GoodHope-2004

Chemical variables comments

April 2006

Marta Álvarez.

Facultad de Ciencias, Univ.Vigo, Campus Lagoas-Marcosende s/n, 36200, Vigo, Spain.

martaalvarez@uvigo.es

After the whole bottle data from the cruise, temperature, salinity, dissolved oxygen, nutrient salts, $pH25_T$ and alkalinity (TA) is compiled we can perform a more detailed analysis of the quality of our chemical data.

First we calculate NTA, the normalised to salinity TA, NTA=35·TA/Sal so that, we can observe changes in TA not associated with salinity. Then we perform a visual analysis station by station of the pH and NTA profiles with depth and also their relation with potential temperature (Tpot), salinity (Sal) or nutrients, nitrate (NO_3), phosphate (PO_4) and silicate (SiO_2) and oxygen (O_2). In this way we can detect bad fired bottles which show anomalous values for all the biogeochemical variables.

After this first visual inspection, a residual analysis is done to detect any bias in the data from wrong calibrations, technical problems or bad calculations:

- every biogeochemical variable is calculated as a multiple linear regression from Tpot, Sal, pressure, latitude and longitude. See following table.

Table. *Regression coefficients, R^2 and standard error of the residuals for each variable.*

	NO_3	PO_4	SiO_2	O_2	pH25T	NTA
Intercept	-53.64	-8.76	-1589.76	4124.65	10.65	2368.09
Tpot (ºC)	-2.18	-0.15	-4.36	9.30	0.03	-2.96
Sal (psu)	3.01	0.31	42.86	-115.10	-0.09	-0.66
Pres (dbar)	-0.00147	-0.00012	0.01101	0.01524	0.00003	0.01
Lat (º)	0.31	-0.01	-3.42	-0.81	0.002	-0.39
Lon (º)	-0.16	0.02	1.63	-1.22	-0.01	-0.94
R^2	0.87	0.87	0.92	0.57	0.88	0.93
STD	2.96	0.22	10.12	26.52	0.04	6.41

- the mean and standard deviation of the corresponding residuals for each station are calculated.

- within each station samples far from the mean plus/minus one std are further inspected to check any inconsistent value.

- the tendency of the residual mean along the cruise track for the set of variables is inspected to detect any problem for each station subset and variable.

Here are some figures showing the residuals distribution along the GH-2004 cruise for different variables.

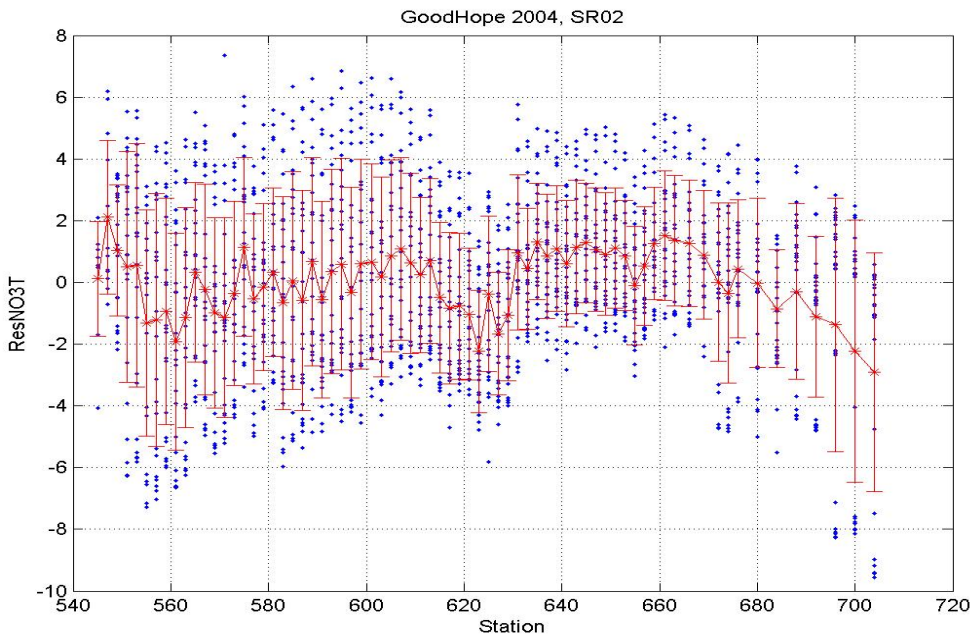

Figure 1. NO_{3T} (umol/kg) residuals as blue dots and the mean and standard deviation residual in red for every station along the GH2004 cruise. Note the spikes up and down and some descending and ascending trends in the red line.

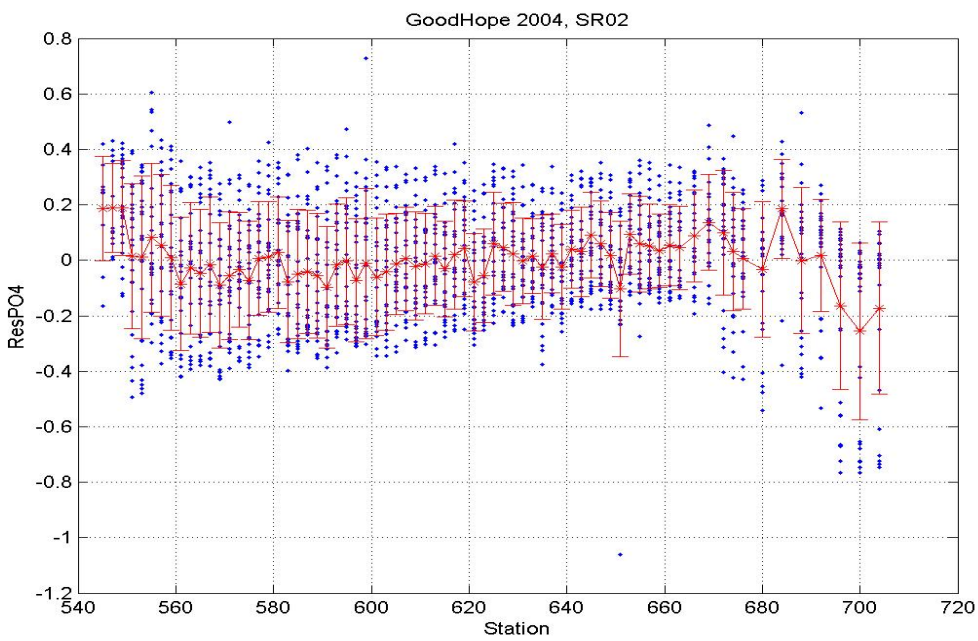

Figure 2. PO_4 (umol/kg) residuals as blue dots and the mean and standard deviation residual in red for every station along the GH2004 cruise. Note some outliers.

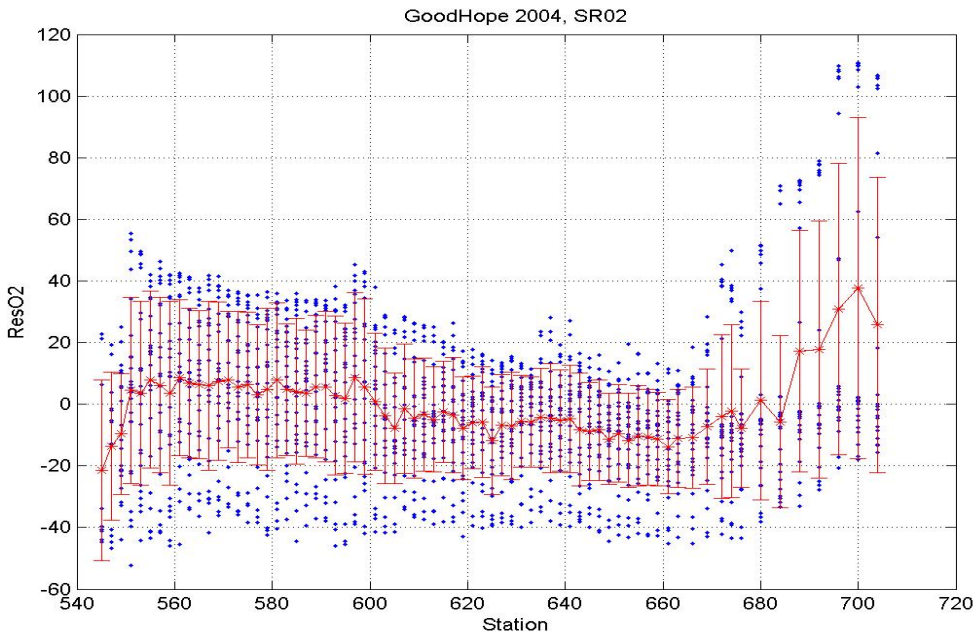

Figure 3. O_2 (umol/kg) residuals as blue dots and the mean and standard deviation residual in red for every station along the GH2004 cruise.

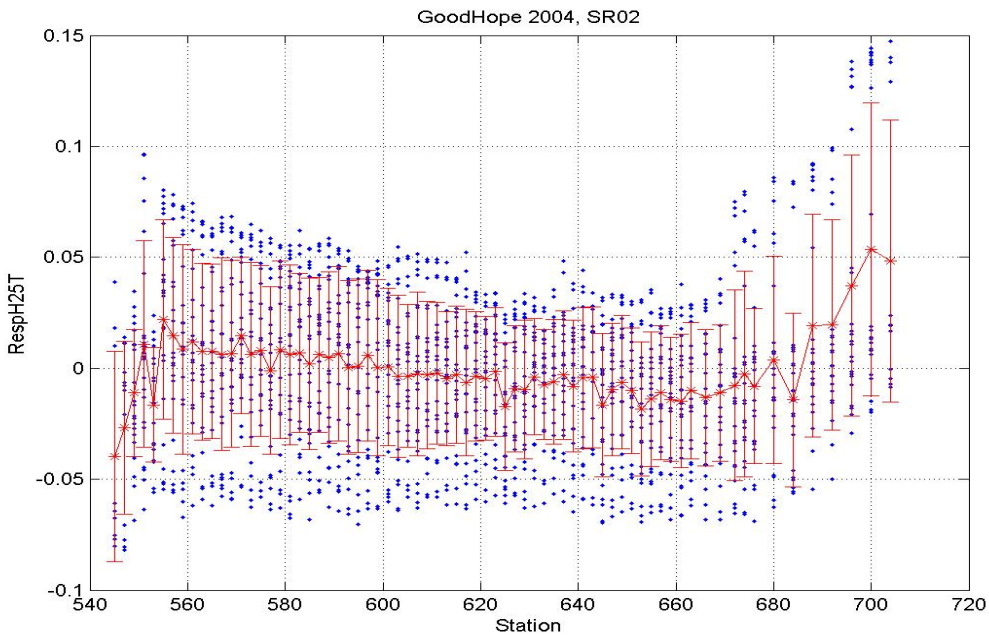

Figure 4. Individual $pH25_T$ residuals as blue dots and the mean and standard deviation residual in red for every station along the GH2004 cruise. Note the similarity between the pH and O_2 mean residual trends.

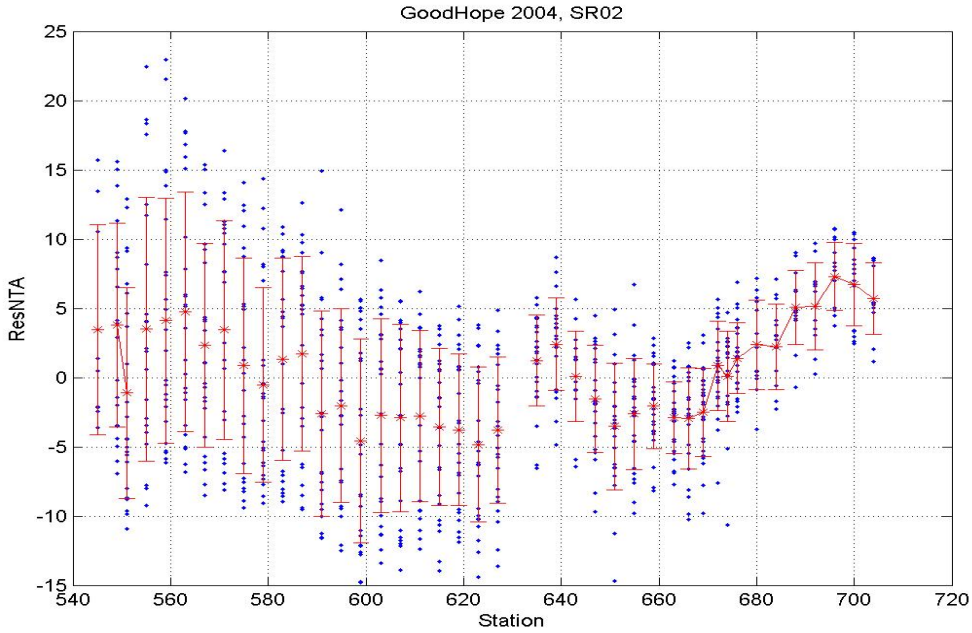

Figure 5. Individual NTA (umol/kg) residuals as blue dots and the mean and standard deviation residual in red for every station along the GH2004 cruise.

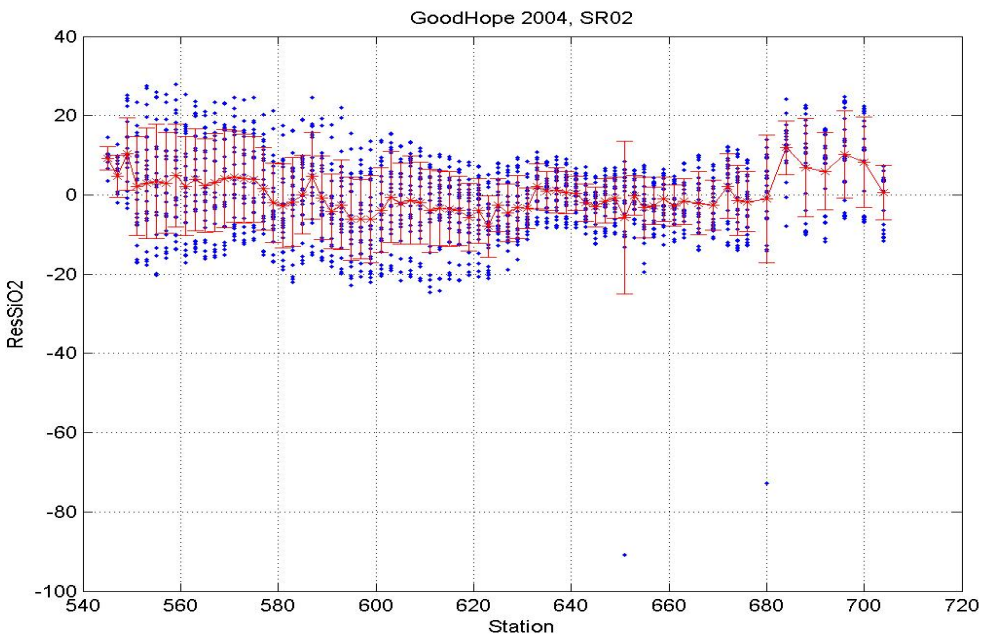

Figure 6. Individual SiO_2 (umol/kg) residuals as blue dots and the mean and standard deviation residual in red for every station along the GH2004 cruise. Note the outliers.

The residuals for the conservative variables NO (=O_2 + 10.6· NO_3) and PO (= O_2 +170·PO_4) are also shown, as they are conservative their trends could better reflect any problem with the analysis or data.

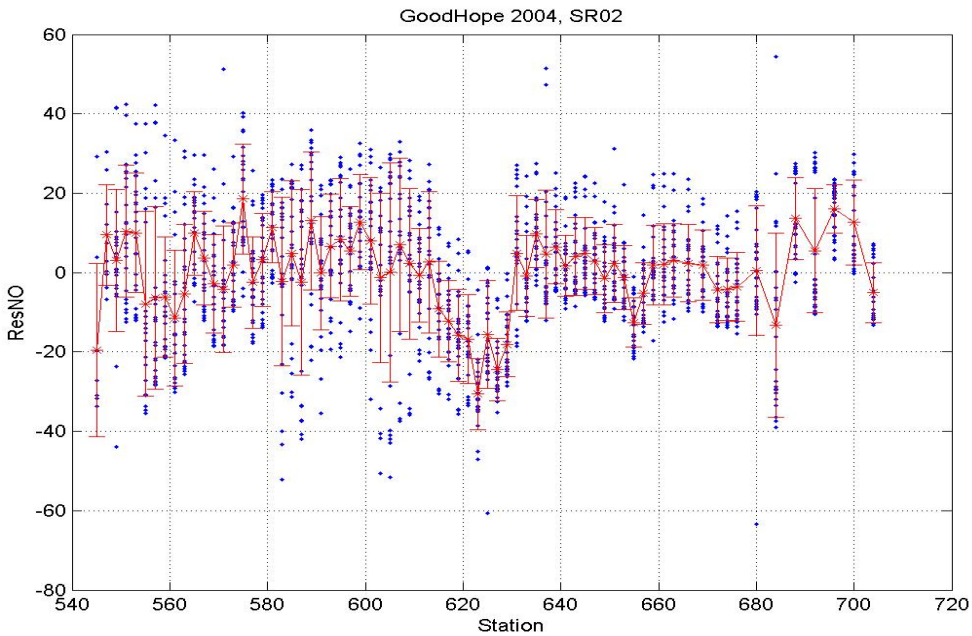

Figure 7. Individual NO (umol/kg) residuals as blue dots and the mean and standard deviation residual in red for every station along the GH2004 cruise. Note some outliers also seen in the NO_3 figure and the trends and spikes very much resemble that of NO_3.

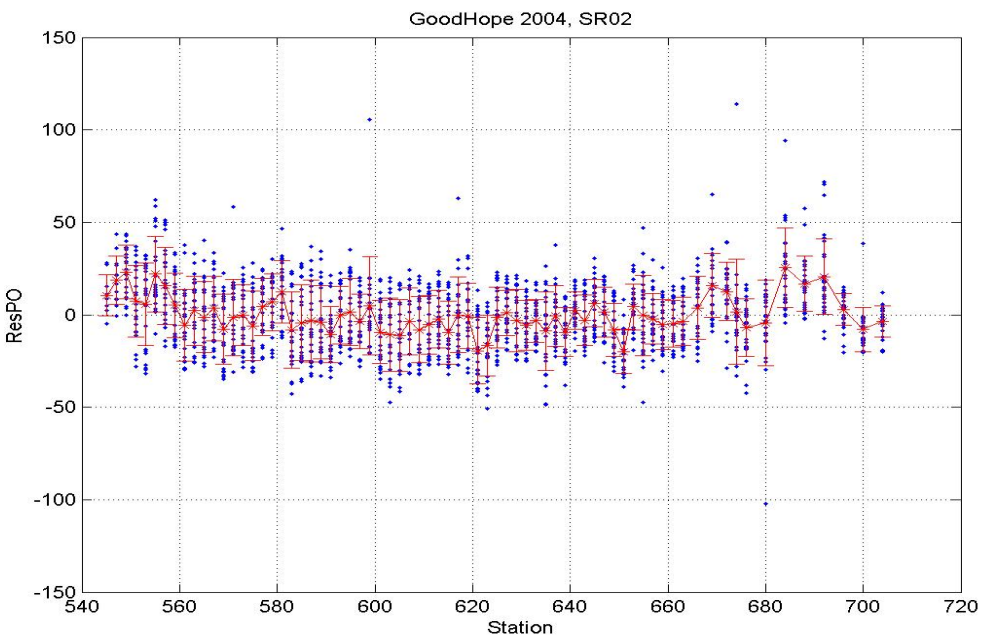

Figure 8. Individual PO (umol/kg) residuals as blue dots and the mean and standard deviation residual in red for every station along the GH2004 cruise. Note some outliers also seen in the PO_4 figure and the trends and spikes very much resemble that of PO_4.

Following are some plots showing just the mean residual by station, to check the consistency in the trends between variables. For example, note the consistency in the trends of O_2 and pH. On the other hand, PO_4 and NO_3 have similar trends but much noisier and inversely correlated to O_2 and pH. The SiO_2 and NTA trends are correlated but SiO_2 is noisier and they resemble somehow the pH residuals.

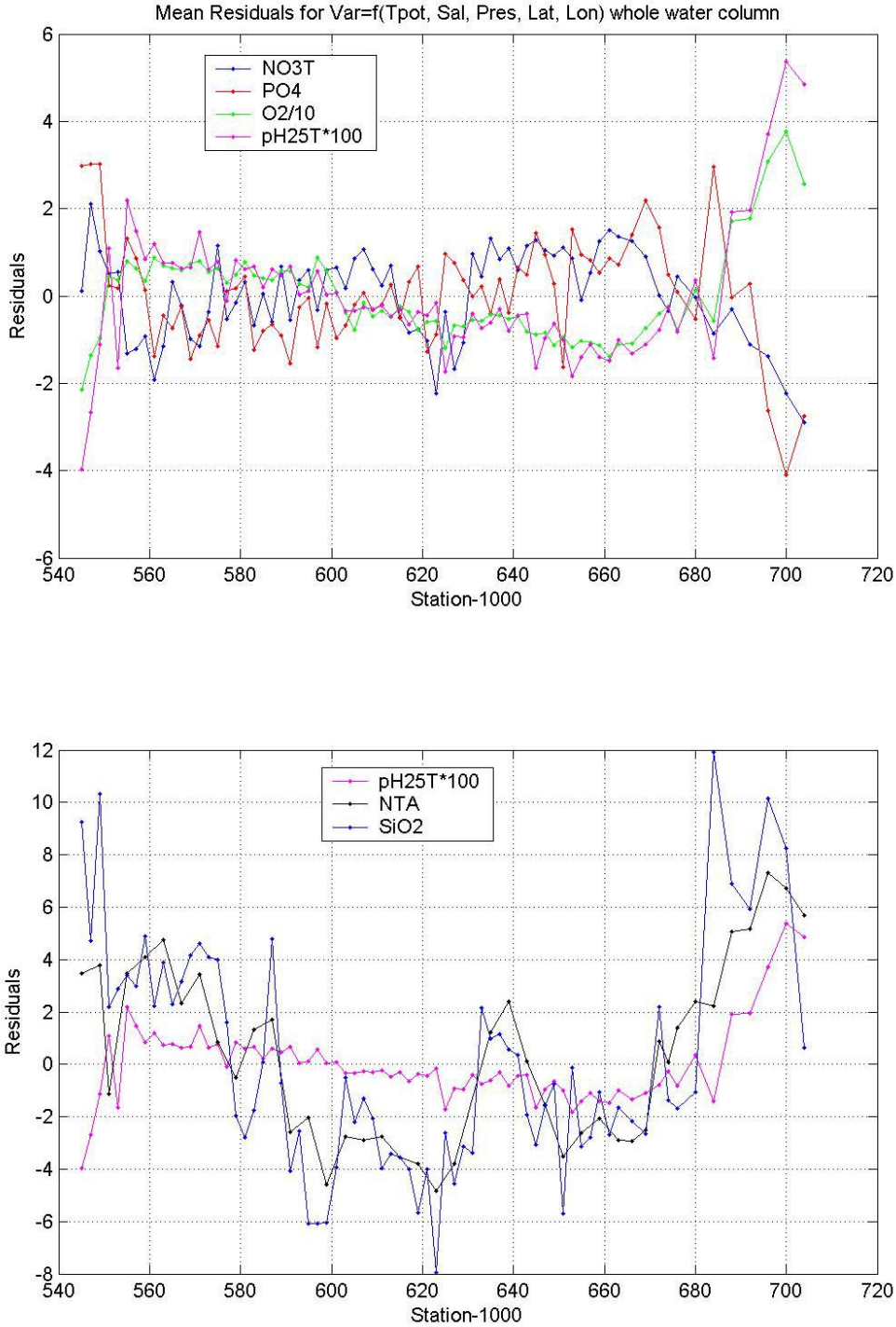

Figure 9. Mean residual by station for different variables, some of them are scaled, see the legend, all in umol/kg except pH. Note the consistency or inconsistency in the trends.

Here are the trends for the mean residuals of NO, PO and O_2. The NO and PO residuals are noisier than the O_2 ones so pointing to likely wrong adjustments in the calibration for the NO_3 and PO_4 analysis.

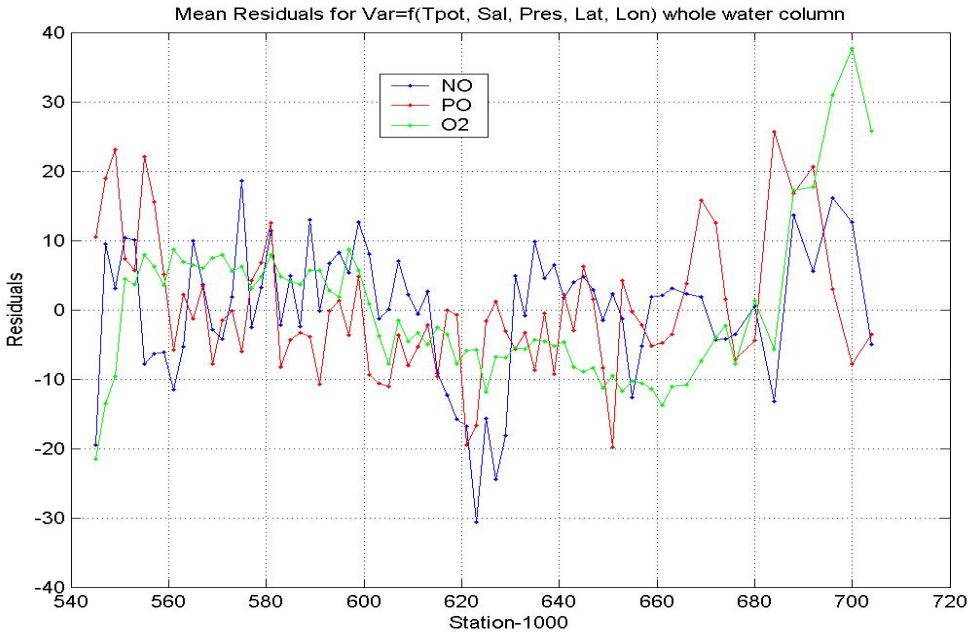

Figure 10. Mean residual by station for NO, PO and O_2, all in umol/kg. Note the consistency or inconsistency in the trends.

Here are some regression lines between the mean residuals by station. They give us an idea of the correlation between them and help to detect some “wrong” stations, those one far from the regression line. The R^2 for each line in the following subplots are 0.88, 0.35 and 0.48 with $p<0.001$ for all of them.

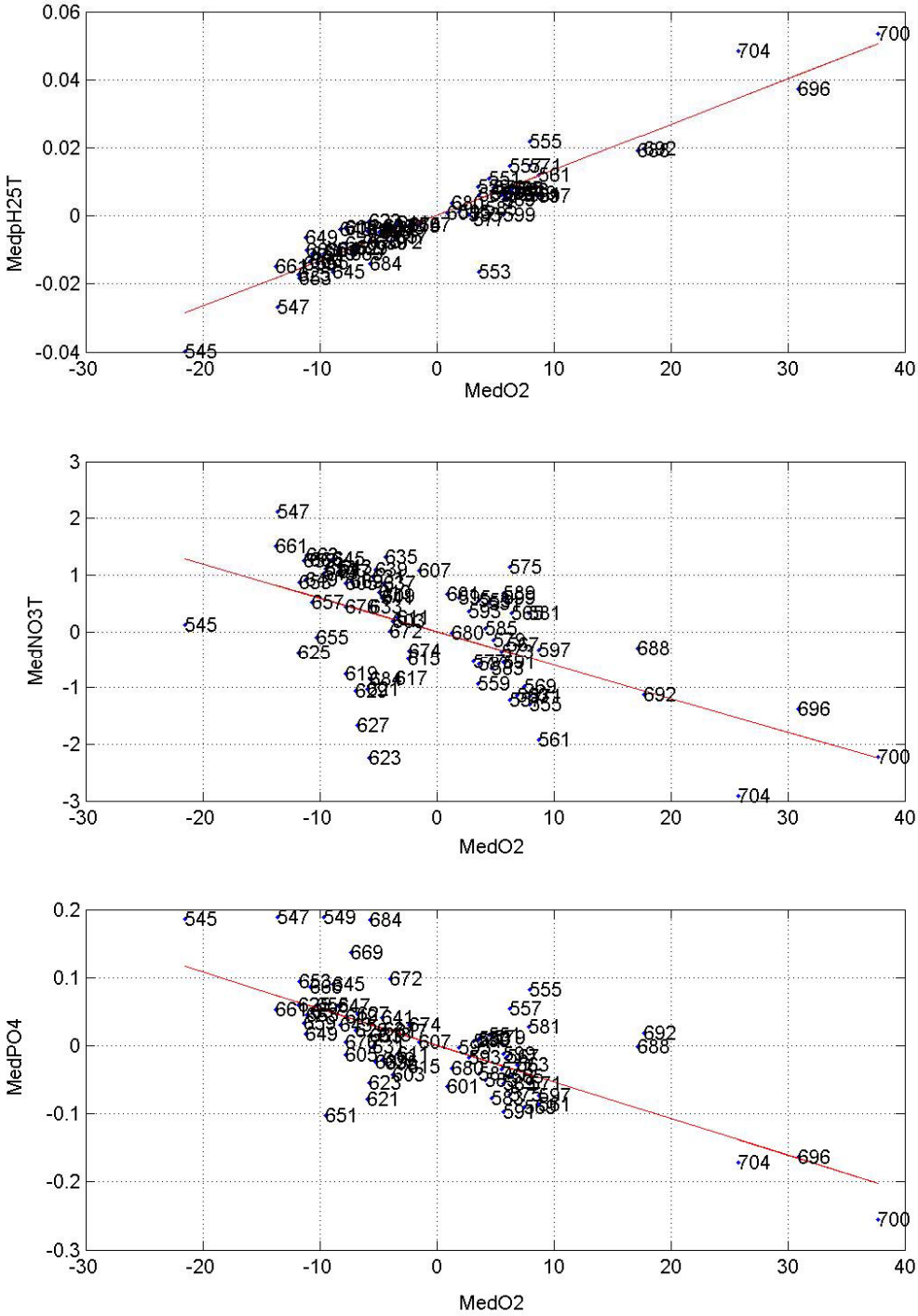

Figure 11. Correlation between mean residuals by station for different variables (in umol/kg, except pH). Numbers represent the station number – 1000.

More regression lines between the mean residuals by station. They give us an idea of the correlation between them and help to detect some “wrong” stations, those one far from the regression line. The R^2 for each line in the following subplots are 0.13 ($p<0.01$), 0.39 and 0.69 with $p<0.001$for the second and third ones, respectively.

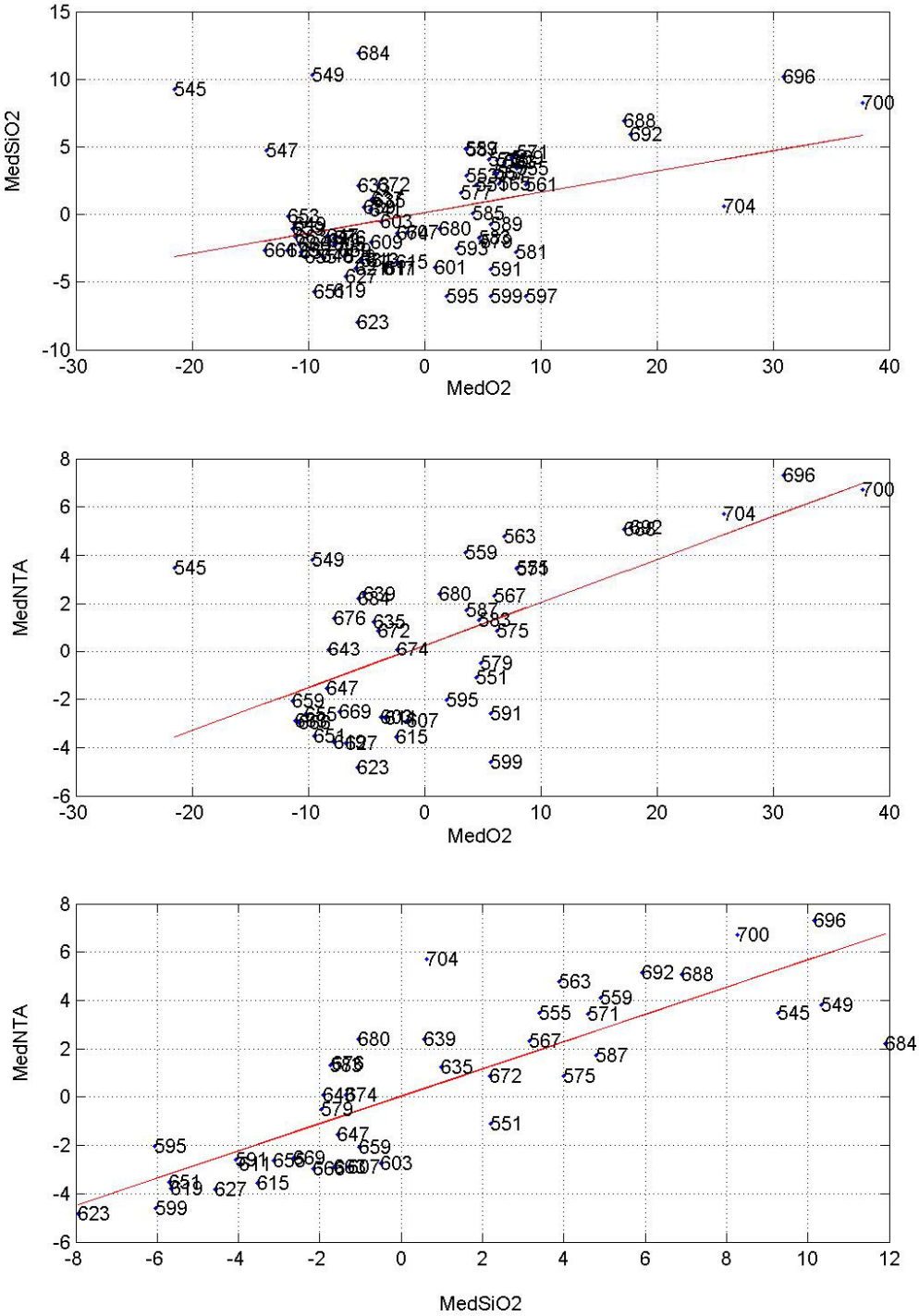

Figure 12. Correlation between mean residuals by station for different variables (in umol/kg). Numbers represent the station number – 1000.

Finally, I would suggest a careful checking of the mean residual trends along GH2004 for the different variables, specially the nutrients and confirm whether the spikes correspond to any change in the standards or substandards. There is a natural variability in the nutrient content of the water masses, it reflects mixing and biological activity, and consequently a correlation between those changes (mean residuals by station) is expected between the NO_3 and PO4 and O_2. The regression lines shown before could help to detect "wrong" stations, as I did with the final set of TA measurements from station 680 till the end.

Here is a list of some samples I inspected visually and considered dubious.

Station	Bottle	Comments	
545	all	NO3-PO4 relation low	
549	19	O2 low	
549	12	NO3 low	
551	14	O2 high	
557	21	NO3 &PO4	
557	22	NO3 &PO4	
559	2	bad bottle?	
561	between 700-1000		O2 change
561	21	PO4 low	
571	18	NO3 & PO4 high	
573	5	PO4 high	
575	6	PO4 low	
575	1	PO4 low	
575	21	NO3 & PO4 high	
579	1000 m	O2 spike	
583	22	PO4 high	
599	17	PO4 high	
617	2	PO4high	
617	12	PO4high	
619	20	PO4 high	
625	24	NO3 low	
633	24	NO3 high	
635	24-21	PO4 low	
651	5	PO4 low	
659	1-2	PO4 ??	
674	22	PO4 high	
674	21	PO4 high	
680	1	NO3 low	
680	1	bad bottle	
688	14	PO4 high	

Best regards to all.

Marta.