

Département Océanographie et Dynamique des Ecosystèmes
Laboratoire Environnement Ressources Provence Azur Corse

Rosalie FUCHS
Ivane PAIRAUD

Janvier 2014 - RST.ODE/LER-PAC/14-25



Ifremer

Study of high frequency measurements from the MesuRho mooring

Fiche documentaire

Numéro d'identification du rapport : RST.ODE/LER-PAC/14-25 Diffusion : libre : <input checked="" type="checkbox"/> restreinte : <input type="checkbox"/> interdite : <input type="checkbox"/>		date de publication : 27/01/2014 nombre de pages : 39 bibliographie : oui illustration(s) : oui langue du rapport : Anglais
Validé par : ANDRAL Bruno Adresse électronique : bruno-andral@ifremer.fr		
Titre de l'article		
Contrat n°		Rapport intermédiaire <input type="checkbox"/> Rapport définitif <input checked="" type="checkbox"/>
Auteur(s) principal(aux) : FUCHS Rosalie PAIRAUD Ivane	Organisme / Direction / Service, laboratoire LER PAC	
Encadrement(s) : PAIRAUD Ivane		
Cadre de la recherche : Projet FP7 PERSEUS		
Destinataire :		
Résumé <p>La station haute fréquence MesuRho appartenant au réseau MAREL et située à l'embouchure du Rhône est présentée. Les mesures présentant parfois des signes de biais, de dérive ou de doubles, une interface graphique à été mise en place afin de visualiser rapidement les données et éventuellement de les corriger. Les jeux de données d'origine et corrigées 2009-2013 sont présentés, les méthodes de corrections mise à disposition sont détaillées. Enfin, l'étude d'un évènement de tempête (coup de vent d'Est) entre le 4 et le 7 Mars 2013 est présentée à titre d'exemple, et quelques perspectives en vue d'améliorer la gestion des données sont proposées.</p>		
Abstract <p>The high frequency MesuRho station, part of the MAREL network and located at the mouth of the Rhone River is presented. Measures sometimes show bias signs, drift or double, so that a graphical interface was implemented to quickly visualize data and eventually correct them. Original data set and corrected data for 2009-2013 are presented, the methods of corrections provided are detailed. Finally an example of a storm event study (under strong south east wind) between the 4th and the 7th of March 2013 is presented and some prospects to improve data management are proposed.</p>		
Mots-clés MesuRho, code qualité, correction des données, analyse et traitement de données, filtrage, ACP, Fourier, évènement de vent d'Est		
Words keys MesuRho, quality code, data corrections, data and treatments analysis, filtering, PCA, Fourier, storm event		

TABLE OF CONTENT

1. General presentation	7
1.1. MesuRho Mooring localization and context.....	7
1.2. Sensors' characteristics	8
1.2.1. SMATCH multi-parameters probes.....	9
1.2.2. ADCP (currents and waves)	10
2. Comments on data and previous treatment methods	12
2.1. Overview of SMATCH probes measurements	12
2.2. Problems identification	15
2.3. The Coriolis Data Base	15
2.4. Some correction methods	16
2.4.1. Bias correction	16
2.4.2. Double correction	17
2.4.3. Drift correction.....	19
3. Development of analysis tools	21
3.1. Matlab Buoy Interface (MBI) development.....	21
3.2. Data analysis tools included in the interface	21
3.2.1. Linear regression	21
3.2.2. Filters	22
3.2.3. Fourier series (from Emery and Thomson (2004))	24
3.2.4. Principal Component Analysis	26
4. Data treatment (from SMATCH multi-parameters sensors).....	29
4.1. Temperature.....	29
4.2. Salinity	29
4.3. Fluorescence (chlorophyll-a or phycocyanine).....	30
4.4. Turbidity.....	30
5. Example of study : storm event (5-7 March 2013).....	31
6. Future improvements	33
References	35
Appendix.....	36

[List of figures](#)

Figure 1-1: Position of the high frequency acquisition MesuRho buoy in the Gulf of Lions.....	7
Figure 1-2: ADCP with the NEMO waves module.....	10
Figure 2-1: Instruments immersion date depending on their ID number for SMATCH probes	12
Figure 2-2: Temperature data from ABIN mails and Coriolis base	12
Figure 2-3: Salinity data from ABIN mails and Coriolis base	13
Figure 2-4: Fluorescence data from ABIN mails and Coriolis base	13
Figure 2-5: Turbidity data from ABIN mails and Coriolis base	14
Figure 2-6: Dissolved Oxygen data from ABIN mails and Coriolis base	14
Figure 2-7: Depth/Pressure data from ABIN mails and Coriolis base	14
Figure 2-8: Chlorination data from ABIN mails and Coriolis base	15
Figure 2-9: Quality Control Code from Coriolis web base	15
Figure 2-10: Example of biased salinity data series.....	16
Figure 2-11: Example of bias corrected salinity data series.....	17
Figure 2-12: Example of double correction on salinity data series with the 'From date' option	17
Figure 2-13: Example of double correction on salinity data series with the 'From data' option	18
Figure 2-14: Salinity data series with drift signal	19
Figure 2-15: Drift correction method illustration.....	19
Figure 2-16: Salinity data series corrected from drift	20
Figure 2-17: Comparison of in situ data from CTD with corrected salinity data series .	20
Figure 3-1: The Matlab Buoy Interface (MBI) showing surface temperature data compared with CTD data	21
Figure 3-2: Example of linear regression on sub-surface temperature data series.....	22
Figure 3-3: Example of QC filter on salinity data series.....	23
Figure 3-4: Example of simple and double exponential smoothing on temperature data	23
Figure 3-5: Example of Fourier series on salinity data, with the 15th first components	25
Figure 3-6: Details on the 6th first components of the Fourier series.....	25
Figure 3-7: ACP data matrix	26
Figure 3-8: Eigenvalues of the diagonal matrix and cumulative explained variance	27
Figure 3-9: Variables and observations representation in the reference defined by the two first principal components/axis	28
Figure 4-1: Surface and bottom temperature data with control quality code colours	29
Figure 4-2: Surface and bottom salinity data with control quality code colours.....	29
Figure 4-3: Surface and bottom fluorescence data with control quality code colours	30
Figure 4-4: Surface and bottom turbidity data with control quality code colours.....	30
Figure 5-1: Rhone river flow and SMATCH probe salinity data (a), weather station data (b, c) during the storm event	32
Figure 5-2: Wave heights and surface elevation derived from ADCP data (a), SMATCH probe fluorescence and turbidity data at the sub-surface (b) and at the bottom (c)	32
Figure 5-3: MODIS Chlorophyll satellite data before and after the storm event	33

[List of tables](#)

Table 1-1: multi-parameters SMATCH probe characteristics	9
Table 1-2: ADCP RDI's Workhorse 600 kHz characteristics	10

1. General presentation

1.1. MesuRho Mooring localization and context

The Rhone River (situated in the North Western Mediterranean Sea in the Gulf of Lions) being the main source of fresh water input in the Mediterranean Sea, brings large quantities of particles, from natural to anthropogenic origins, which could affect the coastal marine environment (Pinazo *et al.* (2013), Lorthiois (2012), Moutin *et al.* (1998)). Since 2009, the french institute IFREMER manages the acquisition of in situ physico-chemical data in front of the Rhone River mouth via the MesuRho instrumented observatory in order to characterize the Gulf of Lions inputs (Figure 1-1). The high frequency acquisition of the MesuRho mooring enable to study sedimentary dynamics, the influence of extreme events (storms, flood events) on particulate/dissolved organic and inorganic inputs into the marine coastal area as well as sediment resuspension events. The MesuRho mooring is part of the IFREMER MAREL (Mesures Automatisées en Réseau pour l'Environnement et le Littoral) network and of several national and european research programs such as the observation network MOOSE (Mediterranean Ocean Observing System on Environment), the MISTRALS MERMEX program (Marine Ecosystems Response in the Mediterranean Experiment), the AMORAD ANR (Amélioration des MODèles de prévision de la dispersion et d'évaluation de l'impact des RADionucléides au sein de l'environnement), and the FP7 PERSEUS project (Protecting European SEas and borders through the intelligent USE of surveillance).



Figure 1-1: Position of the high frequency acquisition MesuRho buoy in the Gulf of Lions

The issues raised in the different projects are mainly related to Rhone River inputs characterization and to the sedimentary dynamics at the river mouth. As a LERPAC lab (IFREMER) initiative, associated with the LDCM lab (IFREMER Brest), the MesuRho project was launched in 2009 with several partners: the service of Headlights and Tags of the Departmental Direction of Equipment in the PACA region (DDE13), The Institute for Radiological Protection and Nuclear Safety (IRSN), The National Institute of Sciences of the Universe (INSU), The National Centre for Scientific Research (CNRS), The Mediterranean Institute of Oceanography (MIO), The Centre for Maritime and River Technical Studies (CETMEF), The European Centre for Research and Teaching of

Environmental Science (CEREGE) and The Laboratory of Climate Sciences and the Environment (LSCE).

A monitoring committee involving each body has been established to ensure the proper functioning of the system.

The MesuRho station has been operational since June 2009 and is located at the Buoy Float Immersed (BFI) maritime buoyage Roustan East ($43^{\circ} 19.2$ N, $4^{\circ} 52$ E , depth 20 m), one of the two buoys indicating the Rhone prodelta . Located at the mouth of the river, this station is configured to collect physicochemical data in near real time and at high frequency (about 30 min) in the fresh waters/marine waters transition zone. Since its installation, the instrumentation is enriched over the years to obtain a more complete set of data.

The station was originally equipped with a weather station associated with a PAR sensor (Photosynthetic Available Radiation) in the air, and a multi-parameters probe SMATCH (temperature, pressure (depth), conductivity (salinity), turbidity, fluorescence (chlorophyll), dissolved oxygen) in the subsurface. In June 2010, a second multi parameters probe SMATCH and an Acoustic Doppler Current Profiler (ADCP) were added at the bottom. Finally, in 2012, a STPS sensor associated with a nitrate probe was added in sub-surface, as well as a benthic station for measuring dissolved oxygen in the sediment. In the near future, atmospheric and subsurface sensors of radioactivity will be attached to the buoy. The weather station and the SMATCH probes are operated by IFREMER, the ADCP is operated by IFREMER and IRSN, the nitrate sensor is operated by the MIO and the benthic station is operated by LSCE.


The SMATCH probes, the ADCP, the PAR sensor, the weather station and the benthic station are connected by a cable to a controller disposed in the atmosphere, powered by solar panels. The measurements are transmitted to the Coriolis data center via GPRS six times a day (about 1 transmission/4h) .

1.2. Sensors' characteristics

The IFREMER LERPAC and LDCM labs have five multi-parameters SMATCH probes, with two probes at sea when the three others are in calibration and/or stored. They also have three ADCPs for sensors rotation.

1.2.1. SMATCH multi-parameters probes

Table 1-1: multi-parameters SMATCH probe characteristics

	<u>Sensor</u>	<u>Characteristics</u>
	Depth	range 0 to 20m accuracy < 0.06m resolution 0.006m
	Temperature	range -5°C to +35°C accuracy < 5m°C in the range 0-20°C resolution 1m°C at 10°C
	Conductivity	range 0 to 70 mS/cm accuracy < 0.05 mS/cm in the range 0 to 60 mS/cm resolution 0.0012 mS/cm
	Salinity	range 2 to 42‰ at 20°C accuracy ±0.1‰ resolution 0.0011‰
	Dissolved Oxygen	Saturation: range 0 to 120% accuracy < 5% resolution 0.01% Concentration: 0 to 16mg/l accuracy < 5% resolution 0.01%
	Turbidity	Range (configurable at works) 25, 125, 500 or 750 FTU / 2.5 FTU on demand Linearity < 2% deviation 0-750 FTU Light source Wavelength: 880 nm
	Fluorescence	Chlorophyll a - Turner Design optical sensor Minimum detection limit: 0.03 µg/l Sensitivity: 0-5µg/L or 0-50

		$\mu\text{g/L}$ or 0-500 $\mu\text{g/L}$ Wavelength: Excitation 460 nm Emission 620-715 nm
	pH (not available at MESURHO)	Range 0 to 14pH Resolution 0.0003pH
	Time	Internal clock with calendar (clock drift < 1mn/month)

1.2.2. ADCP (currents and waves)

The ADCPs are Teledyne RDI's Workhorse Monitor 600 kHz associated with the waves array and the NEMO wave processing array for near real time processing and transmission of waves and currents (Figure 1-2). Sensor characteristics are described in

Table 1-2

Figure 1-2: ADCP with the NEMO waves module



Table 1-2: ADCP RDI's Workhorse 600 kHz characteristics

Water profiling	Typical range 50m 600 kHz	
<u>Vertical resolution</u>	<u>range</u>	<u>Std. Dev.</u>
0.5 m	38 m	14.0 cm/s
1 m	42 m	7.0 cm/s
2 m	46 m	3.6 cm/s
4 m	51 m	1.8 cm/s
Long Range Mode		
<u>Vertical resolution</u>	<u>range</u>	<u>Std. Dev.</u>
4 m	66 m	3.6 cm/s
Profile Parameters		
Velocity Accuracy	0.3% of water velocity relative to ADCP $\pm 0.3\text{cm/s}$	
Velocity resolution	0.1 cm/s	
Velocity range	$\pm 5\text{m/s}$ default, $\pm 20\text{m/s}$ max	
Number of depth cells	1-128	

Ping rate	2 Hz (typical)
Echo Intensity Profile	
Vertical resolution	Depth cell size (user configurable)
Dynamic range	80 dB
Precision	±1.5dB
Transducer and Hardware	
Beam angle	20°
Configuration	4 beam, convex
Internal memory	Two PCMCIA card slots; no memory card included
Communication	Serial port selectable by switch for RS-232 or RS-422. ASCII or binary output at 1200-115,200 baud
Environmental	
Standard depth rating	200m; optional to 500m, 1000m, 6000m
Operating temperature	-5° to 45°C
Storage temperature	-30° to 60°C
Software	
	WinSC—Data Acquisition System; WinADCP—Data Display and Export
Standard Sensors	
Temperatures (mounted on transducer)	Range -5° to 45°C, Precision ±0.4°C, Resolution 0.01°
Tilt	Range ±15°, Accuracy ±0.5°, Precision ±0.5°, Resolution 0.01°
Compass (fluxgate type, includes built-in field calibration feature)	Accuracy ±2°5 Precision ±0.5°5 Resolution 0.01° Maximum tilt ±15°

2. Comments on data and previous treatment methods

2.1. Overview of SMATCH probes measurements

– SMATCH ID

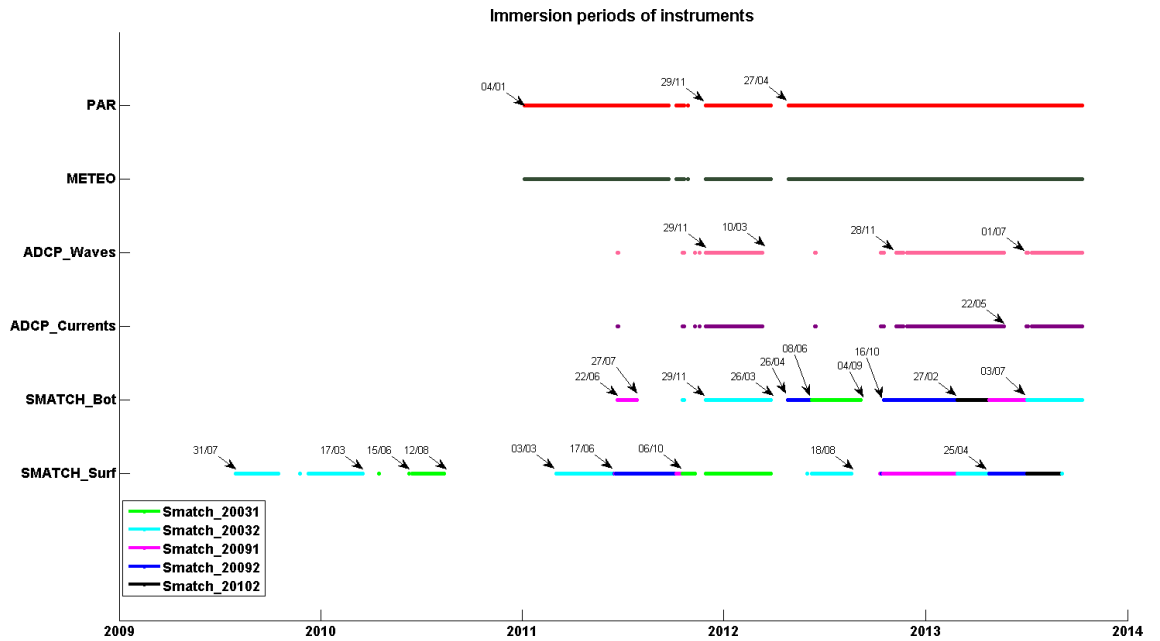


Figure 2-1: Instruments immersion date depending on their ID number for SMATCH probes

– Temperature

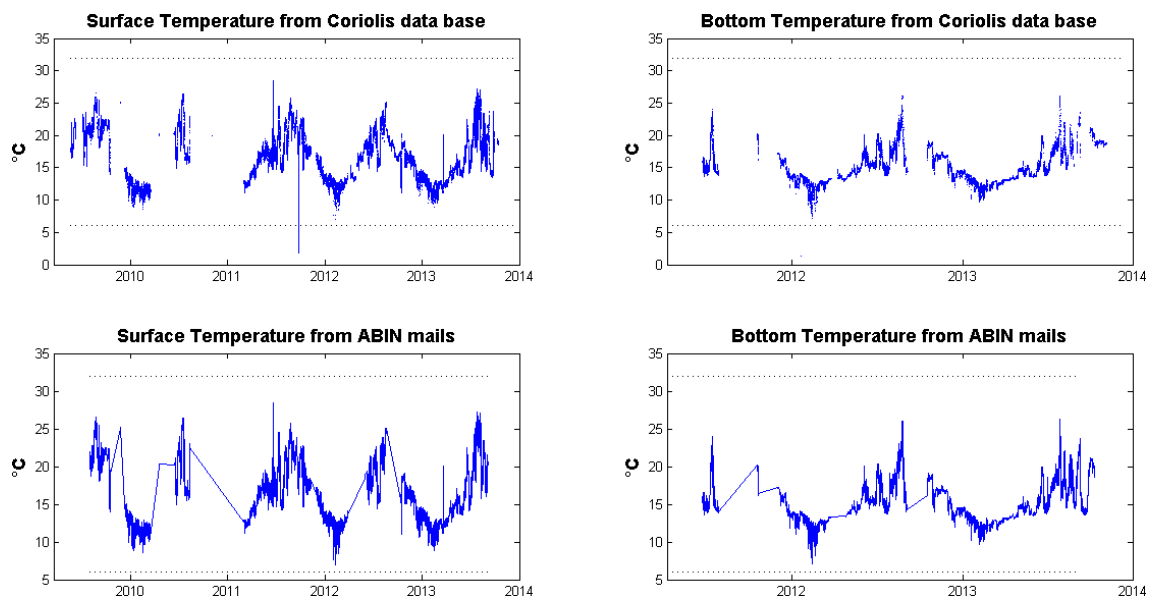


Figure 2-2: Temperature data from ABIN mails and Coriolis base

– Salinity

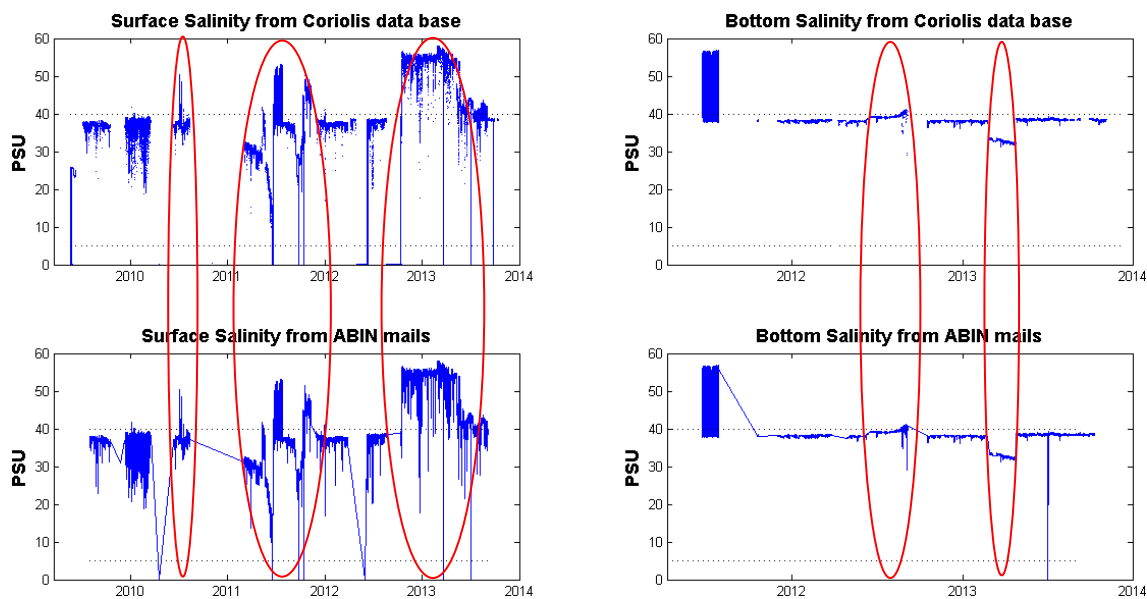


Figure 2-3: Salinity data from ABIN mails and Coriolis base

– Fluorescence

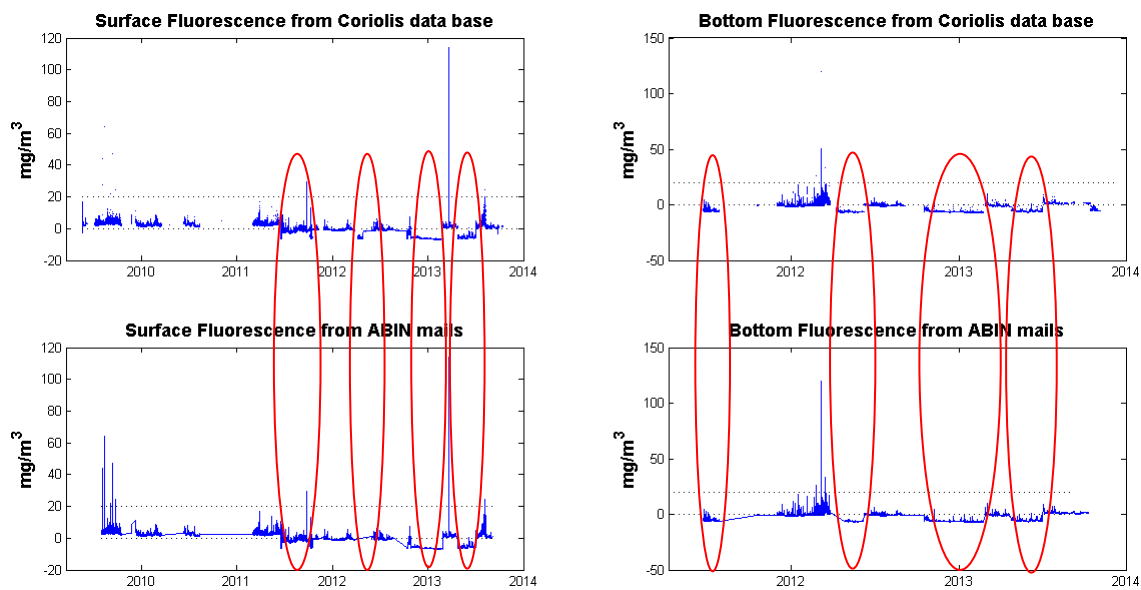


Figure 2-4: Fluorescence data from ABIN mails and Coriolis base

-Turbidity

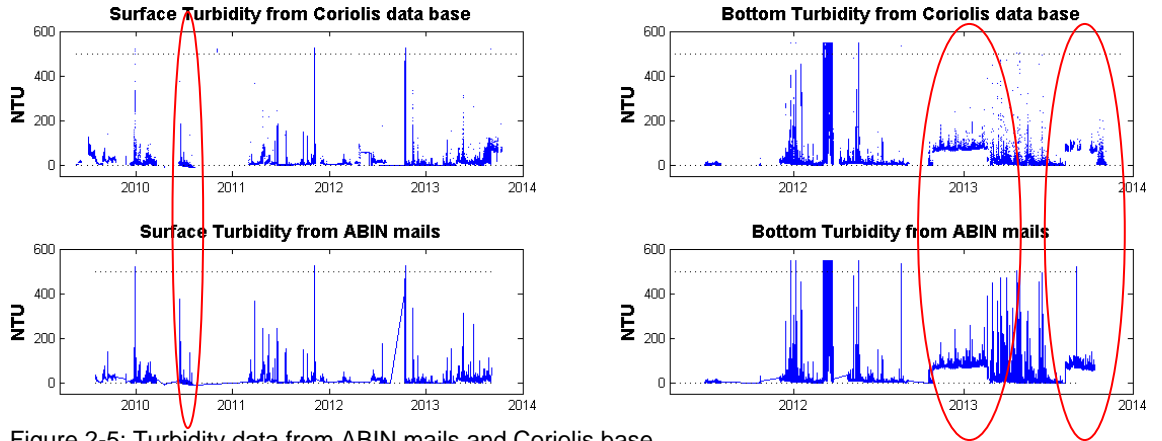


Figure 2-5: Turbidity data from ABIN mails and Coriolis base

-Dissolved Oxygen

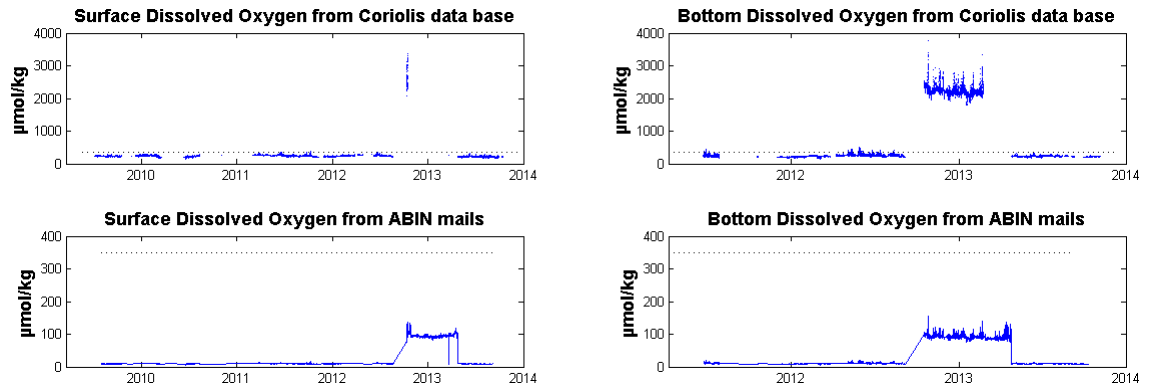


Figure 2-6: Dissolved Oxygen data from ABIN mails and Coriolis base

-Depth/ Pressure

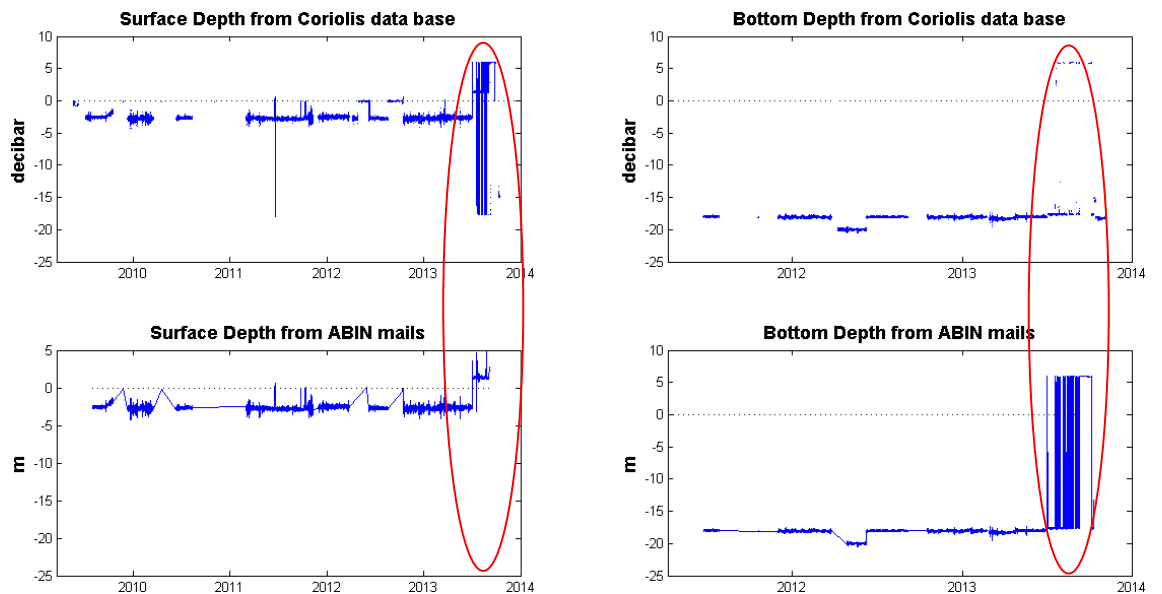


Figure 2-7: Depth/Pressure data from ABIN mails and Coriolis base

-Chlorination

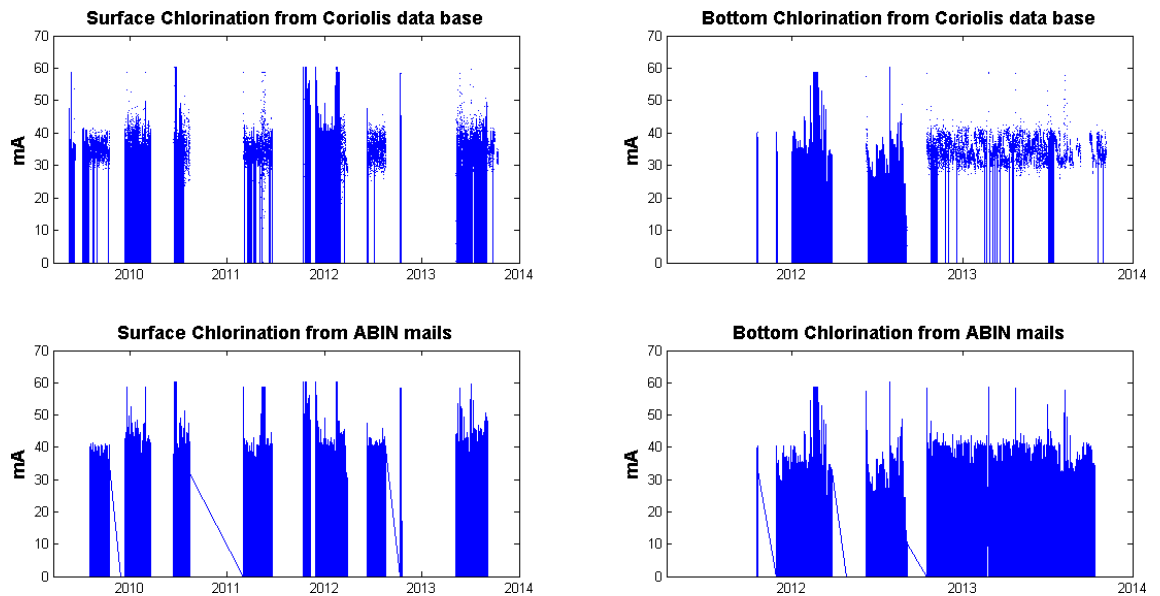


Figure 2-8: Chlorination data from ABIN mails and Coriolis base

2.2. Problems identification

Red circles in the figures of section 2.1 indicate some measures bias (positive and negative, Figure 2-3, Figure 2-4, Figure 2-5), sensors drift (Figure 2-3) or double measurements. Previous studies tackled that question and suspected interferences between the different kind of instruments (SMATCH, ADCP, GPRS automate,...), isolation default or cable damages to be responsible for these bad data profiles.

2.3. The Coriolis Data Base

The MesuRho mooring measures are transmitted 6 times a day to the Coriolis data-centre and stored in a data base, which is accessible at <http://www.ifremer.fr/co-en/>. The Coriolis web base includes a Quality Control (QC) Code (Figure 2-9) associated with data sets, depending on valid bounds known in the concerned area. As an example, the salinity valid bounds at MesuRho station are 5-40 (PSU) as it is in the Mediterranean Sea and in front of Rhone River mouth. It happens that measures are inside valid bounds but look uncertain as shown above (large drift, double measures...). In these cases, it would be useful to have the possibility to update, at least, QC codes, especially for data that may be used for models validation.



Figure 2-9: Quality Control Code from Coriolis web base

2.4. Some correction methods

A previous study (Lachaise, 2013) identified three 'intuitive' methods enabling to correct temporal data time series in order to obtain more or less homogeneous and more realistic series. Since dedicated oceanographic campaigns are sparse, very few in situ data comparisons between values from mooring sensors and from external sensors can be done. As consequences, corrections using following methods should be use with caution, considering the corrected data series for qualitative aspects more than for precise quantitative applications (like model validation,...) when additional data needed for correction validation are missing.

In the future, if several annual data time series are available, it would be possible to correct data series providing uncertainties bounds depending on previous data series at the same period. Neural network methods are also good candidates for correcting data, and/or fill missing values based on water masses characteristics.

2.4.1. Bias correction

The method implemented to correct biased data is close to a Moving Average Method applied on both reference series and the series that have to be corrected. The reference series is divided in intervals on which the maximum (or minimum) is calculated on the percentile 95 of the data contained in the interval (Figure 2-10). Then, the mean of all maximum found on intervals is calculated and kept for future correction. The same methodology is applied to the series that have to be corrected, and each intervals data are corrected from the difference between the mean of reference series max and the max found on the percentile 95 of the considered interval on the series to correct (Figure 2-11).

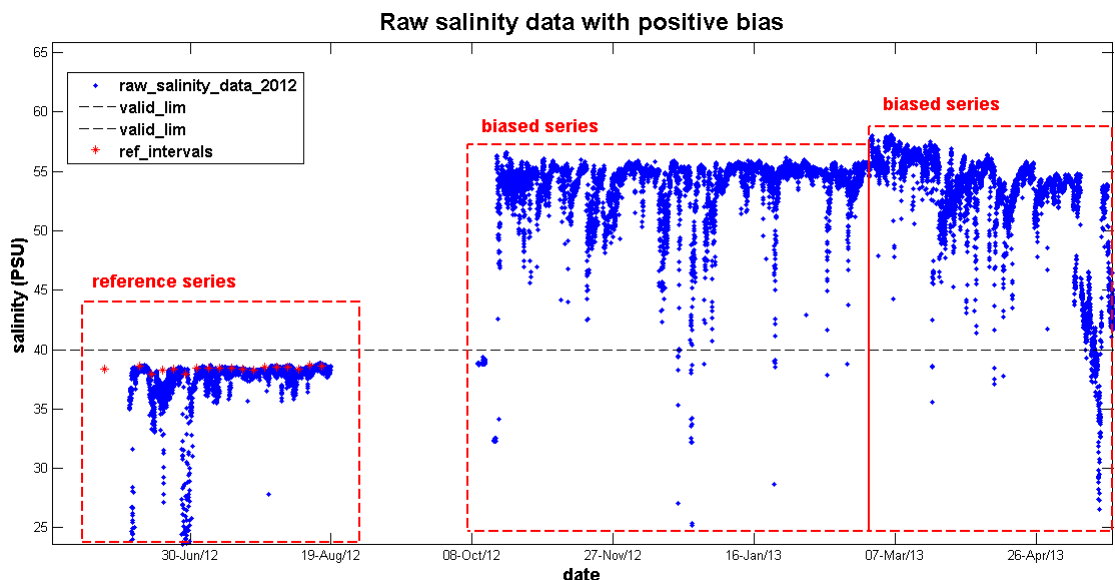


Figure 2-10: Example of biased salinity data series

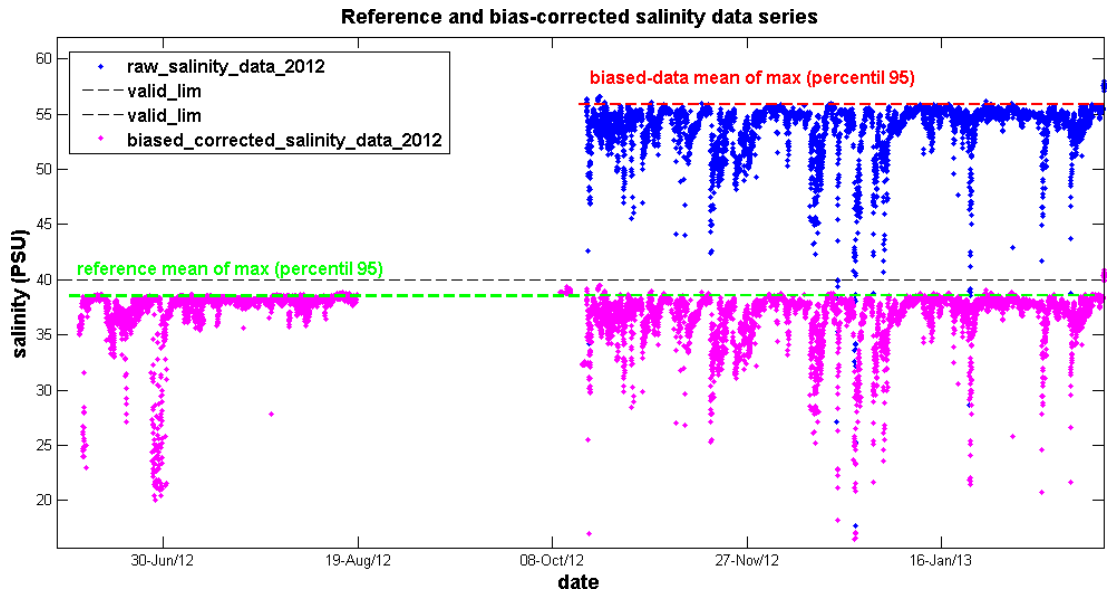


Figure 2-11: Example of bias corrected salinity data series

2.4.2. Double correction

Some data series showed double measures leading to data series overlapping for the same period. In order to retain only one data series, a selection occur depending on dates and differences between the measures.

- **'From date'** option: Starting at a data point with date '*date1*', we look for data points that are present between '*date1*' and $date_next=date1+dt$. If several data points are present in this time interval, the one that have at least $2/3*dt$ time gap and minimum difference with the previous retained measure is retained (Figure 2-12).

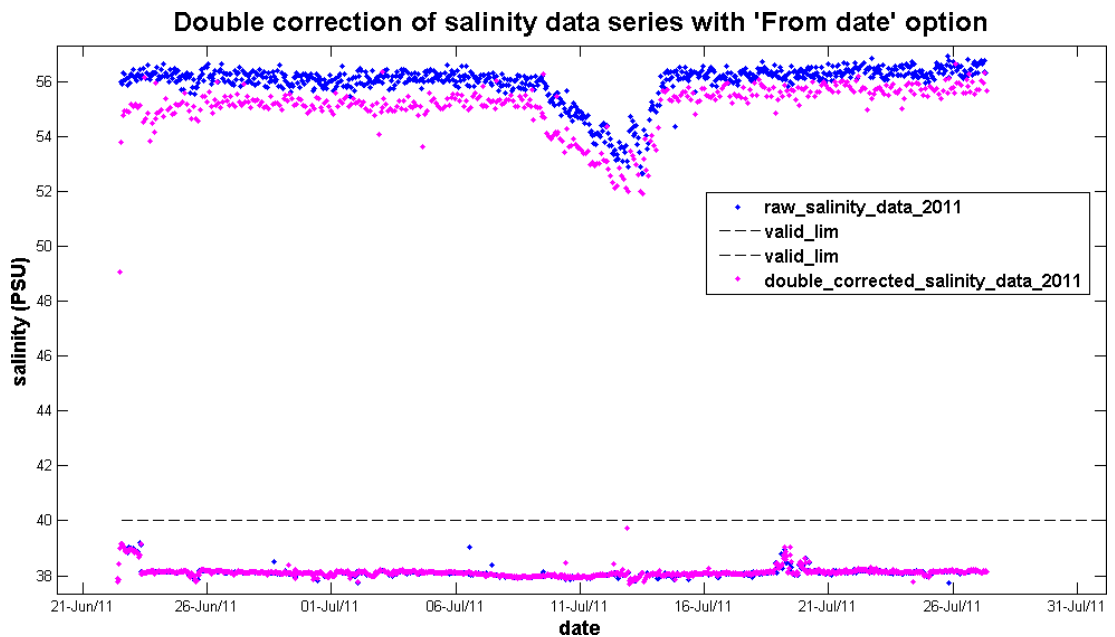


Figure 2-12: Example of double correction on salinity data series with the 'From date' option

- **'From data'** option: with this option, a supplementary condition is applied to the retained points, it must be in the valid bounds of the considered variable (Figure 2-13).

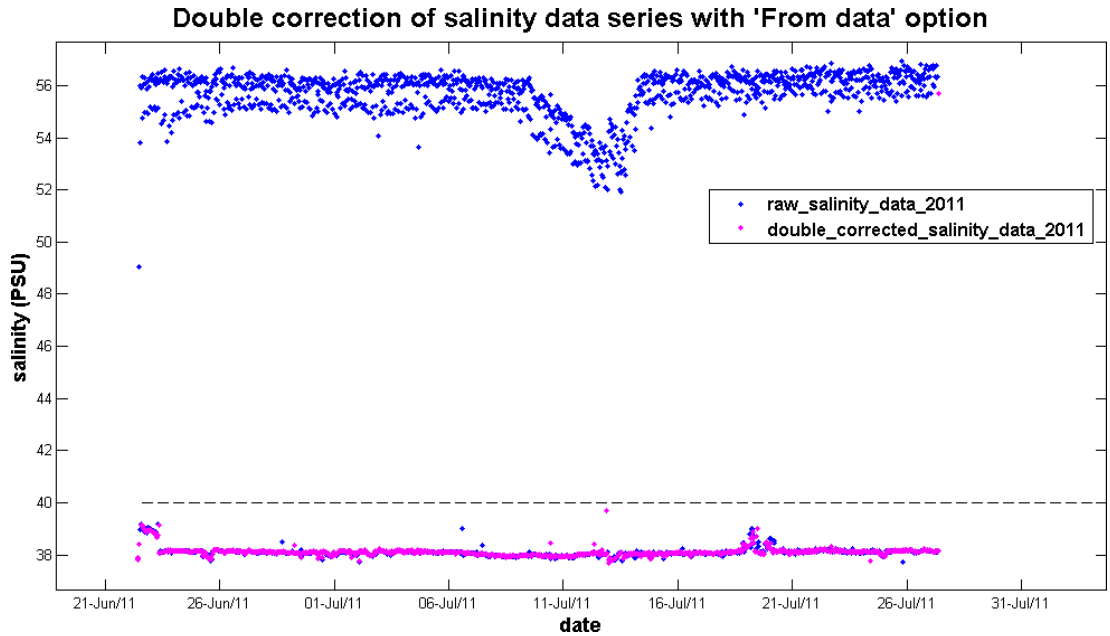


Figure 2-13: Example of double correction on salinity data series with the 'From data' option

2.4.3. Drift correction

For drift correction, the mean of maximum found on percentile 95 of each sub intervals is calculated on a reference data series. On the series that have to be corrected from drift signal (Figure 2-14), a linear regression is made on each subinterval and projected in order to have a horizontal line. The latter is then moved vertically depending on its maximum calculated on percentile 95, in the same manner of the bias correction method (Figure 2-15). When external in situ data are available, a comparison between corrected data series and in situ data can be done in order to evaluate correction relevance (Figure 2-16, Figure 2-17)

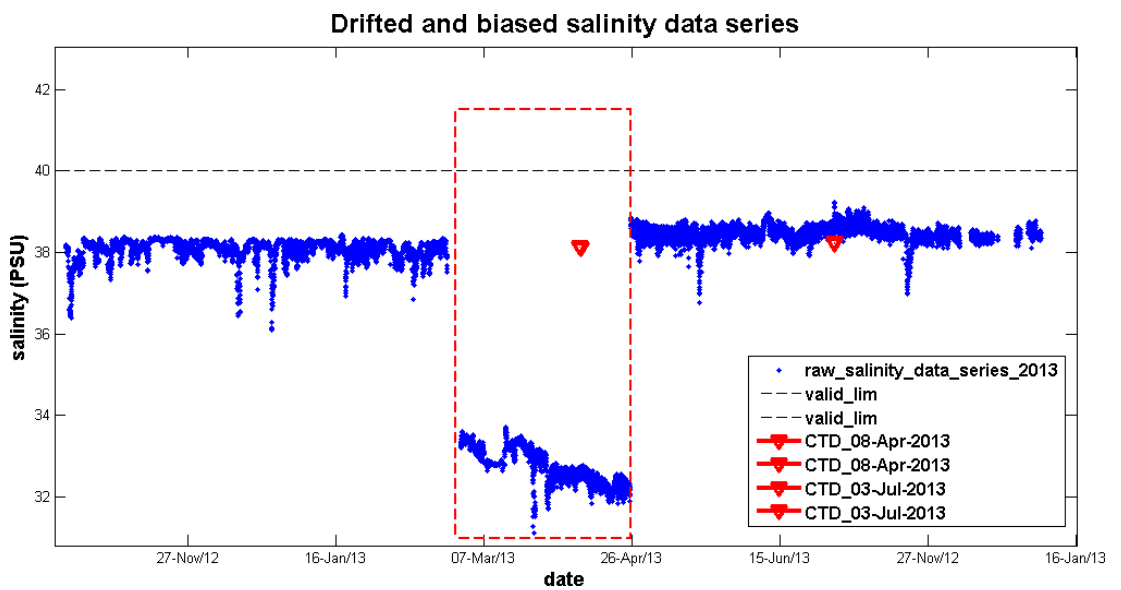


Figure 2-14: Salinity data series with drift signal

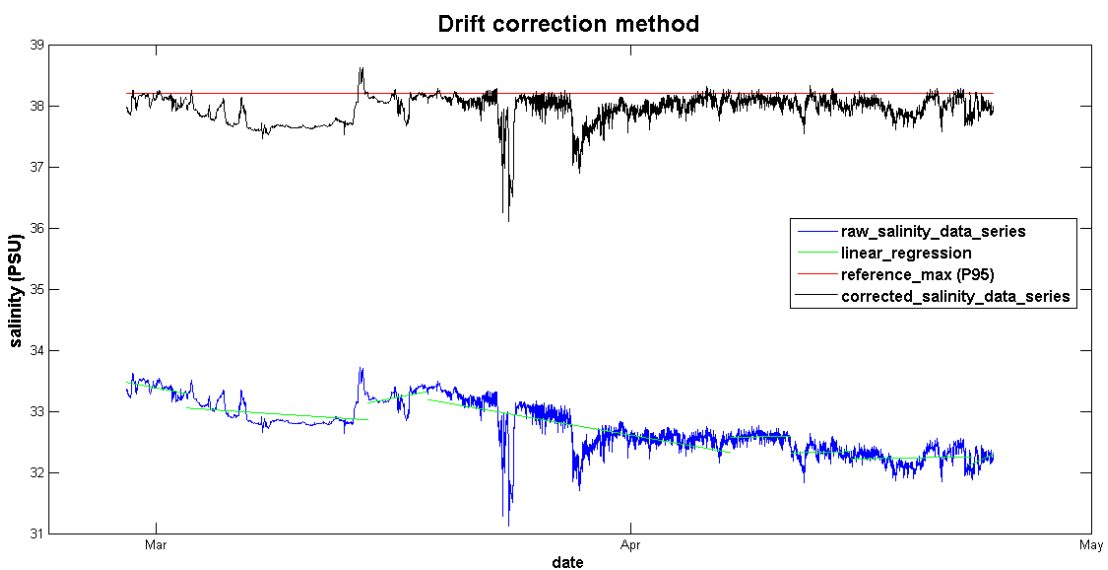


Figure 2-15: Drift correction method illustration

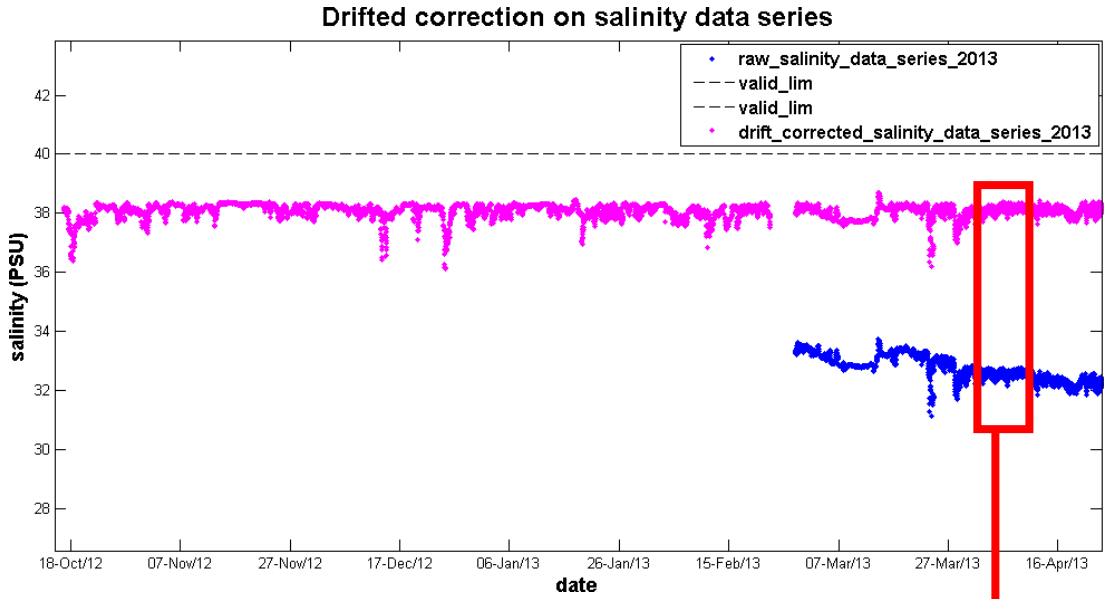


Figure 2-16: Salinity data series corrected from drift

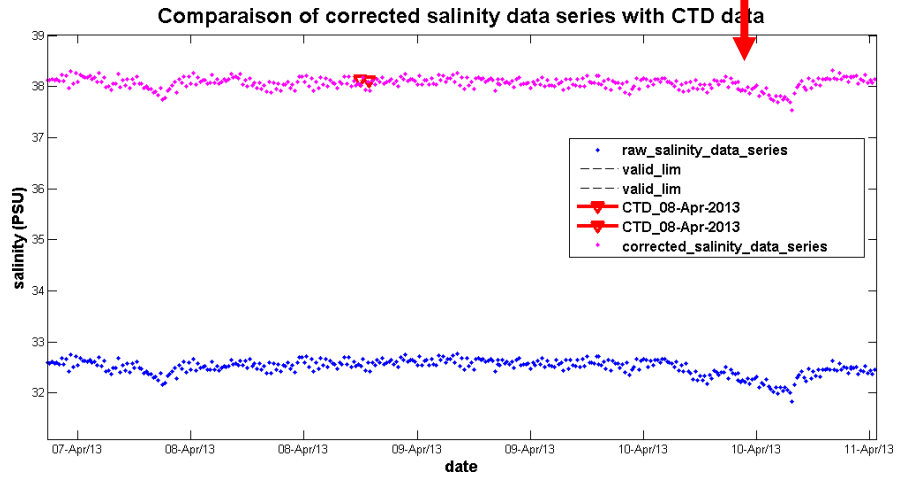


Figure 2-17: Comparison of in situ data from CTD with corrected salinity data series

3. Development of analysis tools

3.1. Matlab Buoy Interface (MBI) development

In order to make easier and faster data visualisation, treatment and analysis, a graphical user interface (GUI) was developed with the Matlab software: the Matlab Buoy Interface (Figure 3-1). From data formatted in the Matlab structure format thanks to dedicated routines, the MBI allows data visualisation, Quality Control (QC) modification, comparisons between SMATCH data and external in situ data (CTD). It gives the possibility to make corrections on data as mentioned above (for bias, drift and double troubles) and can do analysis thanks to several tools such as Fourier Series, Principal Components Analysis, linear regression and several filters, all described below.

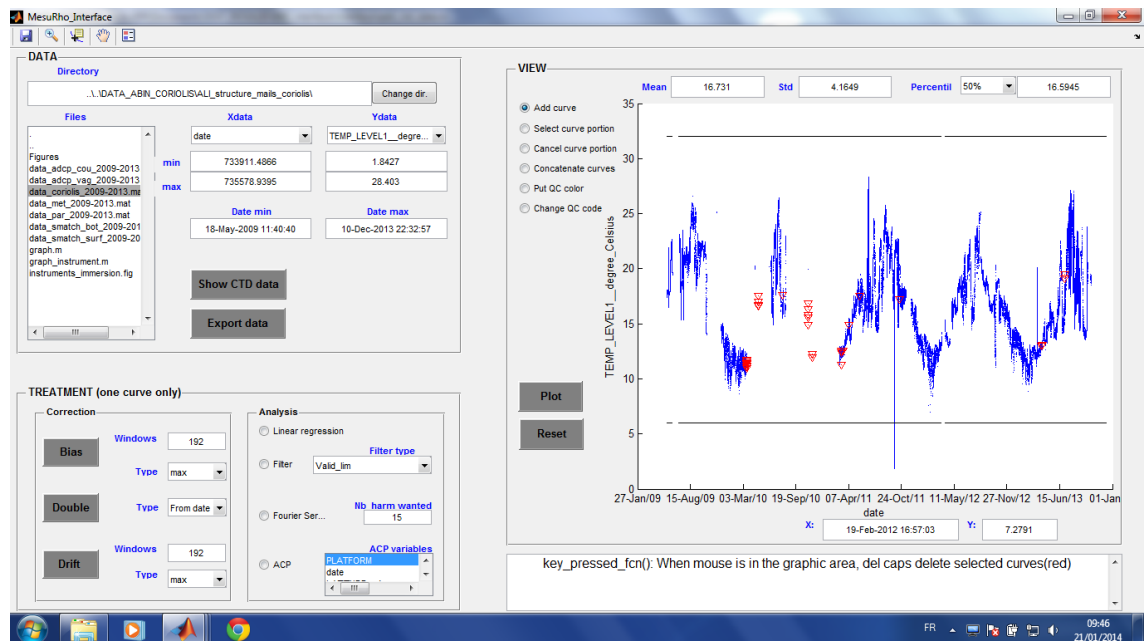


Figure 3-1: The Matlab Buoy Interface (MBI) showing surface temperature data compared with CTD data

3.2. Data analysis tools included in the interface

3.2.1. Linear regression

The linear regression option calculates a straight line that minimizes the quantity J (cost function, by the least squares method):

$$\hat{y}_i = \hat{a} \times data_{xi} + \hat{b} \quad (\text{Eq. 1})$$

$$J = \sum_{i=1}^n [data_{yi} - \hat{y}_i]^2 \quad (\text{Eq. 2})$$

where $data_{xi}$ and $data_{yi}$ are abscissa and ordinate of the data curve respectively, y_i the linear equation of the regression.

The solution of J minimization is

$$\hat{a} = \frac{cov(data_x, data_y)}{var(data_x)} \quad (\text{Eq. 3})$$

$$\hat{b} = \bar{y} - \hat{a} \times \bar{data}_x \quad (\text{Eq. 4})$$

Example:

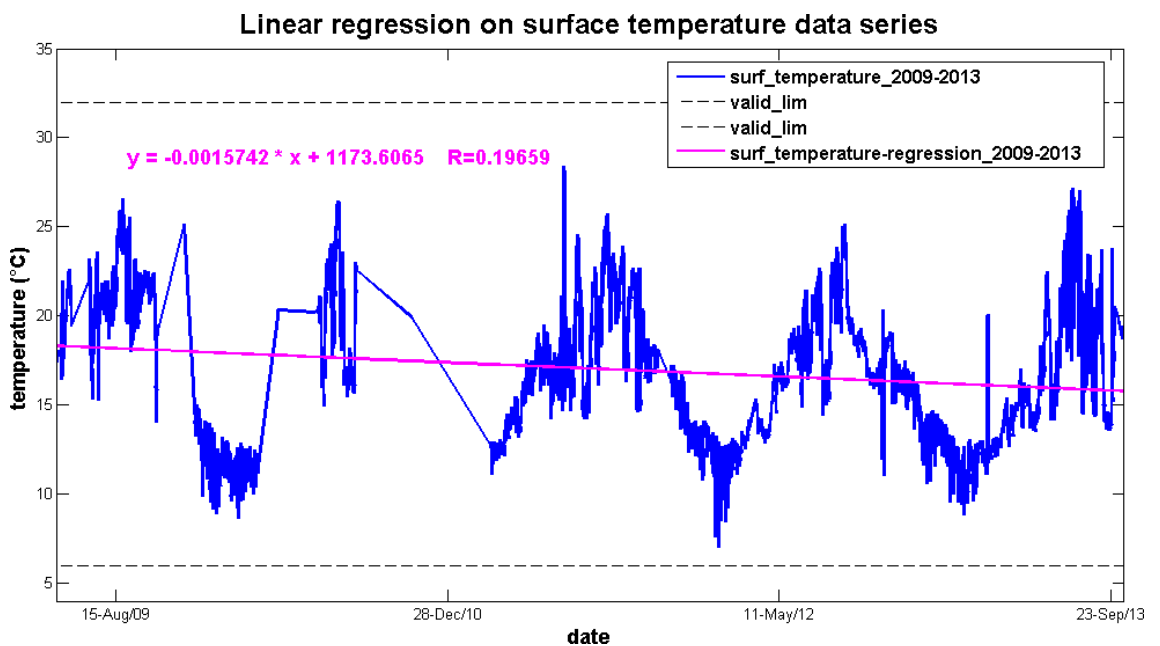


Figure 3-2: Example of linear regression on sub-surface temperature data series

3.2.2. Filters

The MBI is able to filter a temporal data series depending on user choice:

- 'Valid_lim': keeps data that are inside valid bounds
- 'QC': keeps data that have QC code wanted by the user (multiple possible choice)

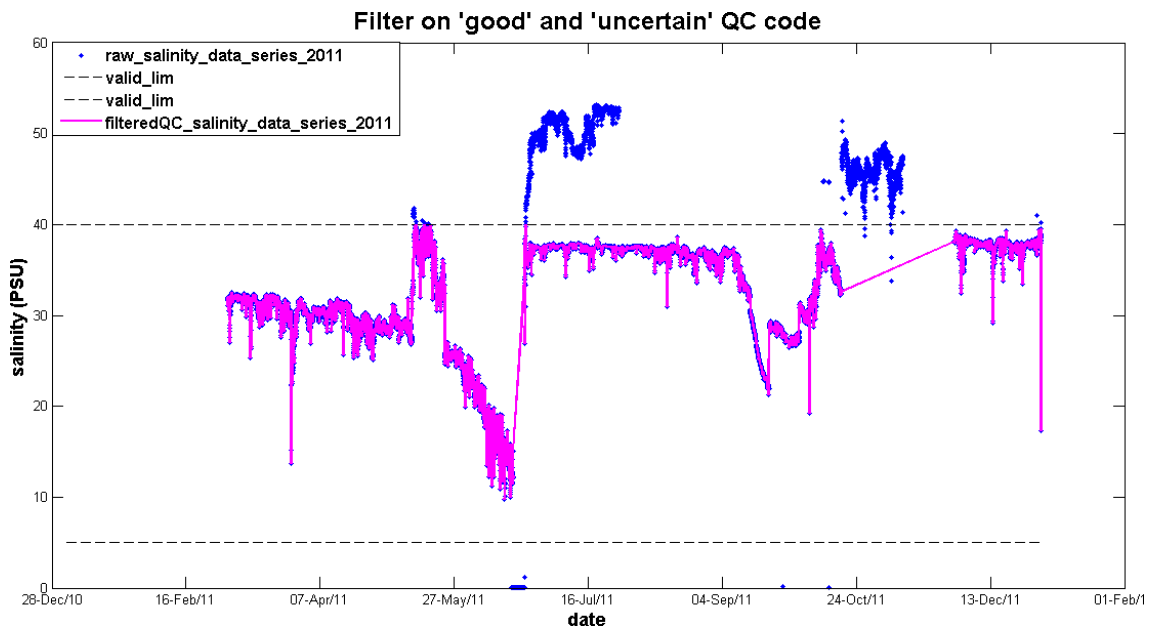
Example:

Figure 3-3: Example of QC filter on salinity data series

- **'Simple exponential'**: makes a simple exponential smoothing with user defined '*alpha*' smoothing parameter, following Eq. 5.

$$\hat{y}_t = \alpha \times y_t + (1 - \alpha) \times \hat{y}_{t-1} \quad (\text{Eq. 5})$$

- **'Double exponential'**: makes a double exponential smoothing (twice the simple exponential smoothing) with user defined '*alpha*' smoothing parameter

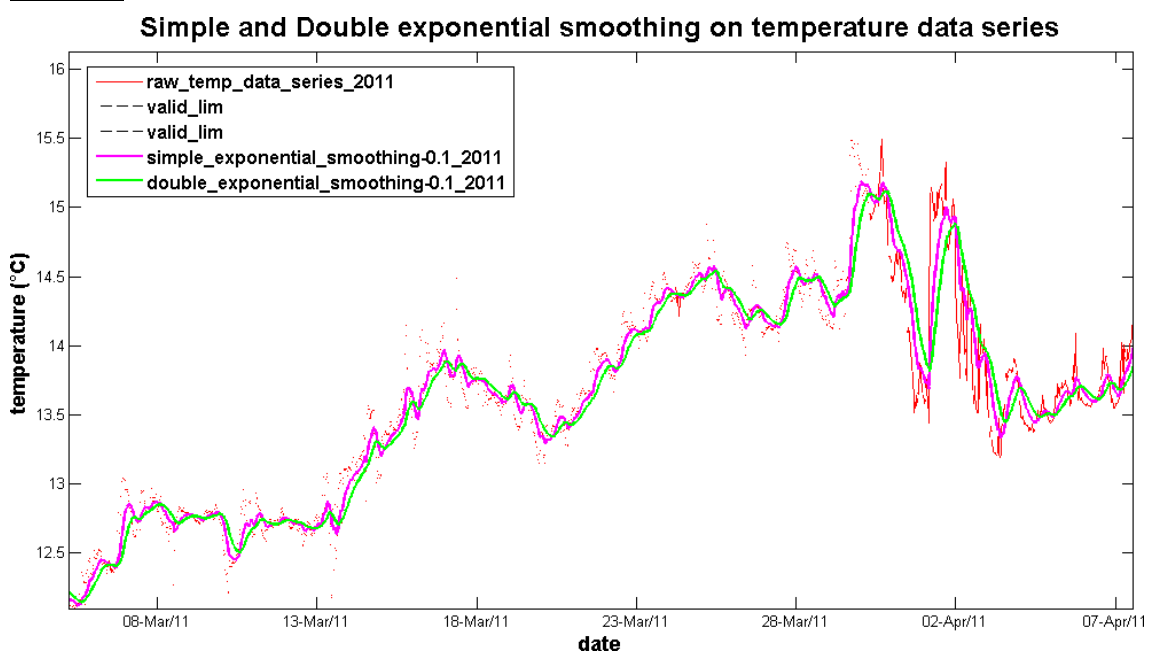
Example:

Figure 3-4: Example of simple and double exponential smoothing on temperature data

3.2.3. Fourier series (from Emery and Thomson (2004))

For many applications, we can view time series as linear combination of periodic or quasi periodic components that are superimposed on a long-term trend and random high frequency noise. The periodic components are assumed to have fixed, or slowly varying amplitude and phases over the length of the record. The trends might include a slow drift in the sensors characteristics or a long term component of variability that can not be resolved by the data series. "Noise" includes random contributions from the instrument sensors or electronics, as well as frequency components that are outside the immediate range of interest (e.g. small scale turbulence). A goal of time-series analysis in the frequency domain is to reliably separate periodic oscillations from the random and aperiodic fluctuations.

Fourier analysis is one of the most commonly used methods for identifying periodic components in near stationary time-series oceanographic data (if the time series are strongly non-stationary, more localized transforms such as Hilbert and Wavelet transforms should be used, as well as Empirical Mode Decomposition).

Fourier's basic premise was that any finite length, infinitely repeated time series, $y(t)$, defined over principal interval $[0, T]$ can be reproduced using a linear summation of cosines and sines of the transform.

The Fourier series is calculated as:

$$y(t_n) = \frac{1}{2} A_0 + \sum_{p=1}^{N/2} [A_p \cdot \cos(\omega_p \cdot t_n) + B_p \cdot \sin(\omega_p \cdot t_n)] \quad (\text{Eq. 6})$$

$$\text{where } \frac{1}{2} A_0 = \bar{y}, B_0 = 0$$

where \bar{y} is the mean of the record, A_p and B_p are constants (Fourier coefficients), and the specified angular frequencies ω_p are integer ($p=1, 2, 3, \dots$) multiples of the fundamental frequency, $\omega_1 = 2\pi \cdot f_1 = 2\pi/T$, where T is the total length of the series.

Using $t_n = n \cdot \Delta t$, the final form for the discrete, finite Fourier series becomes:

$$y(t_n) = \frac{1}{2} A_0 + \sum_{p=1}^{N/2} [A_p \cdot \cos(2 \cdot \pi \cdot p \cdot n / N) + B_p \cdot \sin(2 \cdot \pi \cdot p \cdot n / N)] \quad (\text{Eq. 7})$$

with :

$$A_p = \frac{2}{N} \sum_{n=1}^N y_n \cdot \cos(2 \cdot \pi \cdot p \cdot n / N), \quad p=0, 1, 2, \dots, N/2 \quad (\text{Eq. 8})$$

$$B_p = \frac{2}{N} \sum_{n=1}^N y_n \cdot \sin(2 \cdot \pi \cdot p \cdot n / N), \quad p=1, 2, \dots, (N/2)-1$$

We can also express Fourier series as amplitude and phase (Eq. 9). C_p is the amplitude of the p th component, Θ_p the phase angle of the constituent. The phase angle gives the relative 'lag' of the component in radians measured counter clockwise from the real axis ($A_p > 0, B_p = 0$). The corresponding time lag for the p th component is then $t_p = \Theta_p / 2\pi \cdot f_p$, in which Θ_p is measured in radians so that:

$$y(t_n) = \frac{1}{2} C_0 + \sum_{p=1}^{N/2} [C_p \cdot \cos(2 \cdot \pi \cdot p \cdot n / N) - \theta_p] \quad (\text{Eq. 9})$$

$$C_0 = A_0$$

$$C_p = (A_p^2 + B_p^2)^{1/2}, p=0,1,2,\dots$$

$$\theta_p = \tan^{-1}[B_p / A_p], p=1,2,\dots$$

The relative contribution a given component makes to the total variance of the time series is a measure of the importance of that particular frequency component in the observed signal.

Example:

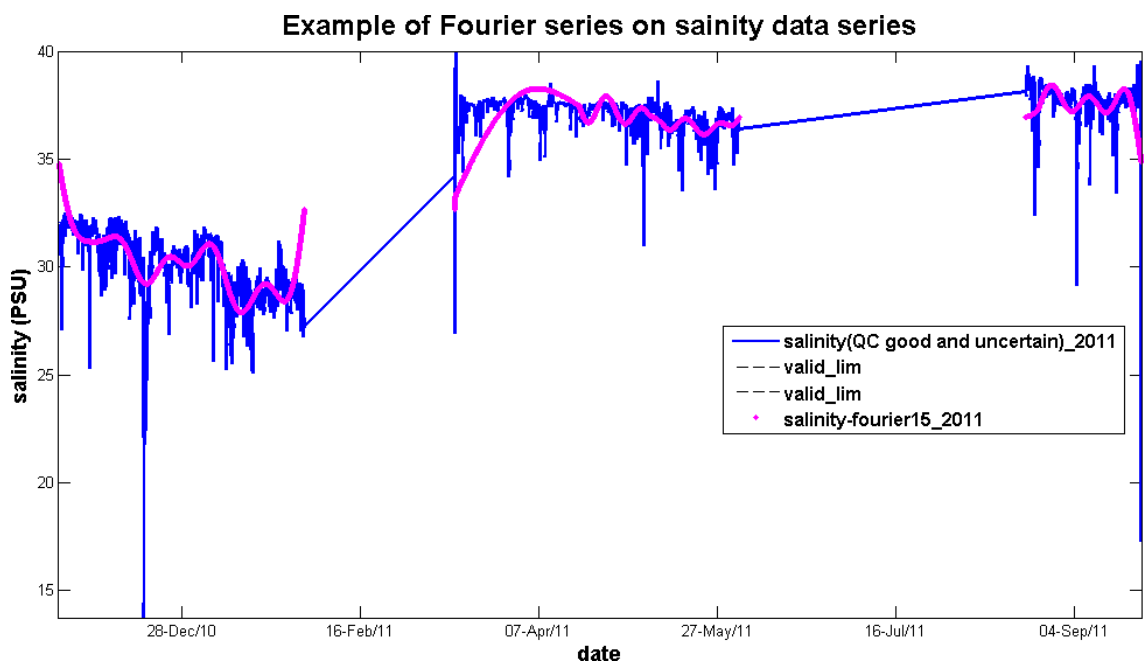


Figure 3-5: Example of Fourier series on salinity data, with the 15th first components

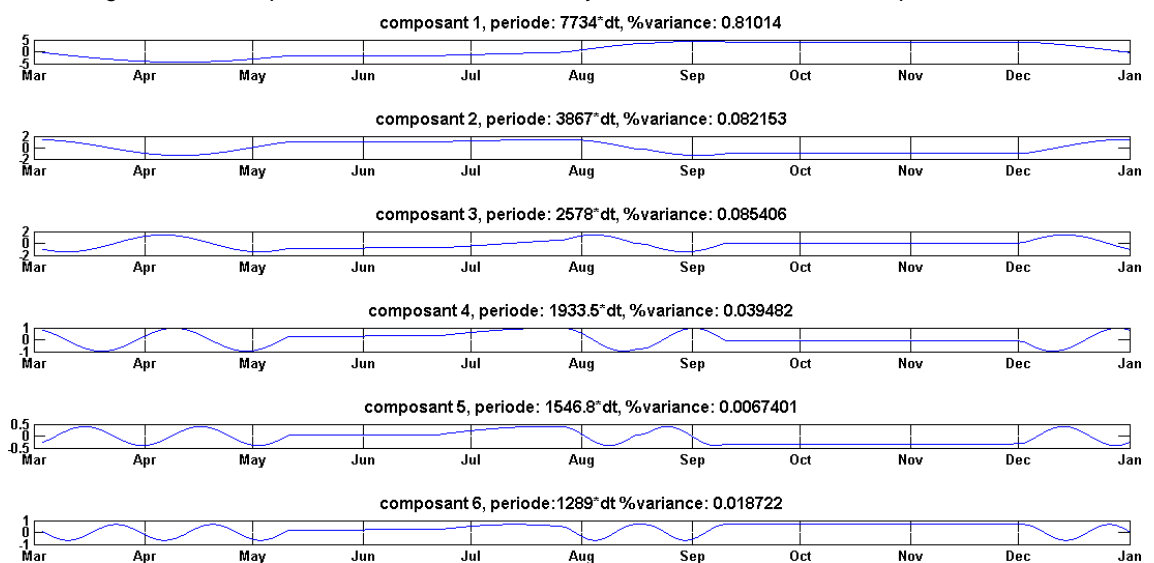


Figure 3-6: Details on the 6th first components of the Fourier series

3.2.4. Principal Component Analysis

Principal component analysis (PCA) is a statistical procedure that uses orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. The number of principal components is less than or equal to the number of original variables. This transformation is defined in such a way that the first principal component has the largest possible variance (that is, accounts for as much of the variability in the data as possible), and each succeeding component in turn has the highest variance possible under the constraint that it must be orthogonal to (i.e., uncorrelated with) the preceding components. Principal components are guaranteed to be independent if the data set is jointly normally distributed. PCA is sensitive to the relative scaling of the original variables.

Methods details from <http://maths.cnam.fr/IMG/pdf/A-C-P-.pdf>:

From a set of p variables observed n times, a data matrix X is constructed (Figure 3-7)

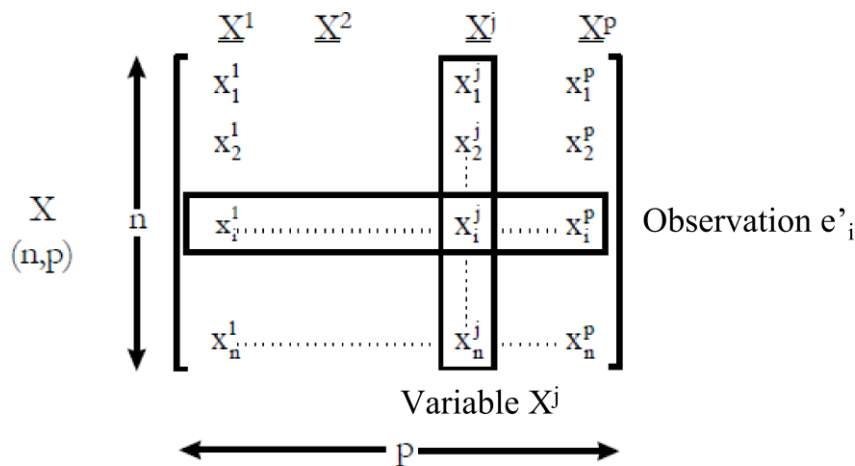


Figure 3-7: ACP data matrix

The data matrix X is then centred (here after annotated 1) or centred and reduced (2) leading to the M matrix, and the variance-covariance (1) or correlation (2) matrix V is calculated as:

$weight = \frac{1}{n} I, \text{ where } I \text{ is the identity}(n, p)$ $V = M' \cdot weight \cdot M$	(Eq. 10)
--	----------

If variables are centred and reduced, diagonal elements of V represent the variables variance and is equal to 1. The trace of V , which represents the total inertia, is equal to p , the number of variables.

The research of axes with the maximum inertia leads to the construction of new variables (which are associated with these axes) of maximum variance. In other words, a change of reference is performed in \mathcal{R}^p in order to be placed in a new

system representation where the first axis provides the maximum total inertia of the data, the second axis the maximum of inertia not taken into account by the first axis, and thereby away.

This reorganization is based on the diagonalization of the variance-covariance (1) or correlation (2) matrix:

$$V = UDU^{-1} \quad (\text{Eq. 11})$$

where D is diagonal, U composed of eigen vectors

Principal components C are then calculated as:

$$\underline{C}^i = U_1^i \cdot \underline{X}^1 + U_p^i \cdot \underline{X}^2 + \dots + U_p^i \cdot \underline{X}^p \quad (\text{Eq. 12})$$

The i^{th} principal component C^i provides coordinates of the n observations on the i^{th} principal axis.

Example:

An ACP was performed on sub surface SMATCH 2011 data set.

Figure 3-8 shows eigenvalues of the diagonal matrix, and cumulative percentage of the total data inertia supported by the axes related to these eigenvalues.

Figure 3-9 shows the representation of data and variables on the two principal axes of the new reference.

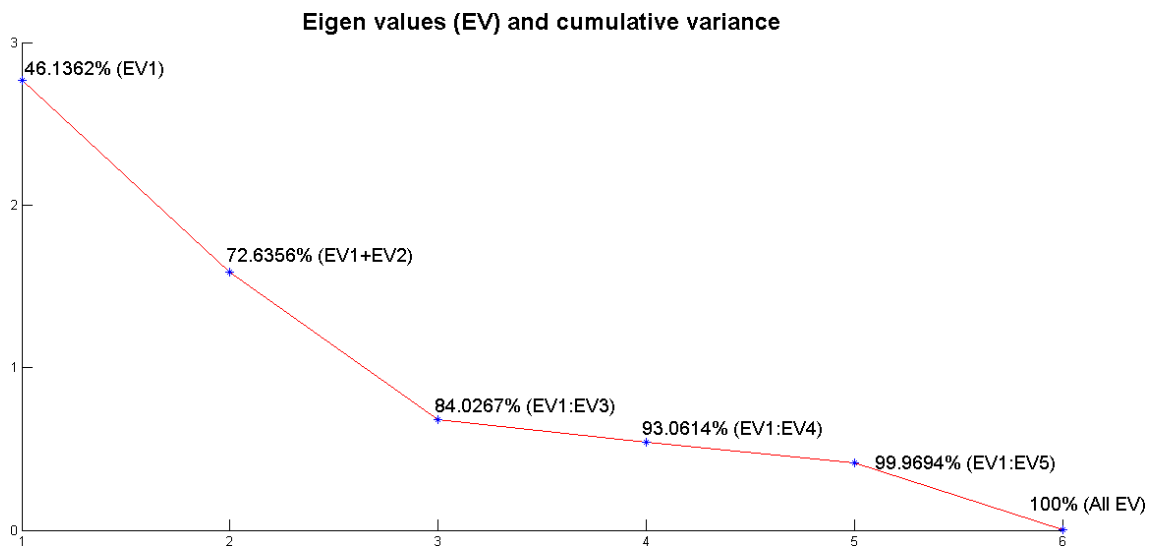


Figure 3-8: Eigenvalues of the diagonal matrix and cumulative explained variance

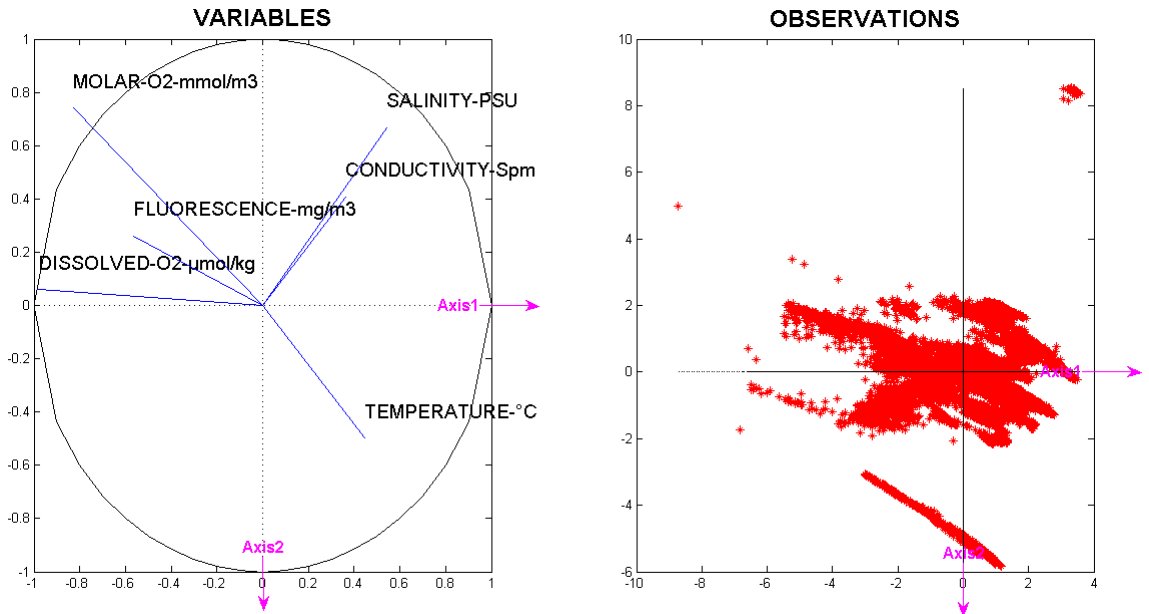


Figure 3-9: Variables and observations representation in the reference defined by the two first principal components/axis

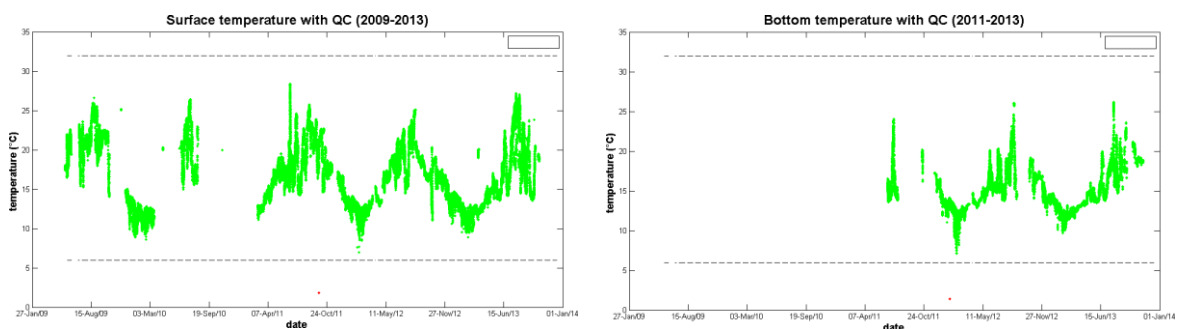
4. Data treatment (from SMATCH multi-parameters sensors)

Here, SMATCH probes data are presented with associated quality code. Some data were corrected on salinity, fluorescence and turbidity variables which shown drift or bias signs. Their quality code were put as 'modified'. When data were suspicious and no chlorination occurred, data were not corrected and put as 'uncertain' for their quality code. Indeed, no chlorination process means that sensors were not cleaned, so we don't really know what was measured (ex: bio fouling presence).



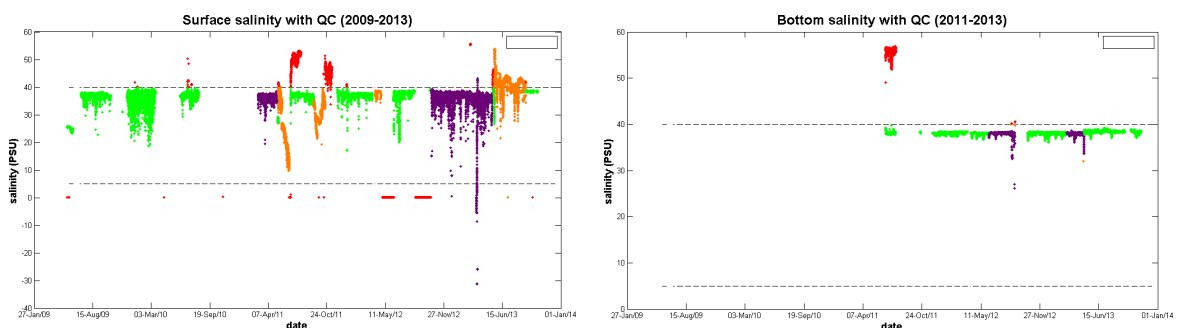
4.1. Temperature

Figure 4-1: Surface and bottom temperature data with control quality code colours



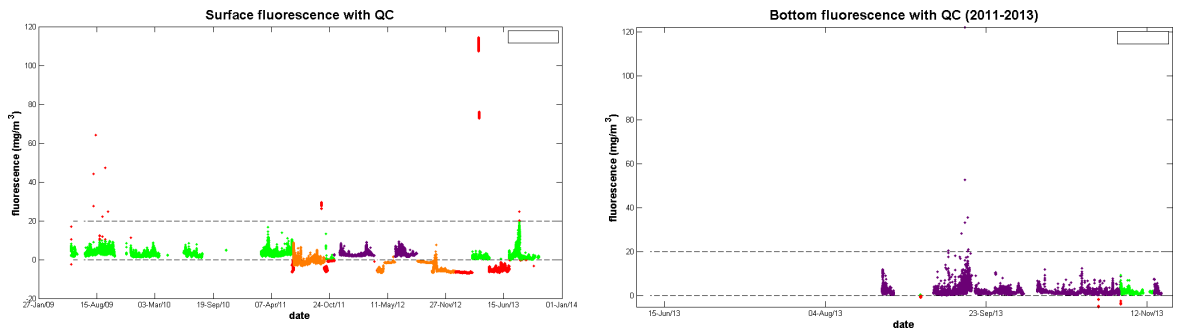
4.2. Salinity

Figure 4-2: Surface and bottom salinity data with control quality code colours



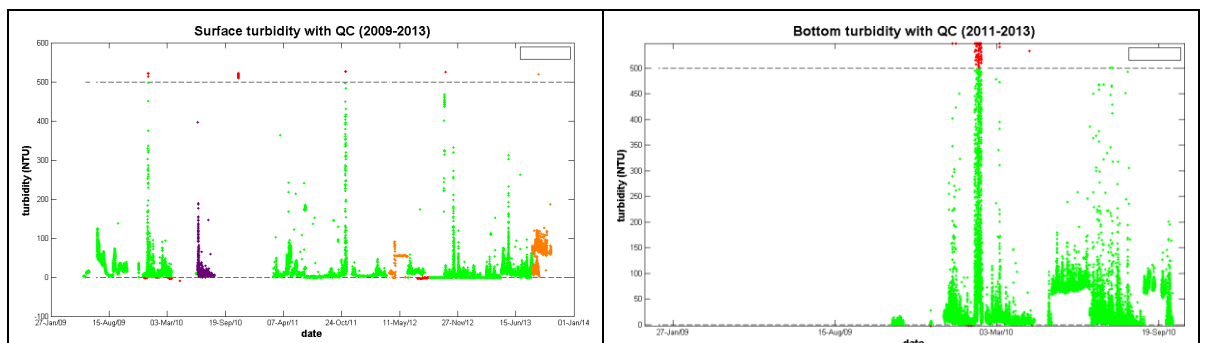
4.3. Fluorescence (chlorophyll-a or phycocyanine)

Figure 4-3: Surface and bottom fluorescence data with control quality code colours



4.4. Turbidity

Figure 4-4: Surface and bottom turbidity data with control quality code colours



5. Example of study : storm event (5-7 March 2013)

The weather station data (Figure 5-1 shows a strong east/south-east (90-110°) wind event that lasted four days, from the 4th of March to the 7th of March 2013. Maximum wind speed reached 20m/s (72 km/h, equivalent to 39 knots). This event was associated with rain (maximum rainfall of 1.54 mm/day on the 6th of March), increasing continuously the Rhone river flow during the storm period (Figure 5-1: Rhone river flow and SMATCH probe salinity data (a), weather station data (b, c) during the storm event). The sub-surface SMATCH probe was situated at 2.7 m +/- 0.5 under the sea surface, with high waves (maximum of 4m as shown in Figure 5-2 (a)) and sea surface elevation.

Strong freshening signal was present only on the 6th of March (Figure 5-1 (a)), while the Rhone River flow increased during four days and wind conditions were stable. Several assumptions could be made to explain this phenomena:

- First, because of sea level variations and high waves, the SMATCH probe was at a variable distance from the sea surface, 2.7m +/- 0.5m during the storm event, which is quite deep to feel the influence of the Rhone River plume.
- The rain doesn't seem to impact salinity at the SMATCH depth as it concerns mainly a few centimetres of the sea surface.
- Figure 5-2 (b) shows strong turbidity on the 6th of March, concomitant to the freshening signal. These facts could lead to the assumption that Rhone River slightly diluted waters may have reached the MesuRho station and could accumulate in the coastal zone due to hydrodynamics under north-west currents, leading to the presence of fresh water at a deeper depth.

In order to better understand the freshening event of the 6th of March 2013, supplementary information and analysis are needed, especially to put measurements in spatial and hydrodynamic contexts. Unfortunately, south east wind brought clouds in the area of interest and no satellite data were available between the 4th and the 7th of March.

Figure 5-3 shows surface chlorophyll satellite data for the 3rd, 8th, 9th and 10th of March 2013. The satellite data resolution as well as clouds presence do not allow to conclude about the freshening event measured by the sub surface SMATCH probe on the 6th of March.

Hydrodynamic models would potentially give precious information about this freshening event:

- where does it come from?
- How deep is the fresh water mass?
- What are the process(es)/process succession or interactions responsible for the fresh water mass reaching the sub surface SMATCH depth at the MesuRho station?

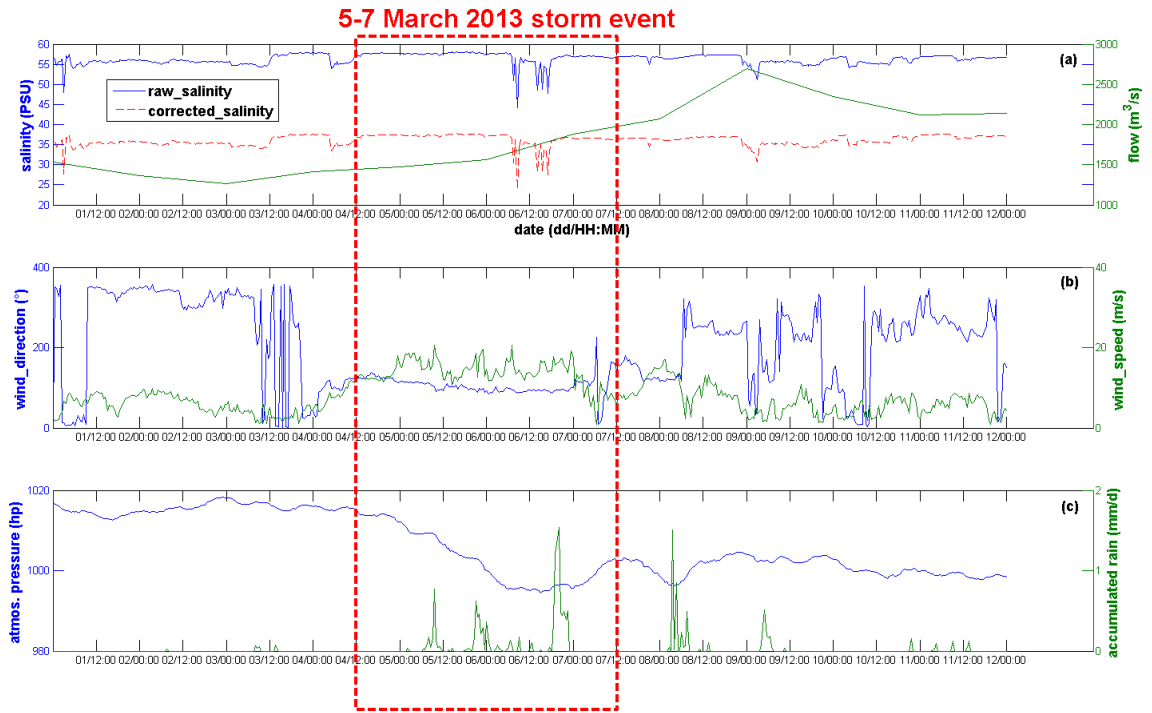


Figure 5-1: Rhone river flow and SMATCH probe salinity data (a), weather station data (b, c) during the storm event

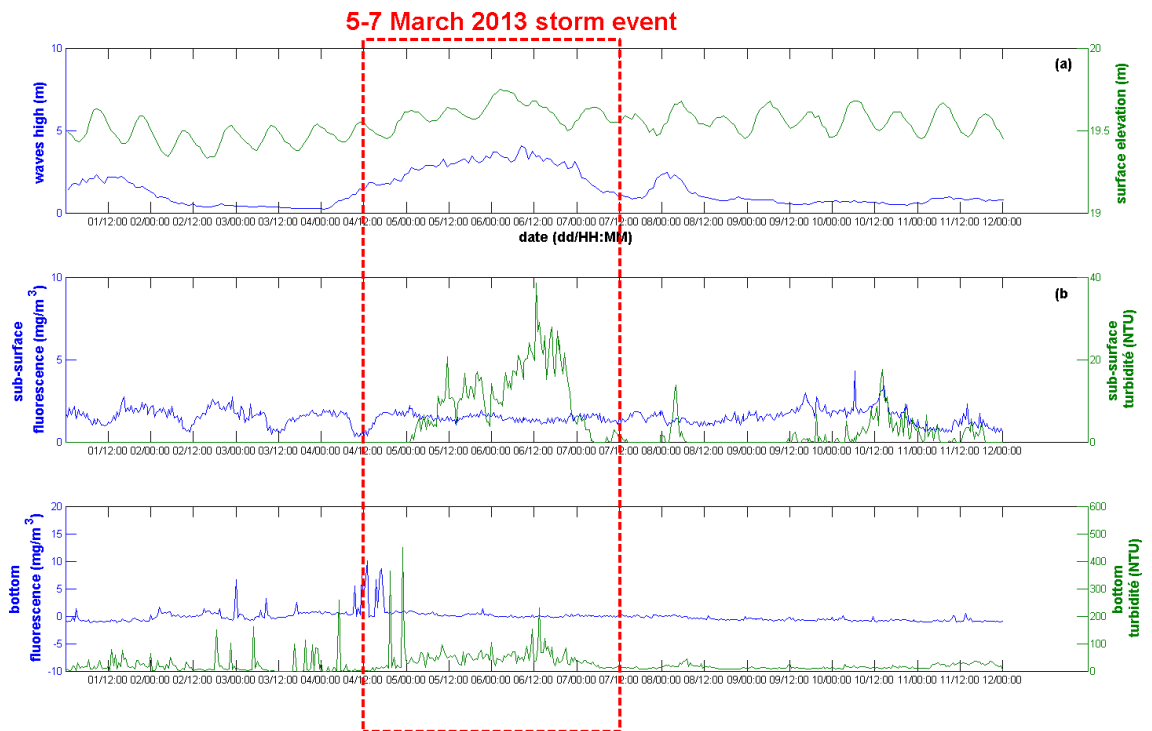


Figure 5-2: Wave heights and surface elevation derived from ADCP data (a), SMATCH probe fluorescence and turbidity data at the sub-surface (b) and at the bottom (c)

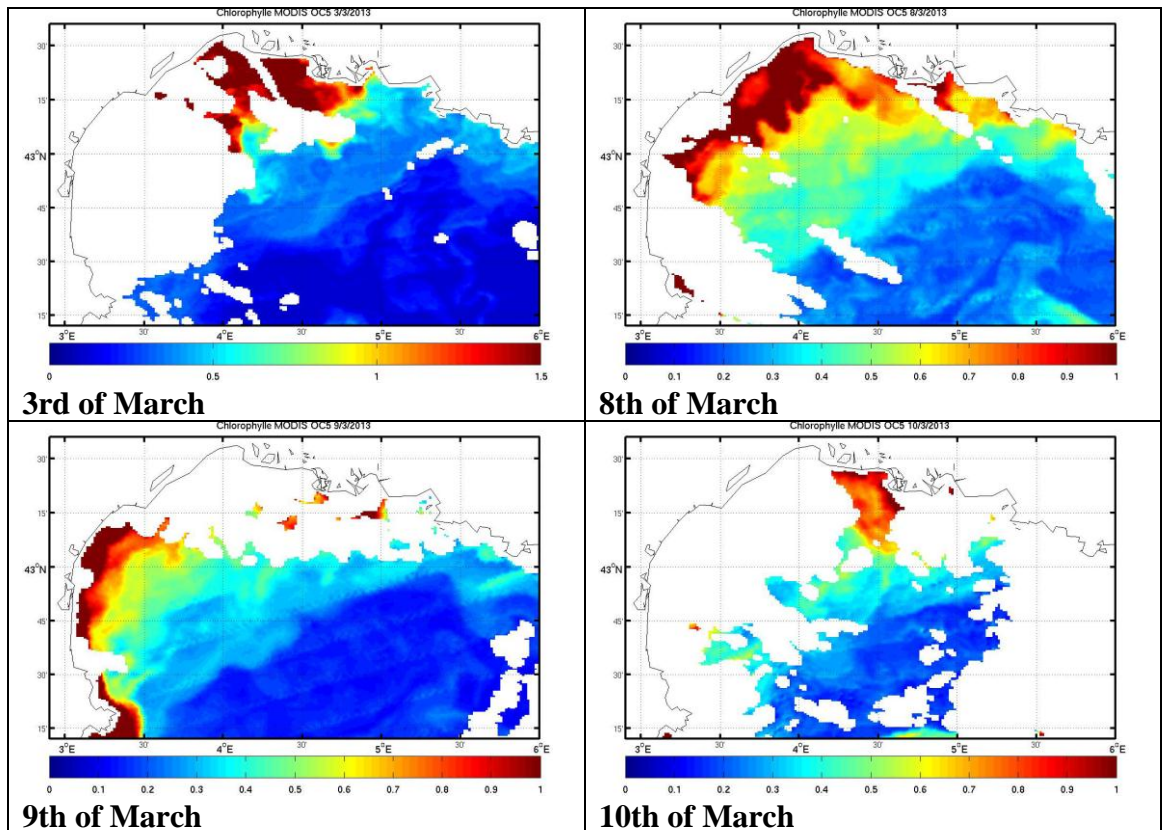


Figure 5-3: MODIS Chlorophyll satellite data before and after the storm event

6. Future improvements

In order to better manage SMATCH probes and resulting measurements, several possible improvements are proposed:

- To integrate sensor id number in the Coriolis web data base. As five SMATCH probes are used in turn on the MesuRho station, the id number record in the web data base would allow to check if observed problems come from a particular probe.
- To look for SMATCH probes sensor calibration in the field, in addition to metrology (laboratory sensor calibration), and for all the sensors, just after the probe deployment and before its recovery, to check for drift signal and magnitude order. This would help for correction tasks and also for optical devices calibration.
- To maintain the list of all interventions on the MesuRho station in a single document/book/online base, allowing a global view, information overlap and monitoring improvements. This can be helpful also for new participants and time saving in many tasks.
- Ideal planning of operations to monitor sensor fluctuations (drift, bias...) would be monthly in situ measurements, as close as possible of the mooring, with external instrument such as a CTD equipped with a fluorometer and an OBS, and water

samples for fluorescence/chlorophyll regression as well as turbidity comparisons, as already done in April and July 2013.

- The MBI interface can be enriched with other analysis methods such as the Hilbert and Wavelet transform and Empirical Mode Decomposition for strongly non stationary time series. Also neural network methods would be a possible candidate, when large data set would be available, for blank filling, water mass discrimination...etc.
- When large data set will be available, corrections on variables which depend on season have to be made based on reference data on same period. Indeed this would allow to provide uncertainties bounds on the corrected data.

References

Aminot, A. and K erouel, R. (2004). Hydrologie des  cosyst mes marins: param tres et analyses. Edition de l'IFREMER

Lachaise, J. (2013). Traitement et analyse des donn es de la station temps-r el MesuRho. Rapport de stage Master professionnel "Sciences et Technologies".

Lorthiois, T. (2012). Dynamique des mati res en suspension dans le panache du Rh ne (m diterran e occidentale) par t l d tection spatiale couleur de l'oc an. PhD thesis, Universit  Pierre et Marie Curie- Paris VI .

Moutin, T., P. Raimbault, H. G., and Coste, B. (1998). The input of nutrients by the Rh ne river into the Mediterranean sea : recent observations and comparaison with earlier data. *Hydrobiologia*, (373/374).

Pinazo, C., Fraysse, M., Doglioli, A., Faure, V. M., Paireaud, I., Petrenko, A., Thouvenin, B., Tronczynski, J., Verney, R., and Yohia, C. (2013). Massilia : Mod lisation de la baie de Marseille : Influence des apports anthropiques de la m tropole sur l' cosyst me marin. Rapport scientifique Ifremer 2013 RST.ODE/LER/PAC/13-14 .

Emery, W.J. and Thomson, R.E. (2004). *Data Analysis Methods in Physical Oceanography*. Elsevier B.V.

Appendix

-Appendix 1: Available variables and names in the Coriolis web data base

Level 0 are atmospheric data, level 1 data belong to sub-surface SMATCH probe, level 3 to bottom SMATCH probe. For ADCP data level are from 0 to 40.

CORIOLIS VARIABLES NAME	VARIABLES
PLATFORM	station number/id
LATITUDE__degree_north	latitude
LONGITUDE__degree_east	longitude
DRYT_LEVEL0__degree_Celsius	Atmospheric temperature
ATMS_LEVEL0__hectopascal	Atmospheric pressure
DEPH_LEVEL0__meter	ADCP depth
GSPD_LEVEL0__meterpsecond	ADCP speed
HEAD_LEVEL0__degree	ADCP direction
DATE	date
PRRD_LEVEL0__millimeterpday	Rain amount
RELH_LEVEL0__p	
SLEV_LEVEL0__meter	Sea surface elevation
VAVH_LEVEL0__meter	Wave high
WDIR_LEVEL0__degree	Wind direction
WSPD_LEVEL0__meterpsecond	Wind speed
DEPH_LEVEL1__meter	ADCP
HCDT_LEVEL1__degree	ADCP
HCSP_LEVEL1__meterpsecond	ADCP
LGH4_LEVEL1__micromole_photonp_m2bs	PAR
SMATCH_AUTOMATE_BATTERY_LEVEL1__volt	SMATCH battery voltage
TEMP_LEVEL1__degree_Celsius	Subsurface SMATCH temperature
PSAL_LEVEL1__psu	Subsurface SMATCH salinity

CNDC_LEVEL1__Spm	Subsurface conductivity	SMATCH
DOX2_LEVEL1__micromolepkg	Subsurface dissolved oxygen	SMATCH
FLU2_LEVEL1__milligrampm3	Subsurface fluorescence	SMATCH
MOLAR_DOXY_LEVEL3__millimolepm3	Subsurface molar oxygen	SMATCH
TUR4_LEVEL1__ntu	Subsurface turbidity	SMATCH

-Appendix 2: Salinity calculation

Salinity is calculated following Aminot and K erouel (2004) formula. It depends on conductivity, temperature and pressure as follows:

$$S = a_0 + a_1 \cdot R_t^{0.5} + a_2 \cdot R_t + a_3 \cdot R_t^{1.5} + a_4 \cdot R_t^2 + a_5 \cdot R_t^{2.5} + \left\{ \frac{(t-15)}{[1+k(t-15)]} \right\} * (b_0 + b_1 \cdot R_t^{0.5} + b_2 \cdot R_t + b_3 \cdot R_t^{1.5} + b_4 \cdot R_t^2 + b_5 \cdot R_t^{2.5})$$

$$\text{where } R_t = \frac{R}{(R_p \cdot r_t)} = \frac{C_{S,t,p}}{(42,914 \cdot r_t)}$$

$$\text{With: } R = \frac{C_{S,t,p}}{42,914} ; R_p = 1 \text{ (for small depth) and } r_t = c_0 + c_1 \cdot t + c_2 \cdot t^2 + c_3 \cdot t^3 + c_4 \cdot t^4$$

where $C = \text{conductivity} (S \cdot m^{-1})$

$t = \text{temperature} (C^\circ)$

$p = \text{pressure}(\text{bar})$

a_i, b_i, c_i et k are chosen such as:

$a_0 = 0,0080$	$b_0 = 0,0005$	$c_0 = 0,6766097$
$a_1 = -0,1692$	$b_1 = -0,0056$	$c_1 = 0,0200564$
$a_2 = 25,3851$	$b_2 = -0,0066$	$c_2 = 0,00011043$
$a_3 = 14,0941$	$b_3 = -0,0375$	$c_3 = -6,9698E-07$
$a_4 = -7,0261$	$b_4 = 0,0636$	$c_4 = 1,0031E-09$
$a_5 = 2,7081$	$b_5 = -0,0144$	$k = 0,0162$
$\Sigma a_i = 35,0000$	$\Sigma b_i = 0,0000$	

- Appendix 3: Dissolved oxygen data treatment

Dissolved oxygen is measured in mg/l and needs conversion in $\mu\text{mol/l}$ as well as salinity and pressure effects correction. The CDOCO applies the following procedure for dissolved oxygen data treatment.

- Measure conversion from mg/l (O_{2_1}) to $\mu\text{mol/l}$ (O_{2_2}) thanks to the coefficient 31,251172 resulting from the oxygen molar mass.

$$O_{2_2} = O_{2_1} \times 31,251175$$

- Salinity (O_{2_3}) and pressure (O_{2_4}) effects correction depending on provided formula by sensor builder. The salinity correction formula comes from Garcia and Gordon (1992).

$$O_{2_3} = O_{2_2} \times e^{\left(S \times \left(b_0 + b_1 \times T_s^{\circ 2} + b_3 \times T_s^{\circ 3} \right) + C_0 \times S^2 \right)}$$

with :

$$b_0 = -6.24097 \times 10^{-3}$$

$$b_1 = -6.93498 \times 10^{-3}$$

$$b_2 = -6.90358 \times 10^{-3}$$

$$b_3 = -4.29155 \times 10^{-3}$$

$$c_0 = -3.11680 \times 10^{-7}$$

$$T_s^{\circ} = \ln((298.15 - t)/(273.15 + t))$$

$$O_{2_4} = O_{2_3} \times \left(\frac{C_{pp} \times p}{1000} \right)$$

where S = Salinity

t = temperature (C°)

p = depth (m)

$C_{pp} = 0,032$ (compensation.coeff (Uchida et al.,2008))

Finally, the concentration expressed in volumetric unit ($\mu\text{mol/l}$) is converted in gravimetric unit ($\mu\text{mol/kg}$) (O_{2-5}) thanks to potential density calculation:

$$O_{2-5} = \frac{O_{2-4}}{dp}$$

where dp is the potential density of water