**SubOcean data processing and visualization : proposed action plan for Hugo**

Jérôme, Sébastien, 3 December 2024.

*NB : We’ve highlighted in red specific aspects requiring clarification with A2PS, clarification which could be handled by Hugo during his training on the software by A2PS.*

**Context**: the SENSE unit owns two SubOcean instruments built by the small French company A2 Photonic Sensors (A2PS) in Grenoble. One of them provides in-situ and real-time measurements of dissolved CH4, dissolved C2H6 and the stable carbon isotopic ratio of CH4 (usually called “SubOcean-CH4”). The second one measures dissolved CH4, dissolved N2O and NH3 (usually called “SubOcean-N2O”). A dedicated software has been developed by the A2PS company, allowing to:

1. Communicate with the instrument in the field,
2. Visualize the instrument data in real-time,
3. Export files,
4. Reprocess files after acquisition.

The latter option allows in particular to apply corrections of the measurements, using a separate dataset including temperature, salinity and dissolved oxygen measurements obtained with a CTD. The three variables are necessary to obtain fully-corrected SubOcean data.

The SENSE unit has the following needs for best use of SubOcean data:

1. Real-time visualization,
2. Offline data treatment and visualization,
3. Offline visualization with other datasets, including with 2D or 3D maps.

Category 1 should be taken in charge by A2PS for improvement of the existing software. Categories 2 and 3 would benefit from custom tools adapted or developed by Hugo.

**SubOcean data**:

ASCII files including:

*Common to the two instruments* :

* Date and time (if I recall well, time is systematically UTC).
* Depth and hydrostatic pressure: The depth is deduced from the hydrostatic pressure by taking into account the latitude entered in the SubOcean program when launching a run, because in the equation *p = ρ g h*, g varies with latitude.
* Carrier gas (“zero air”) flow and total flow, measured by the IQ-Flow regulators. The total flow is the sum of the carrier gas flow and the sample gas flow. The latter includes a large amount of water vapor diffusing through the membrane, in addition to dissolved gases.
* Cavity pressure: set at 30 mbar.
* Cavity temperature (“cellule temperature”): set at 40°C. When starting the instrument from ambient temperature, it usually takes from 30 min to 1 h to reach the set cavity temperature.
* L-shift and Error Standard: variables indicating the quality of the spectral fit. Error Standard is the variable to favor for cleaning the datasets from wrong spectral fits.
* Ringdown time: the variable indicates if the laser spectrometer behaves as expected or if it loses its performance (for instance due to mirrors becoming dirty). For SubOcean-CH4, it should be at about 13 µs (±1). For SubOcean-N2O, it should be around 26 µs (±1).
* Box temperature: temperature inside the pressure tube. Usually cools down after immersion and then slowly stabilizes.
* Carrier gas pressure (in the latest version of SubOcean-N2O): pressure in the small gas tank filled with carrier gas.
* PWM cellule temperature and pressure: not sure what it relates to…
* Laser temperature and laser flux: relatively constant values. No special care.
* Norm signal and value max: not sure what it relates to…

# Metadata for the CSV File

## Column Descriptions

1. \*\*Date\*\*: The date of the measurement (UTC).

2. \*\*Time\*\*: The time of the measurement (UTC).

3. \*\*Date calibrated\*\*: The date when the measurement system was calibrated (UTC).

4. \*\*Time calibrated\*\*: The time when the measurement system was calibrated (UTC).

5. \*\*[CH4] dissolved with water vapour (ppm)\*\*: Methane concentration dissolved with water vapor, measured in parts per million (ppm).

6. \*\*[CH4] dissolved with water vapour (nmol/L)\*\*: Methane concentration dissolved with water vapor, measured in nanomoles per liter (nmol/L).

7. \*\*[CH4] dissolved with constant dry gas flow (ppm)\*\*: Methane concentration dissolved with constant dry gas flow, measured in ppm.

8. \*\*[CH4] dissolved with constant dry gas flow (nmol/L)\*\*: Methane concentration dissolved with constant dry gas flow, measured in nmol/L.

9. \*\*[N2O] dissolved with water vapour (ppm)\*\*: Nitrous oxide concentration dissolved with water vapor, measured in ppm.

10. \*\*[N2O] dissolved with water vapour (nmol/L)\*\*: Nitrous oxide concentration dissolved with water vapor, measured in nmol/L.

11. \*\*[N2O] dissolved with constant dry gas flow (ppm)\*\*: Nitrous oxide concentration dissolved with constant dry gas flow, measured in ppm.

12. \*\*[N2O] dissolved with constant dry gas flow (nmol/L)\*\*: Nitrous oxide concentration dissolved with constant dry gas flow, measured in nmol/L.

13. \*\*[NH3] dissolved (ppm)\*\*: Ammonia concentration dissolved, measured in ppm.

14. \*\*Depth (meter)\*\*: Depth of measurement in meters, calculated from hydrostatic pressure considering latitude.

15. \*\*Hydrostatic Pressure Calibrated (bar)\*\*: Hydrostatic pressure, calibrated and measured in bars.

16. \*\*Carrier gas pressure calibrated (bar)\*\*: Carrier gas pressure, calibrated and measured in bars.

17. \*\*[CH4] measured (ppm)\*\*: Methane concentration measured directly, in ppm.

18. \*\*[N2O] measured (ppm)\*\*: Nitrous oxide concentration measured directly, in ppm.

19. \*\*[NH3] measured (ppm)\*\*: Ammonia concentration measured directly, in ppm.

20. \*\*[H2O] measured (%)\*\*: Water vapor concentration, measured in percentage (%).

21. \*\*Flow Carrier Gas (sccm)\*\*: Carrier gas flow, measured in standard cubic centimeters per minute (sccm).

22. \*\*Total Flow (sccm)\*\*: Total gas flow, measured in sccm, including water vapor and dissolved gases.

23. \*\*Cavity Pressure (mbar)\*\*: Pressure within the cavity, set at 30 mbar.

24. \*\*Cellule Temperature (Degree Celsius)\*\*: Temperature within the cavity, set at 40°C.

25. \*\*Hydrostatic pressure (bar)\*\*: Hydrostatic pressure measured during the run, in bars.

26. \*\*LShift\*\*: Variable related to spectral fit quality. Use in combination with Error Standard for cleaning datasets.

27. \*\*Error Standard\*\*: Indicator of spectral fit quality. Lower values indicate better fits.

28. \*\*Ringdown Time (microSec)\*\*: Laser spectrometer performance indicator. ~13 µs for SubOcean-CH4, ~26 µs for SubOcean-N2O.

29. \*\*Box Temperature (Degree Celsius)\*\*: Internal temperature of the pressure tube. Stabilizes over time after immersion.

30. \*\*Carrier gas pressure (dbar)\*\*: Pressure of carrier gas in the tank, measured in decibars (dbar).

31. \*\*PWM Cellule Temperature\*\*: Not clearly defined. Temperature of the cellule as part of the PWM system.

32. \*\*PWM Cellule Pressure\*\*: Not clearly defined. Pressure within the cellule as part of the PWM system.

33. \*\*Laser Temperature (Degree Celsius)\*\*: Laser temperature, typically constant.

34. \*\*Laser Flux\*\*: Laser flux, typically constant.

35. \*\*Norm Signal\*\*: Not clearly defined.

36. \*\*Value Max\*\*: Not clearly defined

*Specific to each instrument* :

* The concentrations or isotopic ratio measured in the optical cavity (in ppm for trace gases, in percent for water vapor).
* CH4, N2O and NH3 calculated as “dissolved”. Two columns are provided in each case. One noted *“with water vapor”* and the other noted *“with constant dry gas flow”.* The latter assumes that the dry sample gas flow through the membrane is constant at 0.25 ml/min. The first one calculates the dry sample gas flow by taking into account the total gas flow, the carrier gas flow and the water vapor flow based on the water vapor measurement in the cavity.

Anthony Lefebvre (A2PS) provided the following equations, which are embedded in the SubOcean software to lead to the corresponding columns.

“It's the *“dissolved with water vapour (ppm)”* column () and not the *“dissolved with constant dry gas flow (ppm)”* column () that is based on water concentration measurements. As a reminder, this dissolved concentration is simply a matter of normalizing the gas concentration measured in the spectrometer () by the portion of dry gas extracted from the membrane, and taking into account the membrane enrichment factor . In the general case, we have:

: total flow rate through the spectrometer

: dry gas flow rate extracted from the membrane

: flow rate of carrier gas flushing the membrane

: water vapor flow rate extracted from the membrane, calculated from the water concentration

And therefore:

The is obtained with the last formula, and is the most rigorous calculation. It is, however, directly affected by noise on . As the denominator can sometimes be close to zero (), this can have dramatic consequences on the result.

The is obtained with the formula just above, using a fixed of 0.25sccm.”

Note that in case the cavity temperature would not be stable at 40°C (setpoint), the A2PS software includes a correction on the concentrations, using the following equations:

H2O\_corr = H2Omeas/(2.469\*(Tcell-40)/100+1), with Tcell in °C

CH4corr = CH4meas/(0.925\*(Tcell-40)/100+1)

It is unclear to me in which column(s) this correction is applied in the dataset. It should be clarified with A2PS staff. Where is located this correction ? is it directly applied ?

**Post-deployment data treatment**:

Here is a list of routines to be developed and tested by Hugo, which would be much beneficial to us.

1. Data cleaning

The SubOcean datasets often include lines where the spectral fit was deficient. This leads to off-scale concentration measurements. This can possibly affect only one line from time to time (each line corresponds to one second of measurement) or it can also last over a few tens of seconds or even minutes when the bad spectral fit reflects a more serious problem in the instrument behavior.

One should use the *“Erreur standard”* column as the variable to clean up the dataset. However, the value of *“Erreur standard”* is also roughly proportional to the measured concentrations. Roberto Grilli (CNRS) proposes to use a criterion like (Err Std) ^2/[CH4] (Does [CH4] refers to the dissolved with vapour pressure or with constant dry gas flow?) to normalize the variable on the measured concentration (or with N2O if SubOcean-N2O). And then define a threshold above which the corresponding line is flagged as “NaN”.

One could define a specific threshold value based on the various runs made in the context of the GreenFjord expedition.

An option could also be to that the routine includes a cursor on this threshold value, combined with data visualization, so that the experimentalist can select the most suitable value as a function of the general noise of the corresponding measurement series. And then save the cleaned dataset for further data treatment.

1. Smoothing of H2O data for re-calculating the dissolved concentration with water vapor

As H2O is measured using the absorption spectrum in the optical cavity, there is inherent noise in the measurement. Calculating the dissolved gas concentration “line by line” directly and wrongly propagates this noise in the calculated concentrations.

One should consider generating a moving average of the H2O measurements, as its true value on the lee side of the membrane varies mostly at the time scale of minutes (effect of changing hydrostatic pressure for instance).

A routine including a cursor to test “by eye” different lengths of moving average could make sense. A more complicated option would be to calculate the Allan Variance of the concentration datasets based on different lengths of moving average on the H2O values, so that to select the optimal length. But it will work only in the configuration where the concentration does not significantly change in water over the corresponding duration.

I see that H2o measurements have artefacts, should we remove them before doing the moving average ? the moving average is influenced by these extremas Une image contenant texte, capture d’écran, ligne, Tracé

Description générée automatiquement

1. Corrections using off-line CTD data (temperature, salinity, dissolved oxygen) lets do that in the end

The membrane enrichment factor depends on temperature, salinity and dissolved oxygen.

A2PS computes it using the following equation:

=(1.9774+(0.0385-0.00316\*Salinity(psu))\*(WaterTemperature(°C)-2.67))\*(1+0.2286\*(O2diss(%)-0.2)/(0-0.2))

By default (to be checked), according to the A2PS software:

* The water temperature is set at 25°C,
* The water salinity is set at 34 psu,
* The dissolved oxygen is set at 21%.

Usually, a CTD provides the dissolved oxygen concentration both in mg/L or as a percentage of saturation. Therefore, one should use the percentage of saturation measured by the CTD and multiply it by 0.21 to get the right data input for the correction.

The A2PS software allows to provide corrected dissolved CH4 or N2O concentrations using offline CTD data, which are then interpolated over time or over pressure for the correction.

The format of the CTD file must then respect the following one:

* 4 columns, with Temperature, salinity, dissolved oxygen, date or time
* Header for temperature: “Tv+number+C”
* Header for salinity: “Sal+number”
* Header for oxygen: “Sbeox+number+PS”
* Header for date: “mm/dd/yyyy” or header for time “hh:mm:ss

A routine could be developed to offer better performance than the A2PS interface. In practice it would:

* Be able to handle different CTD file formats, without having to specifically extract a priori the 4 columns of interest: **GPT on column name recognition ?**
* Allow to compare the resulting SubOcean concentration profiles using constant values for T, S and O2, with those obtained from the CTD data, in order to evaluate their effect.**So we need to reprocess these concentration gases with these constant parameters ? would need all the equations right ?**

Warning : the CTD sensors reacts without delay. So, the measured values are given on the right time or depth scale. However, in the case of SubOcean instruments, there is a delay between what diffuses through the membrane and what is measured in the optical cavity. The delay depends on the carrier gas flow rate. Typically, with a carrier gas flow rate of 1.5 ml/min, the delay is of the order of 20 seconds. The lower the carrier gas flow rate, the longer the delay (time required to renew the gas content of the optical cavity).

Assume the equation is determined as following

20s = 0.33 min;

delayed\_time = c / Qcarr

0.33 [min] = c / 1.5 [ml /min]

C = 0.33 \* 1.5 = 0,5 mL

0,5/Qcarr =

To be checked with A2PS what equation to use to get the right delay as a function of a given carrier gas flow rate.

This delay implies that:

* The SubOcean data or the CTD data have to be adjusted with respect to time, so that they are put in synchroneity for the correction and for the data comparison,
* When plotting the SubOcean data as a function of depth, the delay must be taken into account. This explains why, when comparing downward and upward concentration profiles in the water column, anomalies usually do not overlap but appear offset.

With the three above sections of post-data treatment, we should then have final corrected datasets, ready for comparison with other ones and for plotting the results with different interfaces.

**Data visualization as a function of depth or of time**:

Anthony Lefebvre at A2PS has developed a MatLab interface allowing to plot all variables of the SubOcean datasets as a function of time. But also to show the different flows on the same plot (carrier gas, water vapor, dry sample gas). And to show the concentration profiles as a function of depth, with a color coding reflecting time (to easily differentiate downward and upward profiles).

It would be useful to rely on another interface:

* Providing easy options to change the graph characteristics, by zooming in for instance, or to change the axis labels and legend,
* Allowing to plot several variables in the same graph, for ease of comparison.

**Data visualization with other variables from different instruments**:

It would be optimal that we could rely on a routine allowing to plot the SubOcean data on the same depth scale as other variables measured with the CTD or with other sensors running in parallel.

Ideally, the routine would allow to zoom in/out and to select different variables coming from the different files, with a color coding based on time to clearly distinguish downward and upward profiles for each variable.

**Data visualization on 3D maps**:

Some of the SubOcean data have been or will be acquired by towing the instruments from a ship, while changing the depth of the deployment with time. This leads to profiles having to be visualized in 3D in the water body.

In an ideal case, an aquatic GPS is attached to the SubOcean frame, allowing to plot the SubOcean data along the 3D GPS coordinates.

The required routine would allow to easily incorporate the GPS coordinates (written in a separate file) as additional columns in the SubOcean dataset, paying attention to synchroneity.

In a less ideal case, the positioning data come from another interface provided by the ship. This will require some data extraction and treatment before incorporation of the 3D positioning in the SubOcean files.

Oceanographers usually use “Ocean Data View” as software to plot their data on top of bathymetry or other background georeferenced data. It would be optimal to rely on a routine interfacing the SubOcean datasets with such a software, handling in particular all aspects of pre-formatting, if required.