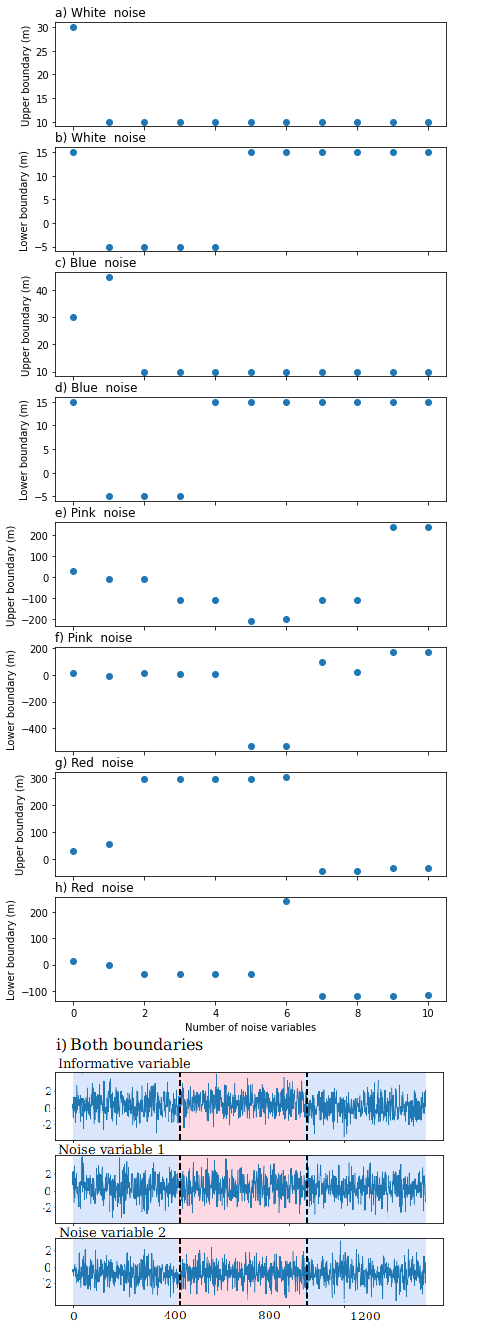


Figure S5: Analysis on synthetic data of the change in (a, c, e, g) upper and (b, d, f, h) lower estimated boundaries when useless variables (noise variables) are added. An “informative” variable for which the breakpoints are known beforehand has been generated along with noise variables of different types: white noise (a and b), blue noise (c and d), pink noise (e and f), red noise: (g and h)[[1]](#footnote-1) (Timmer and Koenig, 1995). White noise presents no autocorrelation whereas the other type of noise does: they are then more likely to perturb the identification process.

Starting with the informative variable, noise variables are sequentially added, and the estimated boundaries (dashed line in i) are compared with the actual ones (colored in red and blue in i). The noise variables and the informative variable have the same standard error (equal to 1). The error is non-zero even without noise variables, underlining the fact that the informative variable presents significant variability (which often occurs in real-world settings). The differences in means between the informative variable sub-signals (colored in i) are respectively 0.6 and -0.3, thus inferior to the standard errors of noise and informative variables. As a result, the informative variable ruptures are comparable in magnitude with the noise variable variance, and identifying the actual breakpoints is not a trivial task. The error in presence of white noise and blue noise remains inferior to 45m whatever the number of noise variables added. The red and pink noise caused more issues, with some strong perturbations occurring when the number of noise variables becomes high. In conclusion, when the noise presents no autocorrelation (a and b), noise variables do not perturb the estimation process. For more complex noises, when there is at least one informative variable for two noise variables, no significant perturbations were recorded. This condition should be easily verified as RUBALIZ is based upon five variables and as no “purely” random variable should be included in the process.

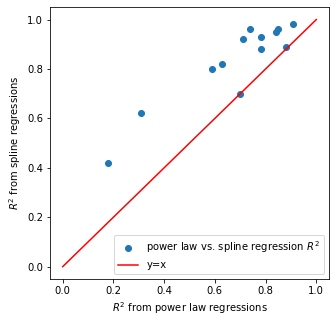


Figure S6: R2 coefficients obtained using linear splines with one node on log-depth and log-PHP as a function of power law regression R2 coefficients for each station. The red line represents a situation where both methods have the same quality of fit, points above the line correspond to stations for which the spline regressions gave a better fit and conversely for points under the line.

Table S2: Median changes in the boundary estimations due to the withdrawal of the fluorescence and [O2] for both upper and lower boundaries. In median the estimated upper boundary moves by 30m (in absolute value) when the two signals are not included and the lower boundary by 101m (in absolute value).

|  |  |
| --- | --- |
| Upper boundary | Lower boundary |
| 30 m | 101 m |

Supplementary references

Timmer, J., & Koenig, M. (1995). On generating power law noise. *Astronomy and Astrophysics*, *300*, 707.

1. Adapted from <http://www.statistics4u.com/fundstat_eng/cc_noise_types.html> (last consulted on 29/08/2022)

   * **white noise**: Independent draws from a Gaussian distribution
   * **pink noise**: The intensity of the noise decreases with the frequency
   * **red noise**: More of low frequency than the average
   * **blue noise**: More of high frequency than the average

   [↑](#footnote-ref-1)