**Additional file 1 - Annex 1** **- MONITOOL sampling sites**

|  |  |  |  |
| --- | --- | --- | --- |
| **Institute** | **Sampling sites** | **Site number (Figure 5)**  **(2 sampling periods)** | **Site with one sampling period** |
| AZTI | Deba | 1 |  |
|  | Lezo | 2 |  |
|  | Museo | 3 |  |
|  | Practicos | 4 |  |
| CEFAS | Belfast | 5 |  |
|  | Fal | 6 |  |
|  | Liverpool |  | X -Technical problem |
|  | X38a | 7 |  |
| DCU | ABW (Alexandra Basin West) | 8 |  |
|  | Dublin Bay Buoy | 9 |  |
|  | Lough Mahon (M69) | 10 |  |
|  | North Channel Great Island (M70) | 11 |  |
| IFREMER | Port-en-Bessin | 15 |  |
|  | Fontenelle | 12 |  |
|  | Saint-Nazaire | 13 |  |
|  | Saumonard | 14 |  |
|  | Terenez |  | X - Additionnal site |
|  | Antifer |  | X - Additionnal site |
|  | Sillon des anglais |  | X - Additionnal site |
|  | Le Croisic |  | X - Additionnal site |
|  | Lazaret |  | X - Additionnal site |
| IPMA | Aveiro |  | X -Technical problem |
|  | Porto |  | X -Technical problem |
|  | Sesimbra |  | X -Technical problem |
|  | Tagus |  | X -Technical problem |
| ITC | Gando | 16 |  |
|  | Jinamar | 17 |  |
|  | Luz | 18 |  |
|  | Taliarte | 19 |  |
| MSS-SEPA | Braehead |  | X - 1 sampling period |
|  | Montrose |  | X - 1 sampling period |
|  | Newhaven |  | X - 1 sampling period |
| UNICA | Molo Dogana | 20 |  |
|  | Molo Ichnusa | 21 |  |
|  | Molo Rinascita | 22 |  |
|  | Sant ‘Elmo | 23 |  |

**Annex 2** - **Statistical process and results**

For each metal, outliers were identified from the linear model, as those samples presenting standardized residuals, greater than 3 (rejection of values above 99.73% of the total values assuming that their distribution follows a normal distribution). In this way, an iterative process was carried out until no more outliers were identified: the process started with the linear modelling of all the results, thus allowing the identification of possible outliers. A new linear model was made by removing the previously identified outliers. The new outliers were then removed, and the process was repeated until all the outliers were identified.

The diagnosis of the model was based on a standard graphical panel with graphical and numerical outputs to judge the normality and homoscedasticity of the residuals (Breusch and Pagan, 1979 ; Goldfeld and Quandt , 1965) (Annex 3). Although there was a consensus on these parameters for the validation of linear models, the assumption of normality of the residuals was considered secondary; the linear model was then considered to be robust to the absence of normality of the given residuals. Nevertheless, in addition to the QQ-plot (which allows a graphical analysis of the normality of the residuals) the Shapiro-Wilk and Lilliefors (Kolmogorov-Smirnov) tests were performed to guide the validation of the model (Knief and Forstmeier, 2020). The most important point was to have a QQ-plot curve that was as linear as possible with no real residuals that stand out (no curved distributed residuals). In this case, even if the results of the normality tests mentioned above were below the validity threshold (p-value < 0.05), the normality hypothesis was accepted . The Shapiro-Wilk and Lilliefors (Kolmogorov-Smirnov) tests were selected considering that the power and robustness of the tests is subject to application conditions, namely the size of the sample population. During data processing, the first step aimed at identifying and excluding outliers (pairs of values), which resulted in fluctuating sizes of metal datasets for linear regressions. To take these varying sizes of datasets into account, it was deemed necessary to consider two normality tests, Shapiro-Wilk and Lilliefors (Kolmogorov-Smirnov), allowing a less critical diagnosis than a result based on the selection of a single test. The idea was to choose a parametric test (Shapiro-Wilk) and a non-parametric test (Lilliefors) (Scherrer, 2007), which have different conditions of application, particularly with regard to the size of the datasets. This approach has also been used to test the homoscedasticity of the model's residuals.

The homoscedasticity of the residuals was also considered, thanks to the graphical output illustrating the square roots of the residuals (standardized residuals) as a function of the predicted theoretical values (fitted-values) of Y and the Breush-Pagan and Golfeld-Quandt tests. Although Koenker's version was considered (less sensitive to the lack of normality than the original version), the Breush-Pagan test stayed sensitive to the lack of normality. The homoscedasticity of the residuals was also tested with the Golfeld-Quandt test.

The validity of the final model for each metal was therefore considered valid when:

1- p-value < 0.05. A low p-Value (<0.05) indicates that the null hypothesis (no effect) can be rejected. In other words, a predictor with a low p-value is likely to be a significant addition to the model, because changes in the value of the predictor will result in changes in the response variable;

2- the residuals follow a normal or near-normal distribution. Although the normality of residuals was checked using the Shapiro-Wilk and Lilliefors tests, a subtle deviation from normality was accepted, following the paradigm of Wasserstein et al. (2019), when the QQ-plot was linear without standing curve-distributed residuals, as linear models are known to be very robust to violations of the normality assumption (Knief and Forstmeier, 2020).

3- the homoscedasticity of the residuals is verified; i.e., there is no increase in the variance of the residuals when the value of X increases.

It should nevertheless be specified that the choice to consider two tests for each criterion offered the possibility to be more objective on the decision (each test having its limits/gaps). Thus, the validation of a model was accepted if one of the 2 tests for each criterion was valid. As mentioned above, normality could be the subject of a less clear-cut decision (linear model is very robust to violations of the normality assumption (Knief and Forstmeier, 2020), also based on the graphical output (QQ-plot). In any case, it is advisable to remain vigilant on this cut-off point of statistical thresholds (i.e. p-value < 0.05), keeping in mind the number of values taken into account, their general appearance by a visual glance (graphical output) and by relativizing the concept of "significant" (Wasserstein et al., 2019).

**REFERENCES**

Breusch, T.S., Pagan, A.R., 1979. A simple test for heteroscedasticity and random coefficient variation. Econometrica: Journal of the Econometric Society 1287–1294.

Goldfeld, S.M., Quandt, R.E., 1965. Some tests for homoscedasticity. Journal of the American statistical Association 60, 539–547.

Knief, U., Forstmeier, W., 2020. Violating the normality assumption may be the lesser of two evils. bioRxiv 498931. https://doi.org/10.1101/498931

Scherrer, B., 2007. Biostatistique, vol. 1, Gaëtan Morin éditeur (816 pp.). ed.

Wasserstein, R.L., Schirm, A.L., Lazar, N.A., 2019. Moving to a World Beyond “*p* < 0.05.” The American Statistician 73, 1–19. https://doi.org/10.1080/00031305.2019.1583913

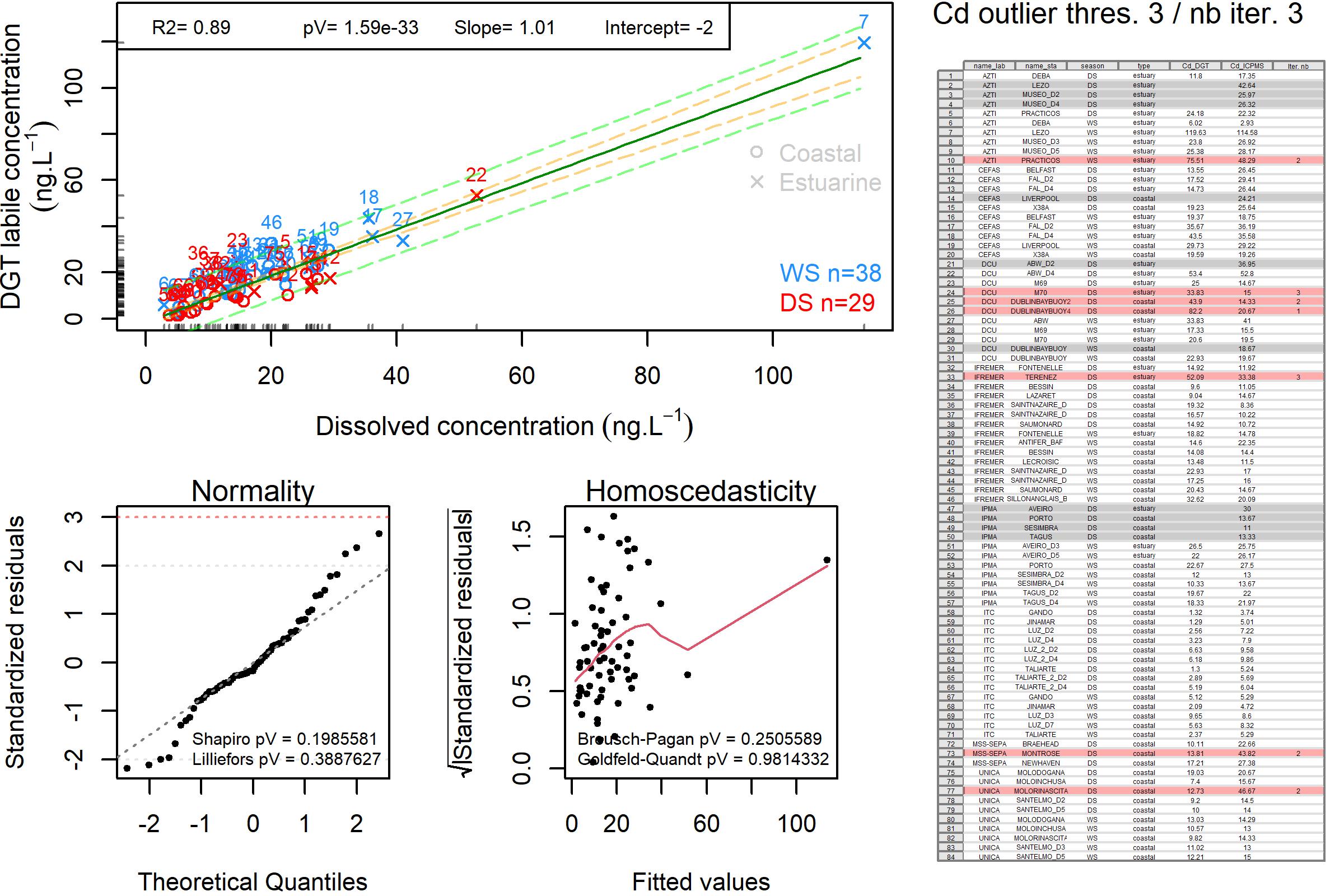
**Annex 3** - **Statistical process results**

Figure S1: Cadmium: Relationship between dissolved concentration and DGT labile concentration. The dark green line represents the linear model and the corresponding confidence interval (CI 95; orange dashed lines) and prediction interval (PI 95 ; green dashed lines) are also depicted. WS: wet season; DS: dry season. The table represents, for each site/season, the coupled DGT and dissolved concentrations (ICPMS) results. The rows with a grey background cannot be used because a result is missing in one or the other method. The rows with red background correspond to the results identified as outliers by the iterative process and the iteration number (nb. Iter.) is specified. The rows with a white background are those that have been used to build the model

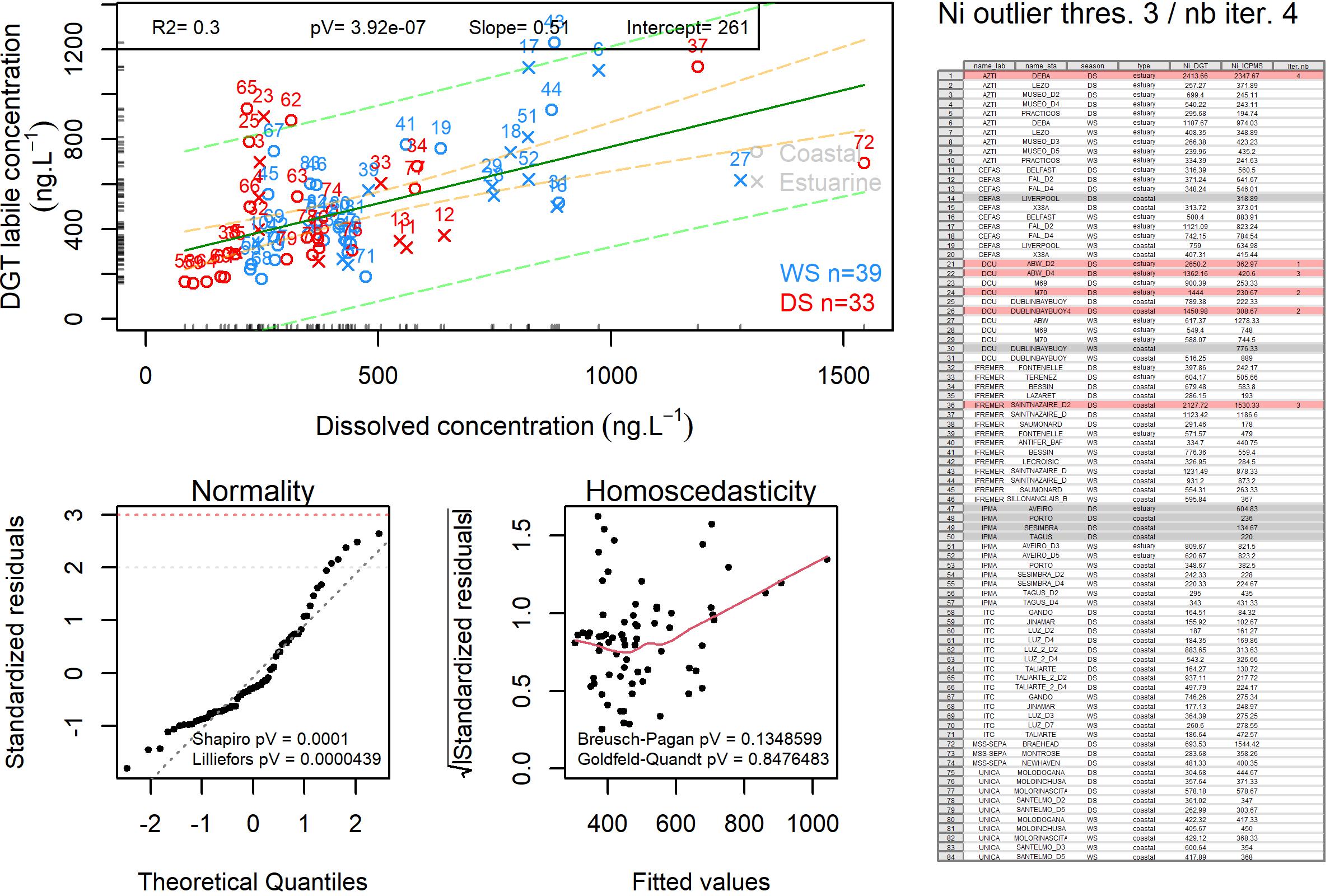


Figure S2: Nickel: Relationship between dissolved concentration and DGT labile concentration. The dark green line represents the linear model and the corresponding confidence interval (CI 95; orange dashed lines) and prediction interval (PI 95 ; green dashed lines) are also depicted. WS: wet season; DS: dry season. . The table represents, for each site/season, the coupled DGT and dissolved concentrations (ICPMS) results. The rows with a grey background cannot be used because a result is missing in one or the other method. The rows with red background correspond to the results identified as outliers by the iterative process and the iteration number (nb. Iter.) is specified. The rows with a white background are those that have been used to build the model

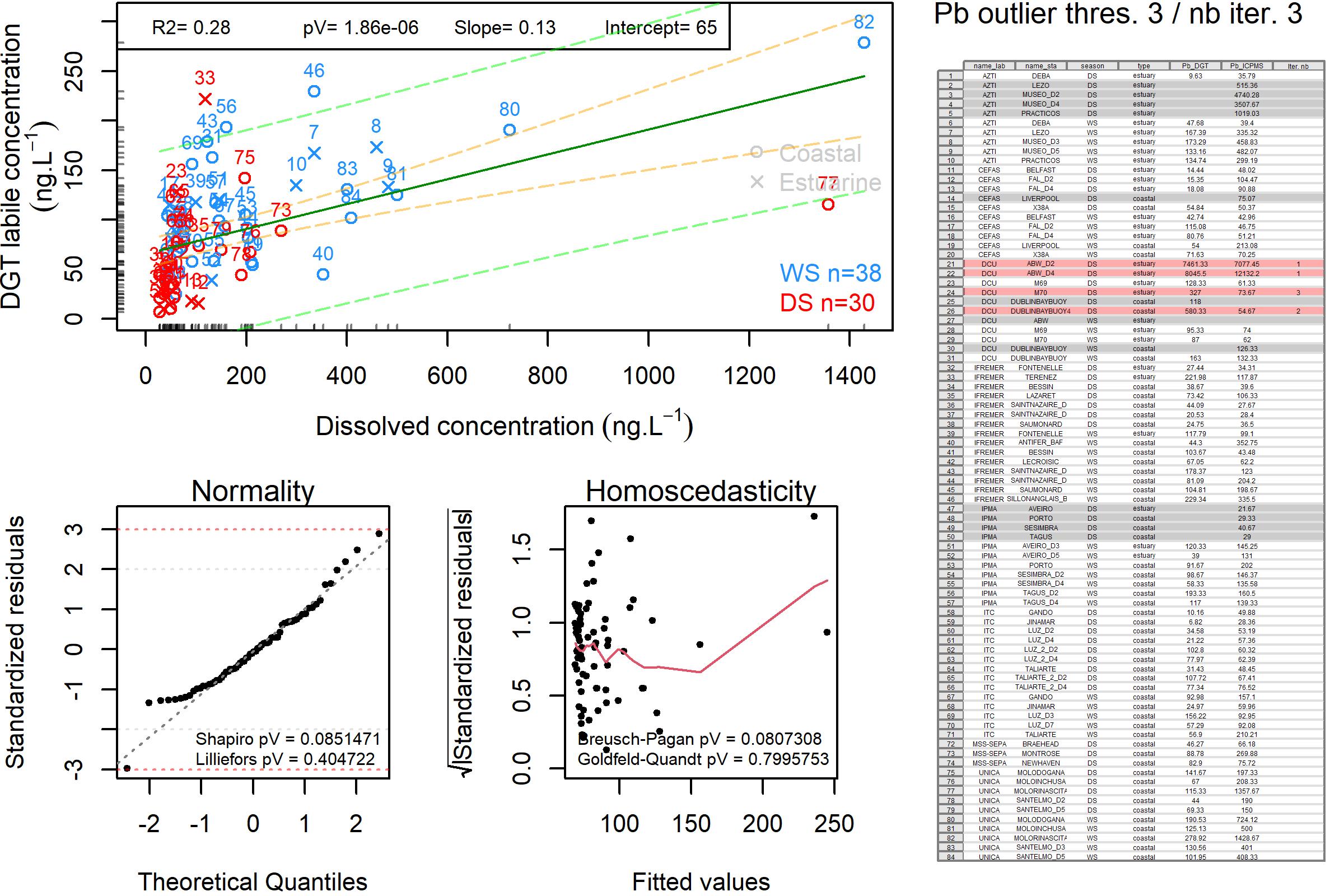


Figure S3: Lead: Relationship between dissolved concentration and DGT labile concentration. The dark green line represents the linear model and the corresponding confidence interval (CI 95 ; orange dashed lines) and prediction interval (PI 95 ; green dashed lines) are also depicted. WS: wet season; DS: dry season. The table represents, for each site/season, the coupled DGT and dissolved concentrations (ICPMS) results. The rows with a grey background cannot be used because a result is missing in one or the other method. The rows with red background correspond to the results identified as outliers by the iterative process and the iteration number (nb. Iter.) is specified. The rows with a white background are those that have been used to build the model

**Annex 4** – **Results of the three statistical methods used as a first approach in the data processing to determine the (A)-EQS DGT**

Preliminary data processing was carried out to explore the dataset and define the most appropriate method and selection of data to be used.

The quantile regression was initially used to visualise the effect of the distribution of the results on the linear regression, by illustrating the outputs by different quantiles of the Xs. It demonstrated the positive effect of removing statistical outliers; better distribution of the lines of the different quantiles (less intersecting or more parallel lines), and especially the slopes and intercepts were very close between the regression line and the median of the quantile regression.

For exploration purpose, comparison of the (A)--EQS DGT values determined using three different statistical approaches was done: i) by multiplying the EQS marine water by the CFDGT (geometric mean of the ratio ([M]DGT/([M]Dissolved concentration) ii) by quantile regression, and iii) by linear regression. The results of these three methods are provided in Table 1.

The quantile regression is not a method intended to make predictions contrarily to linear regression, and using a linear regression is preferred than using only the CF DGT. Thus, the linear regression model was selected as the most appropriate method. (A)- EQS DGT values were closed for Cd using the three different methods, and the linear model provided the lowest (A)-EQS DGT for Ni and Pb, ensuring a more stringent classification and a higher level of protection.

Table S1: Results of the three statistical methods used as a first approach in the data process to determine the (A)-EQS DGT.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Concentration Factor method | | Linear regression model method | Quantile regression model method |
|  | CF DGT  (geometric mean) | (A)-EQS DGT (µg.L-1) | (A)-EQS DGT (µg.L-1) | (A)-EQS DGT (µg.L-1) |
| Cd | 0.91 | 0.18 | 0.20 | 0.21\*  → 0.20 |
| Ni | 1.27 | 10.92\*  → 8.6 | 4.60 | 4.89 |
| Pb | 0.79 | 1.02 | 0.23 | 0.26 |

\* Labile metal concentrations represent free ions and metals dissociating sufficiently fast from inorganic and organic complexes, while not accounting for metals present in strong complexes that, depending metal-specific characteristics and the environment, can represent a high percentage of the metal present. Therefore, as a precaution, when calculated (A)-EQS DGT (10.92 µg.L-1 and 0.21 µg.L-1 for Ni and Cd, respectively) were higher than the existing EQS marine water, these last ones were considered for the (A)-EQS DGT (8.6 µg.L-1 and 0.20 µg.L-1 for Ni and Cd, respectively).