CORRECTIONS ON THE CD139 DATA BASE
Marta Álvarez.

Instituto de Investigaciones Marinas, C/ Eduardo Cabello, Nº6, Vigo, Spain.
malvarez@iim.csic.es

There are two kinds of checking or quality control when detecting and correcting dubious biogeochemical data (BioGeo Var) in a cruise data base: the intra station and the inter station variability. The intra station variability is somehow easier to detect, any sample breaking the trend of any biogeochemical variable against pressure, temperature, salinity, etc… would be considered suspicious and more thoroughly inspected. However, the interstation variability comprises the natural environmental variability, and any errors due to sampling and/or analytical procedures, therefore detecting these "errors" will be the result of a comprehensive and tricky work as we want to preserve the natural variability in the samples.

Correction procedure followed.

The first approximation to a data base searches for any bottle improperly fired and discards the data. Contrasting CTD salinity against salinometer salinity is the best way to trace these misleading bottles.

The second step consists in checking the internal consistency of the data within each station. This is done taking advantage of the natural relationships between physical variables (considered as reference ones) and biogeochemical ones. Eventually, these relationships reflect the water mass properties. Additionally, nutrients, oxygen and carbon data apart of containing information about the water mass inner characteristics, are closely related as they are affected by biological activity, i.e., respiration and photosynthesis. Photosynthesis increases oxygen, reduces nutrients and Total Inorganic Carbon but increases pH, on the contrary, respiration reduces oxygen and pH, but increases nutrients and TIC. On the other side, alkalinity and silicate present a different dynamics as they are mostly affected by the hard pump, the dissolution and formation of hard tissue material.

Taking this in mind, biogeochemical data from each station is carefully examined. A break in the trend of any BioGeo Var against a physical variable should be detected in the other BioGeoVar. Changes should be coherent.

Usually we first calculate the conservative parameters NO ($=O_2 + R_N \cdot NO_3$) and PO ($=O_2 + R_P \cdot PO_4$), using Redfield ratios. The distribution of these parameters against

potential temperature (Tpot) is studied. Then PO against NO, any dubious sample must be examined either in the oxygen or in the nutrient distribution. Then, pH is checked against Tpot and NO3 and finally silicate (SiO4) with Tpot, and then Normalised Alkalinity (NTA = TA·35/Sal) with SiO4.

Suspicious data are flagged and interpolated.

Once each station is coherent itself we proceed to check the interstation variability. As previously said any changes in a BioGeoVar should be accordingly reflected in the others. In order to avoid areas with a high biogeochemical variability and detect the "errors" we preferentially study deep waters where biological activity is supposed to be low, and the variability in the BioGeoVars should be mostly related to the inner characteristics of the water masses.

The influence of several water masses is detected in the data distribution along the CD 139 cruise track. This fact is clearly detected in the complex relationship between Tpot and NO, for example.
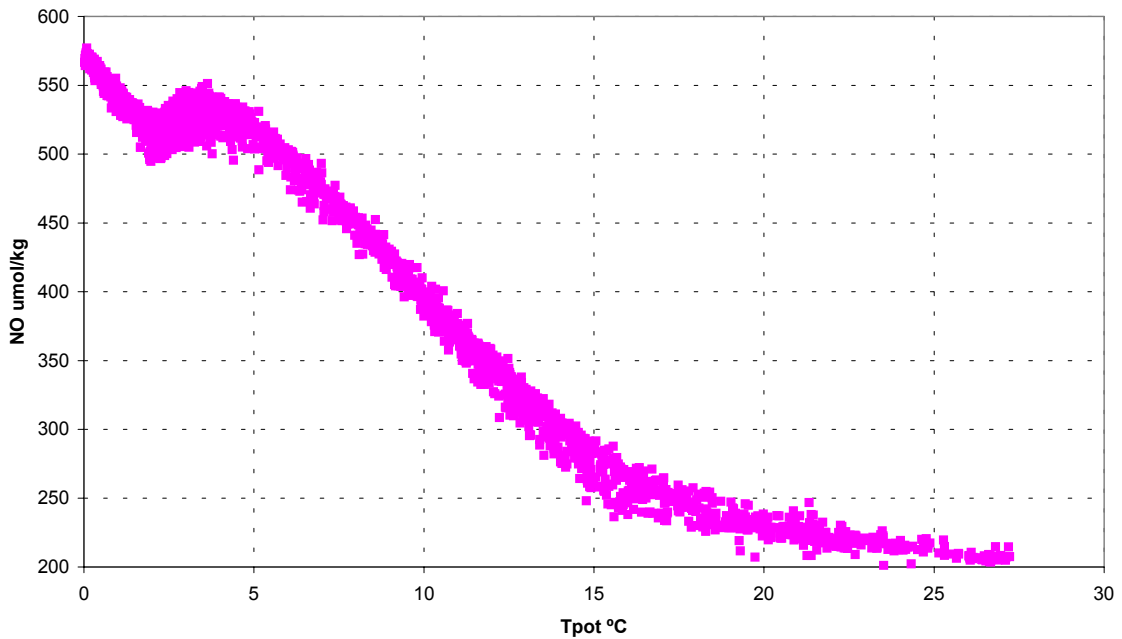
Figure 1: Final Data: NO against potential temperature for the whole CD 139 data set.

With the aim of detecting offsets in the BioGeoVars we calculated a multilinear regression between every BioGeo Var and the following physical variables: pressure, Tpot and salinity for waters below 5ºC.

Then, we calculated the residual of these regressions: ResBioGeoVar = BioGeoVar calculated – BioGeoVar measured.

The residuals should contain information about the biological aging of the water plus any additional offset in the data due to analytical standardisations.

We calculated the residuals for the following BioGeo Var: NO3, PO4, SiO4, pH, NTA, SiO4 and O2.

O2 and pH residuals should be positively correlated, while they should be negatively correlated to NO3 and PO4 residuals. NO3 and PO4 residuals should be positively correlated and their magnitude should practically equal the Redfield ratio (about 16).

Residuals of SiO4 and NTA should be positively correlated.

Additionally we calculated the mean salinity (34.66) and Tpot (2.2) of the samples below 5ºC and then the mean of the following: (Tpot – 2.2) and (Sal – 34.66) quantities in each station. Both give us information about the thermohaline variability between stations.

All these information put together will allow us to detect spurious variations and correct the biogeochemical data.

a) O2 and pH residuals. Figure 2.

The following figure shows the variation of the O2 and pH residuals along the cruise track. Res pH is multiplied by 400 to be compared with the O2 Res. It also shows the variation of 10·(Tpot-2.2) along the section.

Although a bit difficult to see, changes in RespH are correlated with ResO2 and Tpot-2.2. These distributions helped to detect offsets in the pH values of several stations (70, 71, 98, 101, 111, 133, 134) which are corrected.

b) NO3 and PO4 residuals. Figure 3.

This figure shows the distribution of the NO3 raw and the PO4 residuals values multiplied by 16 (mean Redfield ratio) so to get quantities of the same order of magnitude. This figure tells you that there are some suspicious uncorrelated changes in the NO3 and PO4 residuals, and that sharp changes in their magnitude are not reflected in the pH and O2 plot.
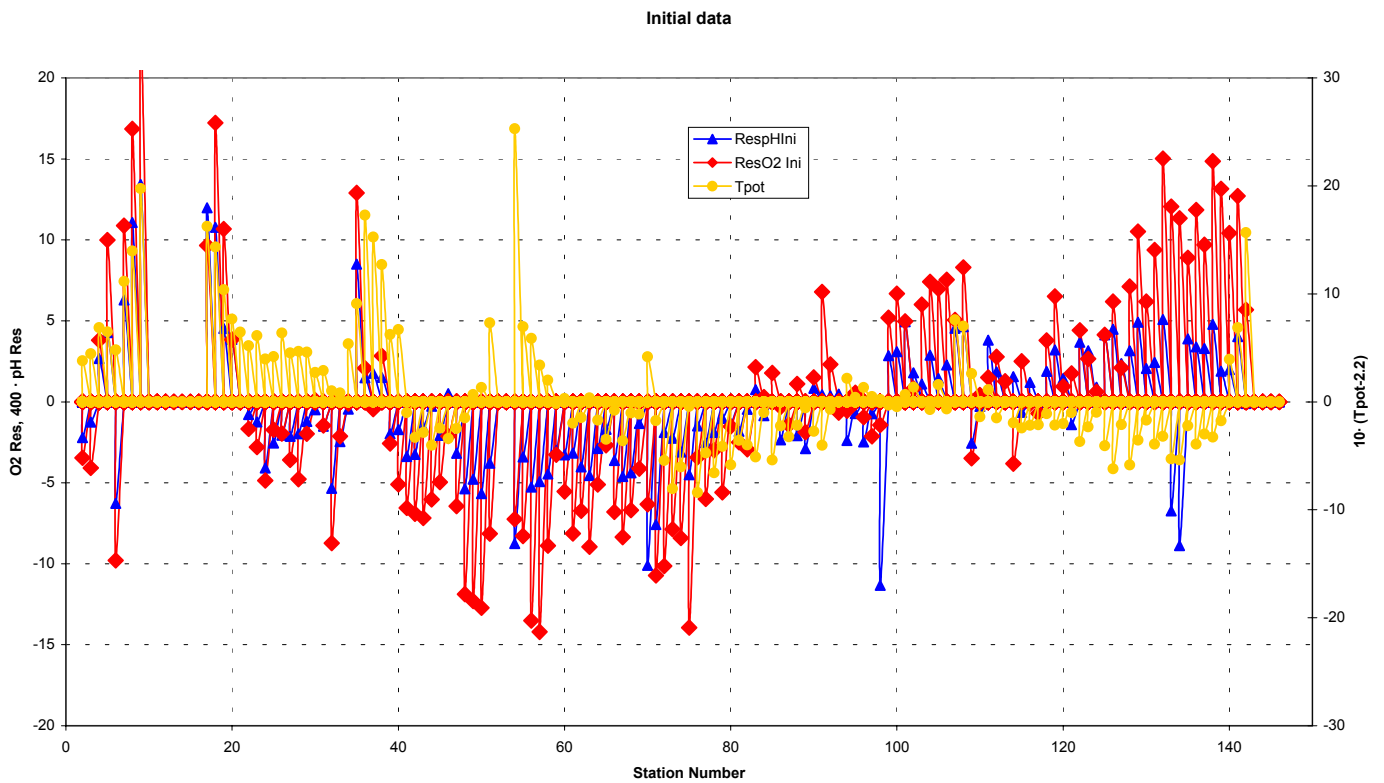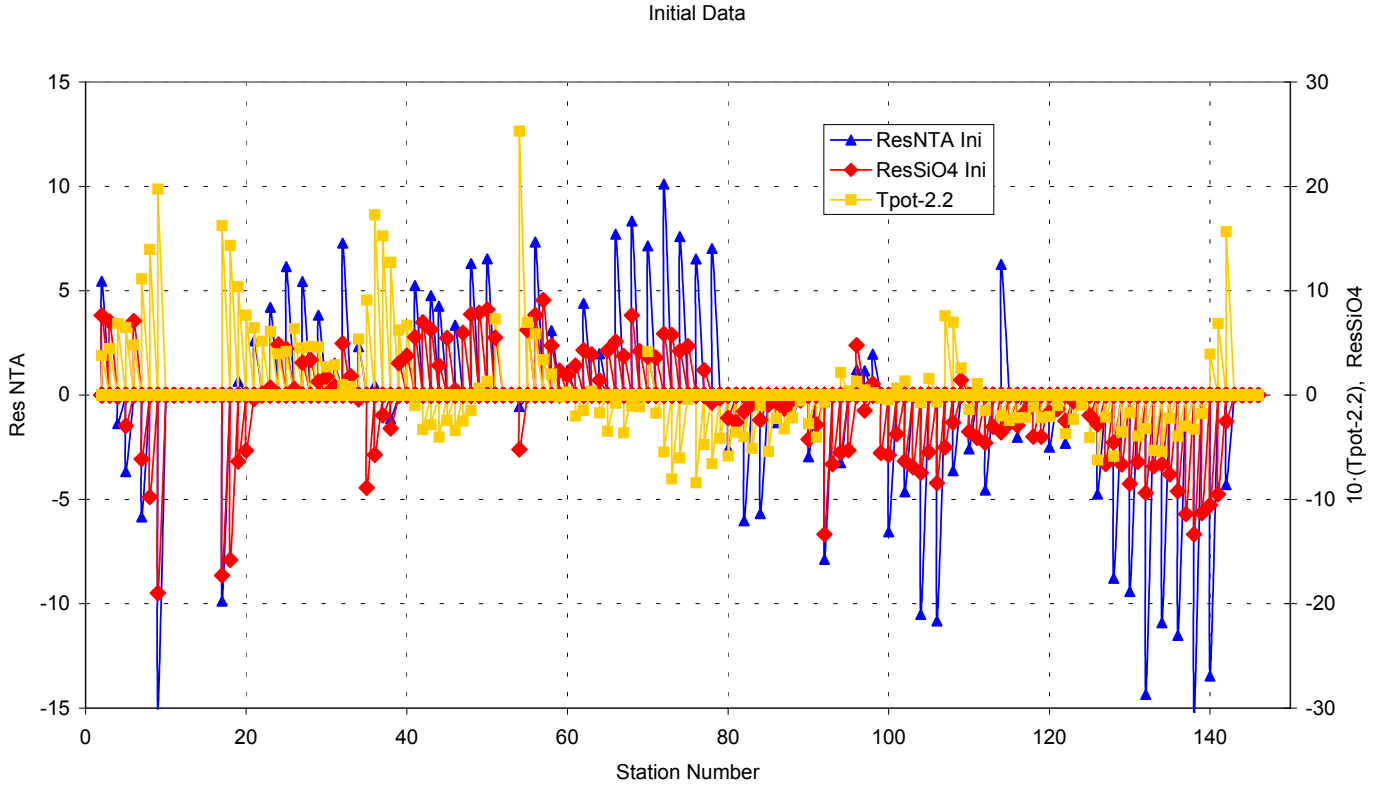
Figure 2. Initial Data: along station variability of the pH and O2 residuals (Calculated-Measured) and mean value of (Tpot-2.2) for waters <5ºC.
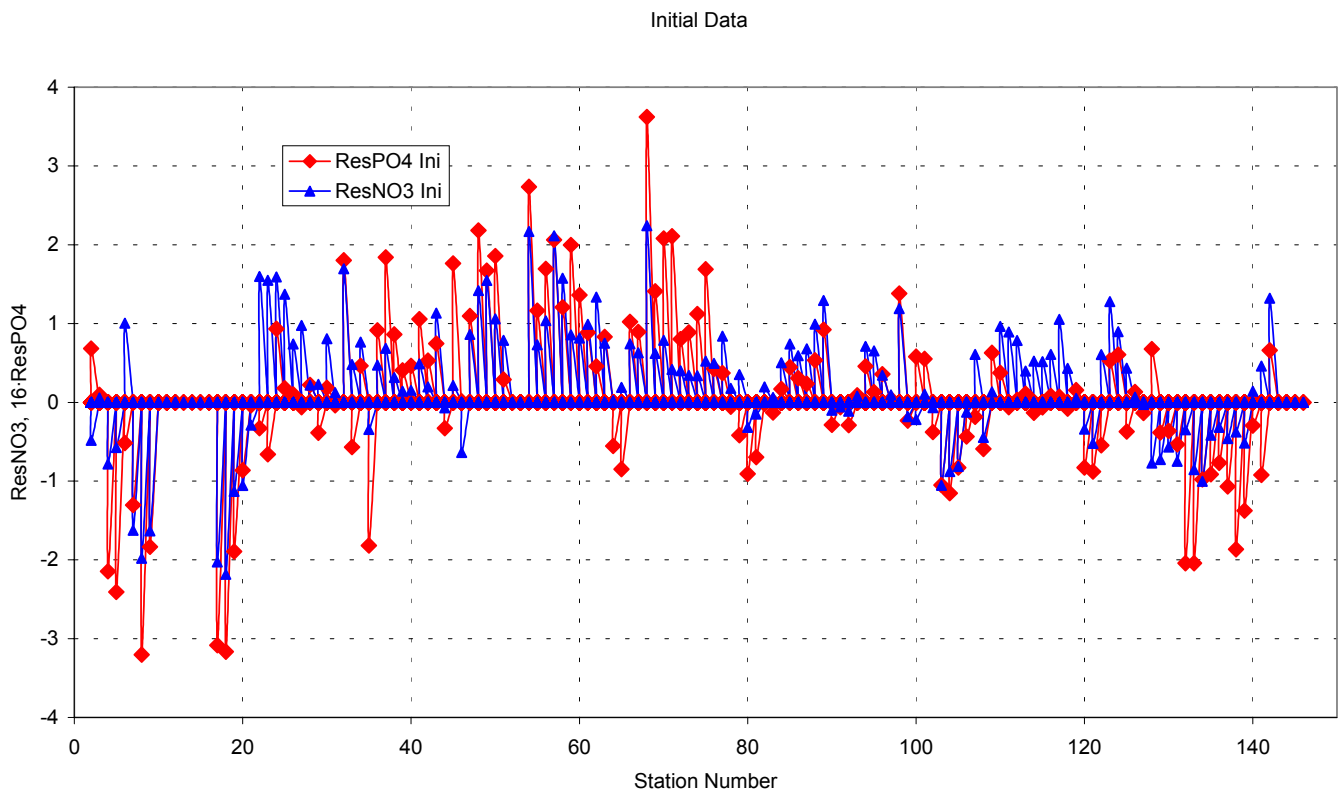
Figure 3. Initial Data: along station variability of the NO3 and PO4 residuals multiplied by 16 (Calculated-Measured) (μmol/kg) for waters <5ºC.

c) SiO4 and NTA residuals. Figure 4.

This figure shows the distribution of the SiO4 and NTA residuals along the CD 139 section. In a first approximation these residuals are mainly well correlated in sign and size and follow the trend in the temperature changes. However a closer look is still needed.

Figure 4. Initial Data: along station variability of the SiO4 and NTA residuals (Calculated-Measured) (both in μmol/kg) and mean value of (Tpot-2.2) for waters <5ºC.

d) Correlations between residuals. Figures 5 and 6.

Figure 5 shows the correlation between O2 residuals and NO3, PO4 (multiplied by 16) and pH (multiplied by 400) residuals. This figure points to the high correlation among the residuals indicating that biological activity or aging is the main reason for this correlation. However, some points (stations) need further investigation.
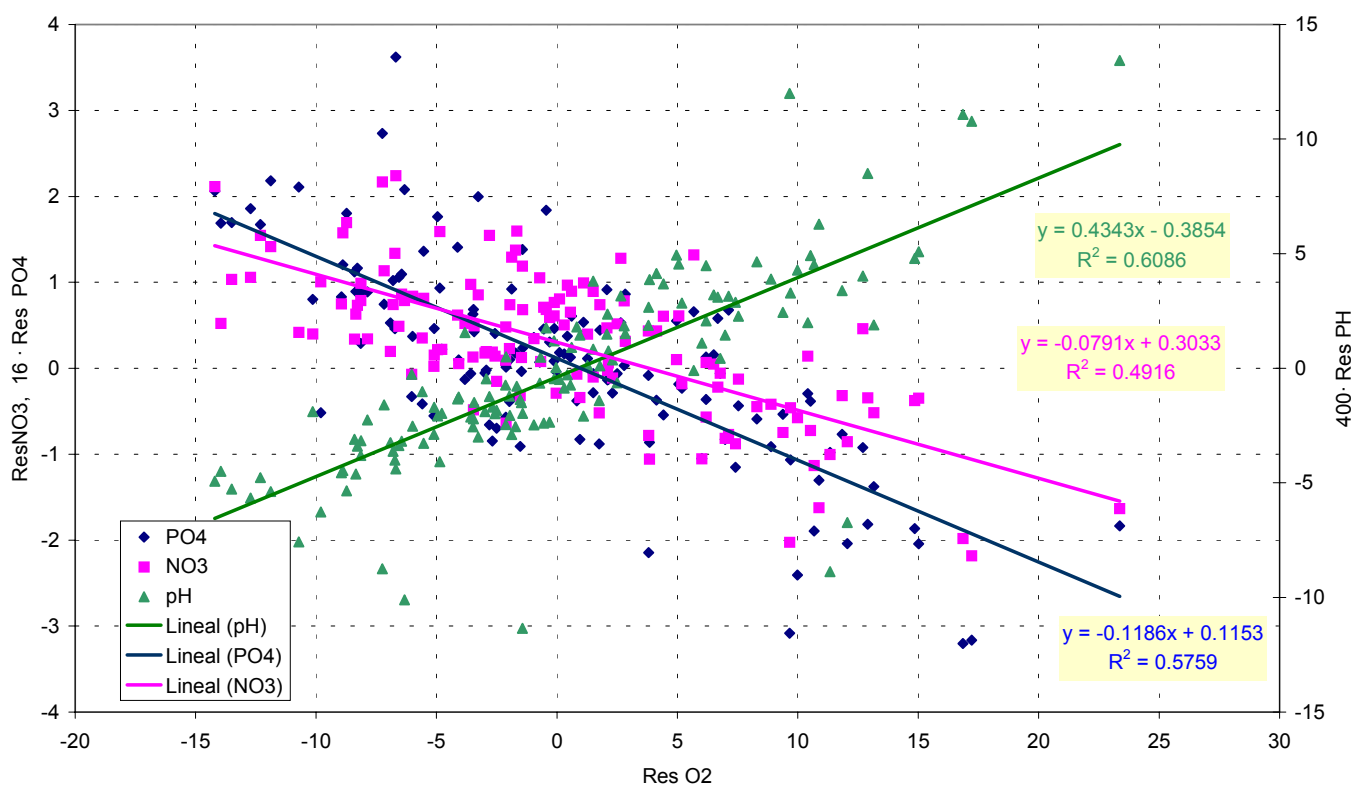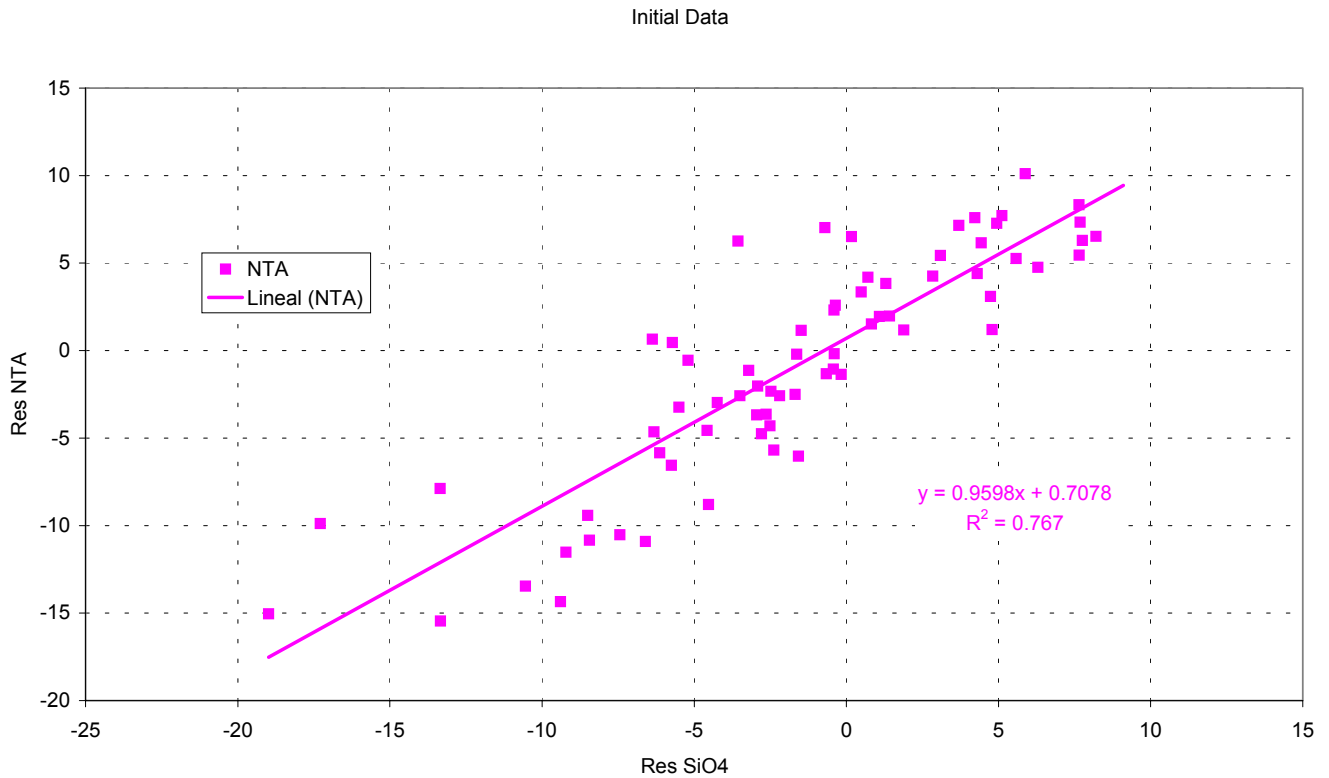
Figure 5. Initial Data: correlation between O2 residuals and NO3, PO4 (multiplied by 16) and pH residuals (multiplied by 400). NO3 and PO4 residuals are in µmol/kg.
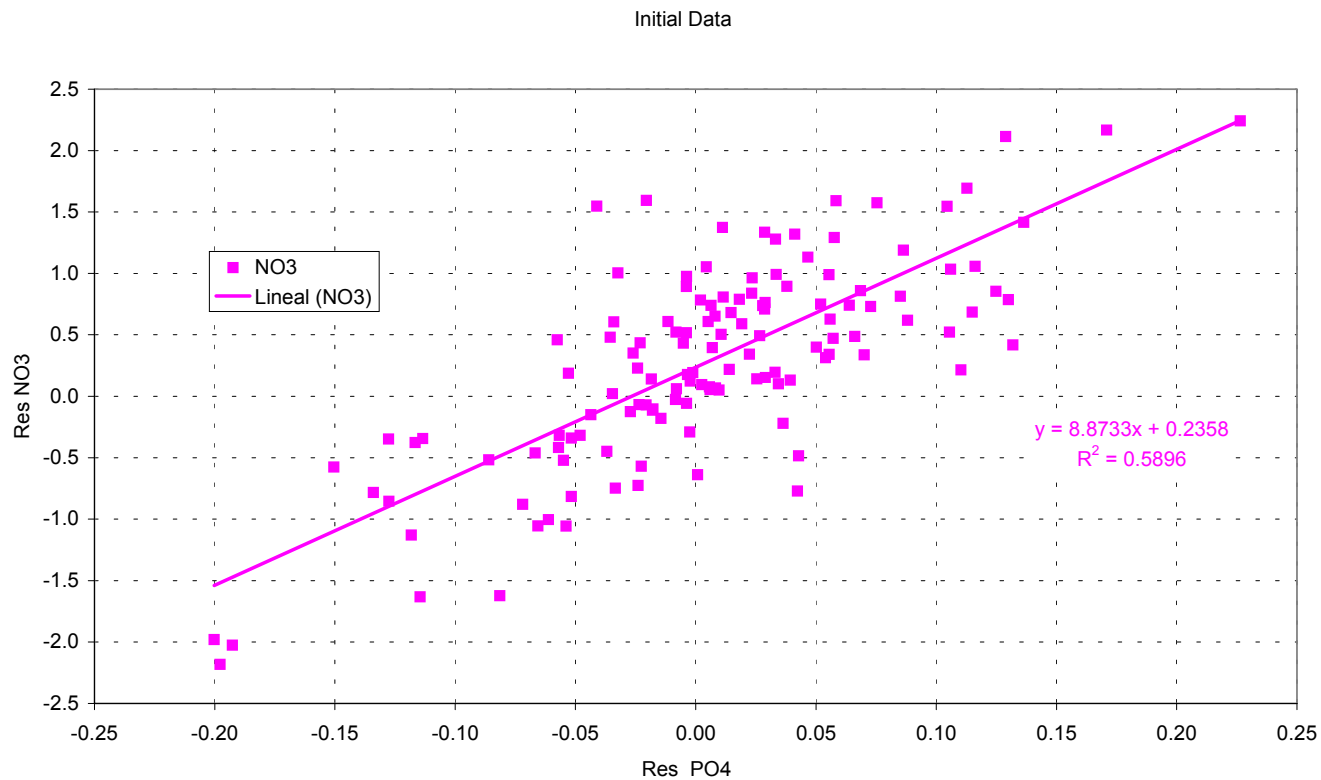
Figure 6. Initial Data: correlation between PO4 and NO3 residuals (µmol/kg).

Figure 6 indicates that the relationship between NO3 and PO4 residuals is low compared to the Redfield value (about 16) and therefore that some stations need more investigation.

Figure 7. Initial Data: correlation between SiO4 and NTA residuals (μmol/kg).

Figure 7 shows the good correlation between SiO4 and NTA residuals and helps to identify some dubious points.

The next step in the correction procedure was to check the trends of the residuals means along the CD 139 track. A group of about 10 stations was selected and all the residuals and Tpot and Sal means plotted. O2 and pH Res are firstly inspected, if they do not have a coherent distribution a closer look to the corresponding station is done. Usually was pH the variable corrected, pH is corrected with an offset added or subtracted to the whole water column data for the corresponding station.

Next, NO3 and PO4 residuals are inspected. In this case if any anomaly detected a factor is calculated by eye to force the mean value of the residuals to follow the O2 and pH trends. PO4 residuals are allowed to be higher than the NO3 ones. I mean, I do not force the PO4 Res to equal the NO3 as in most of the stations the PO4 residuals

(multiplied by 16) are higher than the NO3. The water column data is multiplied by this factor.

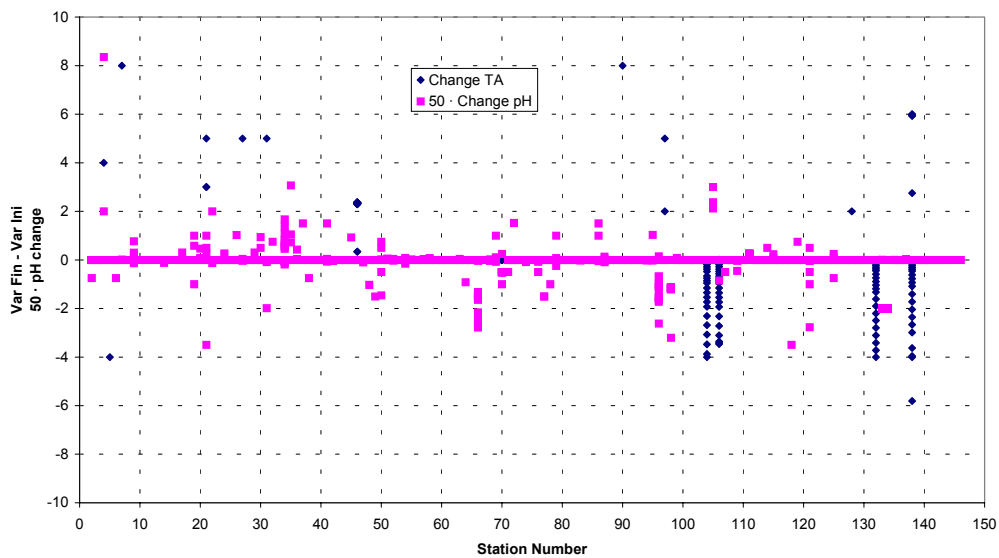Any correction applied over the data is then checked. The new data is plotted, Tpot vs. pH, NO3, PO4, SiO4, NTA; and Tpot vs. NO, PO; pH vs. NO3, PO4; NO3 vs. PO4; NO vs. PO; SiO4 vs. NTA.
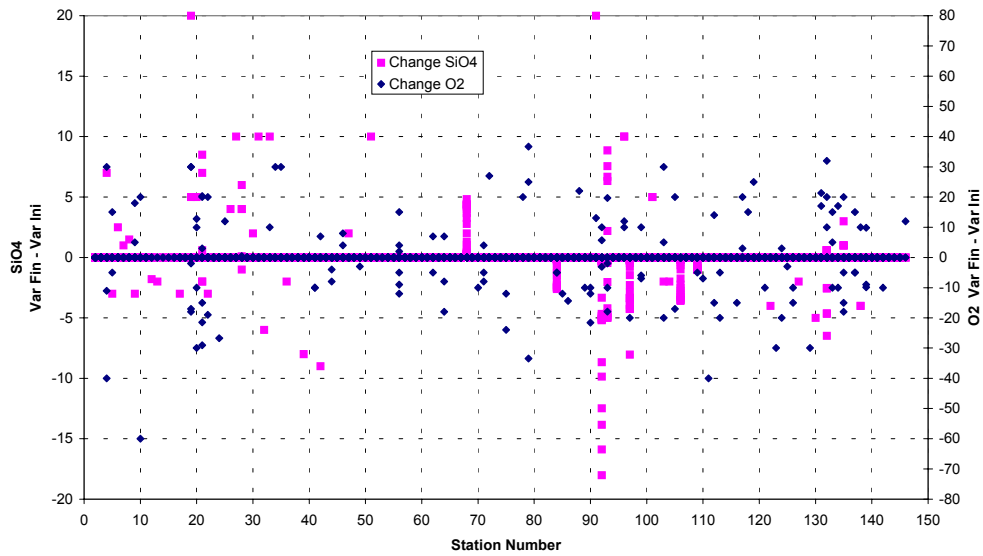
Corrections applied.

The following plots show the absolute corrections applied to the data.
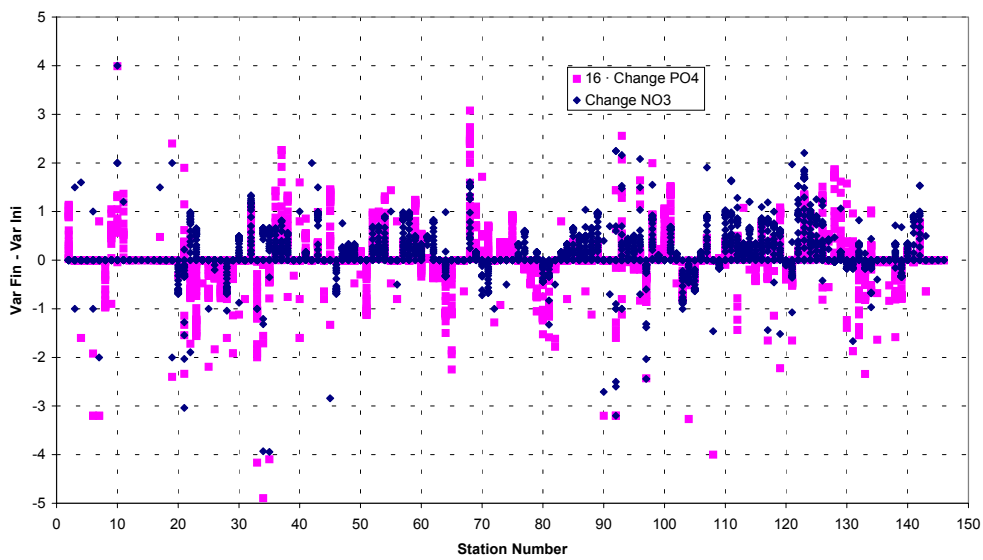
Figure 8. Corrections added to the NO3 and PO4 initial data base in μmol/kg.

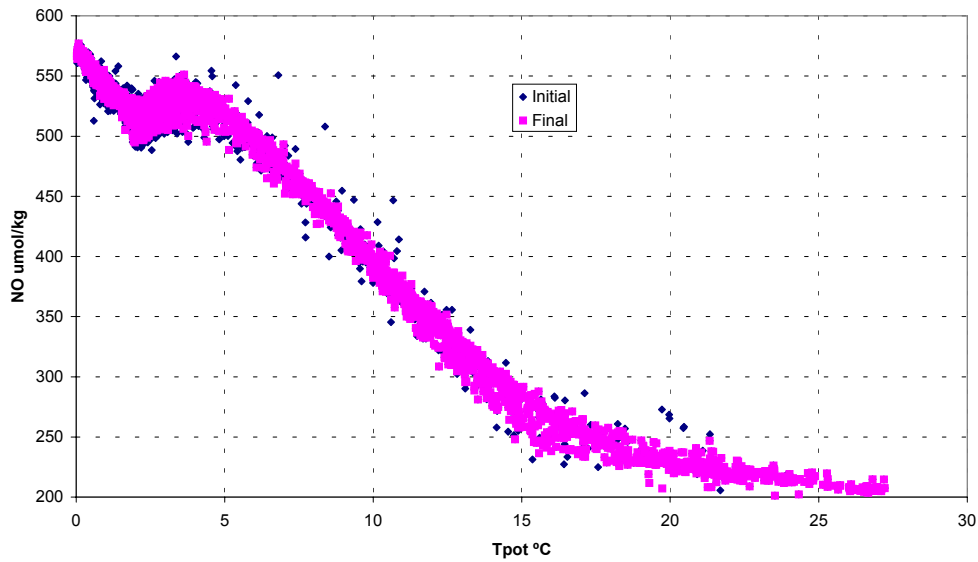
Figure 9. Corrections added to the pH and TA initial data base, pH changes are multiplied by 50, TA is in μmol/kg.
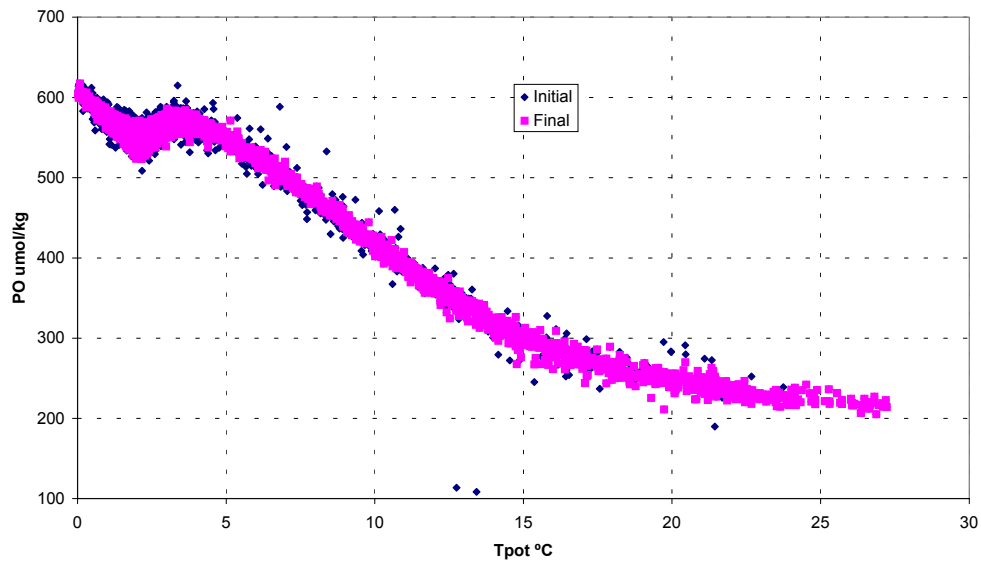
**Figure 10.** Corrections added to the O2 and SiO4 initial data base, both in μmol/kg.

New data set.

The following figures compare the distribution of the BioGeo Var in the initial and final data base.

**Figure 11.** NO (μmol/kg) distribution against potential temperature for the initial and final data base.

Figure 12. PO (μmol/kg) distribution against potential temperature for the initial and final data base.
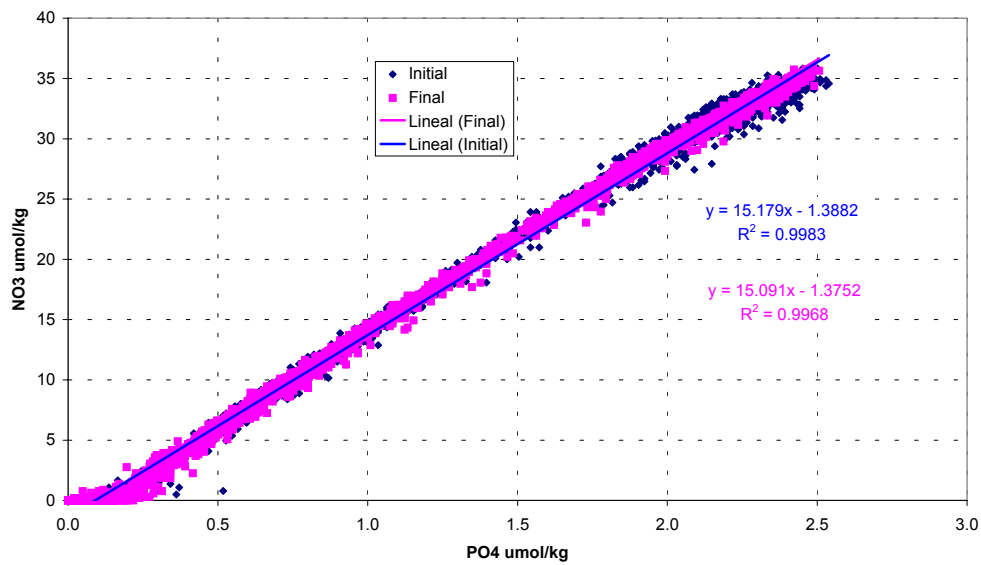
Figure 13. Relationship between NO3 and PO4 (μmol/kg) for the initial and final data base.
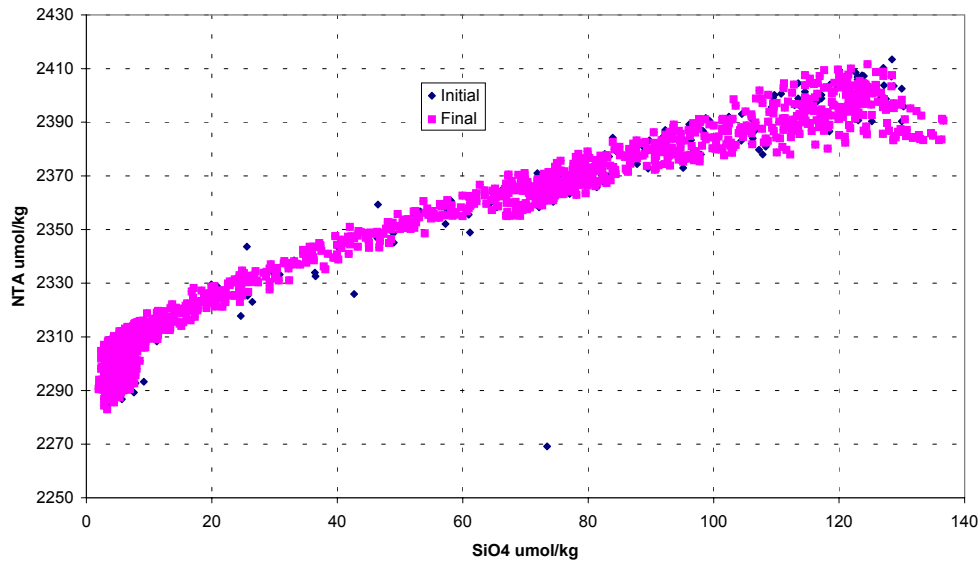
Figure 14. Relationship between SiO4 and NTA (both in µmol/kg) for the initial and final data base.
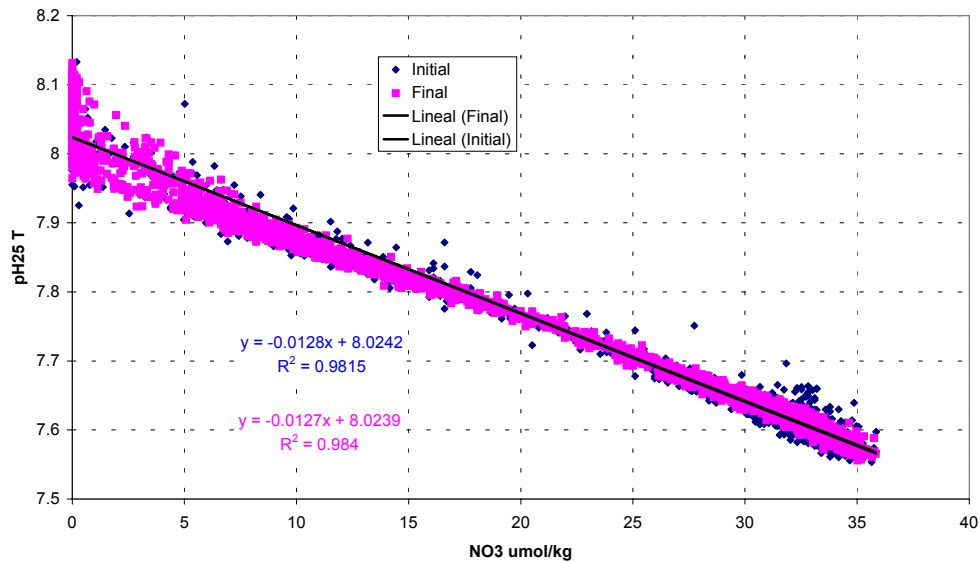
Figure 15. Relationship between NO3 (µmol/kg) and pH25T for the initial and final data base.

The Excel file attached (CD139 Corrections) to this report contains four sheets:

- Initial: contains the initial data base and the calculation of the residuals.
- Final: contains the final data base and the calculation of the residuals.
- Compar Data: compares both data sets, contains the corrections applied to each variable.
- Mean Res by station: contains the mean residual values for each BioGeo Var for the initial and final data sets.

The xls file also contains several graphs…. Have a look.