Bollettino di Geofisica Teorica ed Applicata Vol. 60, n. 3, pp. 517-530; September 2019
DOI 10.4430/bgta0266

Comments on computations about the Mediterranean Outflow composition

C. Millot

Retired, former research director at the National Center for Scientific Research (CNRS), Marseille, France

(Received: 25 November 2018; accepted: 17 January 2019)

ABSTRACT The Mediterranean Outflow (MO) is composed of four major intermediate and deep Mediterranean Waters (MWs) that mix with two major Atlantic Waters (AWs) in the Strait of Gibraltar. There is an ancient debate about whether, all along the Strait, these mixed MWs either totally mix together and give a homogeneous MO or remain individualised and give a heterogeneous MO. I do not address herein this scientific debate, which needs a large amount of data analyses that I will present in papers to come, but I comment on the available computations performed to specify the MO composition. In particular, a recent objective analysis of hydrological profiles aims at specifying, for each sample in either the Mediterranean Sea or the Strait, the percentages of the major MWs and AWs involved in the mixing while claiming using a clustering method classically used to specify water masses characteristics. I show that: i) the performed computations have nothing to do with such a method, ii) the used Euclidean distance cannot provide any sound result, iii) such an analysis always identifies one intermediate MW as having the largest percentage everywhere in the Strait and, more generally, iv) no statistical objective computation can provide sound results in regions where hydrological characteristics are rapidly evolving. I suggest that the sole reliable type of hydrological analyses in such a place is based, as I have done in the past, on the AWs-MWs mixing lines slopes and positions according to processes I am now able to specify, and I conclude that all available data sets account for the MO heterogeneity all along the Strait.

Key words: Mediterranean Outflow, Mediterranean Waters, clustering method.

1. Introduction

The number of major Mediterranean Waters (MWs) identified in the Strait of Gibraltar entrance, namely at Camarinal Sills (~5°45’ W, Fig. 1), as components of the Mediterranean Outflow (MO), as well as the number of major Atlantic Waters (AWs) they mix with, is no more debated since Naranjo *et al*. (2015) specify that their identification is “now in good agreement with the previous study of Millot (2014a)”. The actual debate is on how much the mixed MWs mix together, either mixing totally and giving a homogeneous MO within the Strait itself (from ~6°05’ W) before being split into four veins in the Strait exit (~6°25’ W) due to interactions with bathymetric features there [see for a review Naranjo *et al*. (2015)], or not totally mixing each others and leading to a

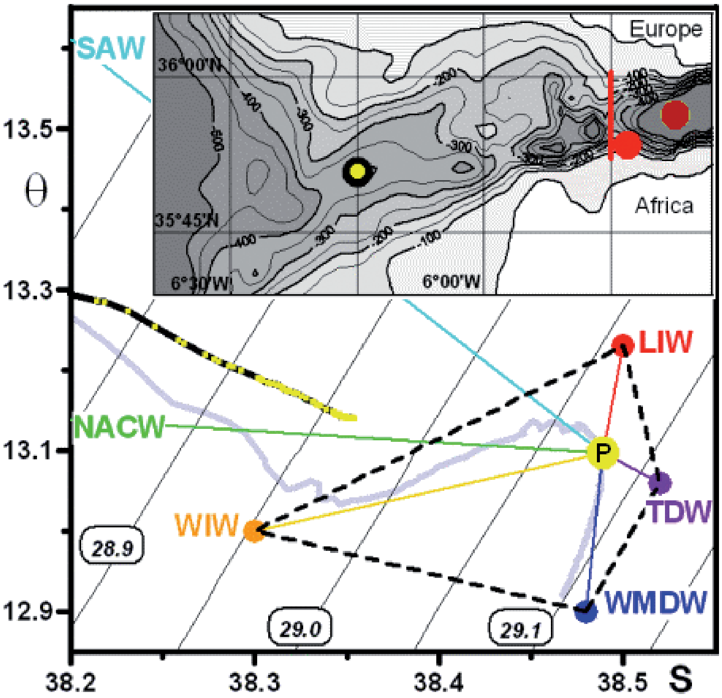


Fig. 1 - The insert shows the Strait of Gibraltar (isobaths in m), the Camarinal Sills (the red bar at 5°45' W), and the data location with: i) a brown dot in the Sea for the profile in gray; note that this sinuous profile (as all profiles east of Camarinal) from the mid 1980's GIBEX (MEDAR Group, 2002) is shifted by $\Delta\theta$=0.12 °C and ΔS=0.04 [consistently with Millot *et al*. (2006); θ are in °C and σ_θ isopycnals are in kg×m^{-3}] to be, for its densest part, within the quadrilateral (dashed lines) formed by the four MWs centroids [defined by Naranjo *et al*. (2015) from 2005-2014 data; their two AWs centroids are out of the figure near S=36-37], ii) a red dot at ~5°43' W for the CTD yo-yo time series represented in Fig. 6, and iii) a yellow-black dot at 6°15' W for the profile #5 from the MO-2009 experiment in black; note that this profile is relatively straight (as all profiles west of Camarinal) and that the 1-db data in yellow evidence a relatively large homogeneity at depth. The coloured lengths between a sample/point (yellow dot P) of the gray profile in the Sea and the centroids visualise $\theta_c^{\,2} + S_c^{\,2}$ (Pythagoras theorem) and are specified in Fig. 2a for a particular yellow point similar to those from the black profile out of the Sea.

heterogeneous MO with the four MWs directly forming the four veins in the ocean (e.g. Millot, 2014a). Strangely, no in situ experiment, no theoretical analysis and no numerical simulation have ever been dedicated to confront the homogeneous vs. heterogeneous hypotheses so that considering the data analyses available up to now deserves specific interest. Before presenting my own analysis of huge data sets collected in 1985-1986, from the (Mediterranean) Sea to the Strait and, in 2009, in the Strait exit, which is done in papers submitted recently, I comment herein on the computations performed by Naranjo *et al*. (2015) that I have been asked to compare with my own approach.

Naranjo *et al*. (2015) analyse CTD profiles in the Sea and the Strait with the aim of quantifying, for each sample, the relative amounts of four MWs (WIW, LIW, TDW and WMDW) and two AWs (SAW and NACW) they a priori identify from historical data by specific θ-S sets located as centroids on a potential temperature-salinity (θ-S) diagram (Fig. 1). It is worth noting that the present paper: i) deals with computational aspects, not oceanographic ones, so that there is no need to specify the MWs and AWs acronyms, ii) does not require any specific background, just to be familiar with θ-S diagrams and aware of the meaning of the potential density anomaly (σ_θ),

iii) does not even need to have the Naranjo *et al*. (2015) article beside, even though I emphasise in my papers to come the extreme value of their data set and how much it supports my own understanding of the processes.

Fig. 1 shows that vertical profiles in the sea (as the gray one) are sinuous, and that samples associated with the MWs are well within the quadrilateral formed by the four MWs centroids (the orange, red, violet, and blue dots); there, and classically for Mediterranean oceanographers, distances on a θ-S diagram (named lengths hereafter) between a given sample (as specified by the yellow dot P) and each of the centroids are visually estimated, leading to the conclusion that, for instance, that sample at a relatively great depth is mainly composed of TDW more or less mixed with the other MWs (without any AW). Vertical profiles in the Strait (as the black one) are relatively straight and samples are out of the quadrilateral; there, and at least to my knowledge, no relationship has ever been made between the lengths and the percentages of this or that water mass.

In section 2.1, I present the computations of Naranjo *et al*. (2015) that they clearly specify within 18 lines in their p. 44, the understanding of which is sufficient for understanding my analysis herein. In both the Sea and the Strait, they link Euclidean distances involving θ, S, and σ_θ between each sample and each centroid (such distances are thus more complex than the lengths in Fig. 1; this will be specified in Fig. 2a with the relative amounts of MWs and AWs in the sample. I analyse the information that can be provided by such a distance definition and I explain why this information can be realistic when samples are in the quadrilateral while it is totally unrealistic when samples are out of the quadrilateral. In section 2.2, I formalise the approach I have used up to now in the Strait (e.g. Millot, 2014a) that is based on the hypothesis of MWs mixing individually with the AWs and leading to relatively straight θ-S diagrams; this has been requested by some referees to illustrate my own understanding of the processes but validation needs a huge amount of data, which is done with an original analysis of ~ 500 profiles in papers to come. In section 3, I discuss the assertions postulated by Naranjo *et al*. (2015) and the fact that their computations have nothing to do with a clustering method they claim to apply since such a method aims at objectively specifying, from Euclidean distances between the samples only, the centroids of an expected set of waters that are thus *a posteriori* defined. I conclude, in section 4, emphasising the fact that, in the Strait where hydrological conditions dramatically evolve in both space and time, an expert subjective analysis is much more sound and efficient than any objective method.

2. Material and methods

2.1. The Naranjo et al. *(2015) computations*

To compute the relative amounts of MWs and AWs they expect for each sample, Naranjo *et al*. (2015) first define θ_c, S_c, and σ_c (a 3-variable set) as the normalised (with the variables ranges) differences between the potential temperature, salinity, and potential density of a given sample with those of the *a priori* defined centroids C's. Then, they define the squared Euclidean distance of the sample to a given C by $D_c=\theta_c^2+S_c^2+\sigma_c^2$, which explains why variables must be normalised, the % of that C in the sample being D_c^{-1} divided by the sum $\sum_{i=1}^{6} D_i^{-1}$ over all six C's, which comes to identify the dominant C with the shortest D_c, hence ordering the different % from comparisons of one distance with another. A Euclidean distance defined with only θ and S (a 2-variable set) would have led to a trivial identification of the dominant water from just a visual sight at a θ-S diagram and a distance measured as a length in Fig. 1, as classically done when analysing such diagrams. Considering a 3rd a priori independent variable (e.g. the depth, the colour or the radioactivity of a given sample) would lead to imagine distances measured as lengths in a 3-dimension diagram that could be intuitively analysed in a similar way. Now, the 3-variable chosen set is, to say the least, strange since σ_c depends on θ_c and S_c (isopycnals are plotted on a θ-S diagram), which has direct consequences analysed with Figs. 2, 3, and 4.

Figs. 2 and 3 compare the distances D_c of any sample with the WIW and LIW centroids, based on arguments that can apply to any other pair of MWs. Fig. 2 considers a sample P on the W-M-L line that is the perpendicular bisector to the WIW-LIW segment, hence that would be visually equidistant from the WIW and LIW centroids, specifies the definition of the variables considered by Naranjo *et al*. (2015) and allows figuring what D_c is for a sample P on the W-M segment (σ_θ < 29.0 kg×m^{-3} as anywhere west of Camarinal). The square of the two identical lengths P-WIW and P-LIW equals $\theta_c^2+S_c^2$ (Pythagoras theorem) so that the distances D_c with both centroids only depend on σ_c^2: all samples on the W-M segment having a σ_c^2 lower with WIW than with LIW are thus computed as composed mainly of WIW and are coloured in orange.

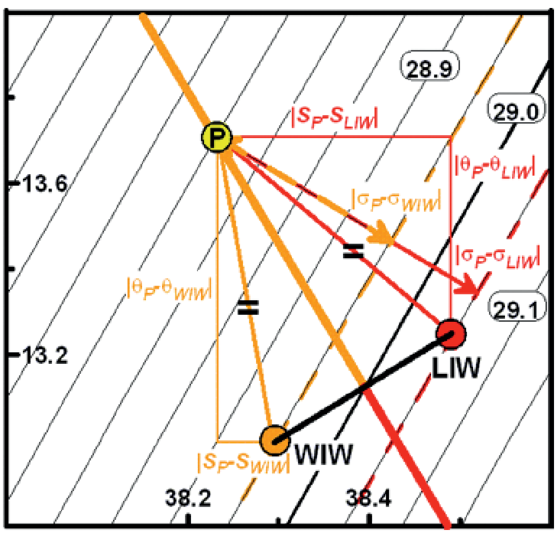


Fig. 2 - The WIW (bright orange) and LIW (bright red) centroids [as defined by Naranjo *et al*. (2015)] have been slightly moved to be located exactly on the 28.95 and 29.05 kg×m^{-3} isopycnals (dashed coloured), respectively, so that point M, in the middle of the WIW-LIW segment, is roughly located on the 29.0 kg×m^{-3} isopycnal while W-L is the perpendicular bisector to that segment. The θ_c and S_c variables are specified by segments parallel to the axes (e.g. the vertical $|\theta_P$-$\theta_{WIW}|$ in orange) for a sample P located on the lightest side (W-M) of this bisector. Both (squared) lengths between P and the two centroids (θ_c^2+S_c^2, identified by the = sign) being equal, the two Euclidean distances D_c only depend on the σ_c values that are specified by the arrow-type segments perpendicular to the isopycnals (e.g. the $|\sigma_P$-$\sigma_{WIW}|$ in orange). For all samples P on the W-M segment, distances D_c will be smaller with WIW, so that the whole W-M is coloured in orange (the same for M-L in red).

More generally, Fig. 3 shows that all samples with σ_θ < 29.0 kg×m^{-3} (hence having a σ_c^2 lower with WIW) that are on the WIW side of the W-M-L line (hence having a θ_c^2+S_c^2 lower with WIW) are orange too (the reverse for LIW in red), hence defining orange and red sectors. But actual sectors where samples are computed as composed mainly of WIW or LIW are in fact larger, even though they cannot be exactly delimited (by the dashed black line) since this needs comparing (θ_c^2+S_c^2) with σ_c^2, which depends on the normalisation used, hence on the ranges considered [that are not specified by Naranjo *et al*. (2015)]; this leads to light orange (yellow) and light red (pink) additional sectors. In any event the whole #5 is orange and thus computed as composed of WIW

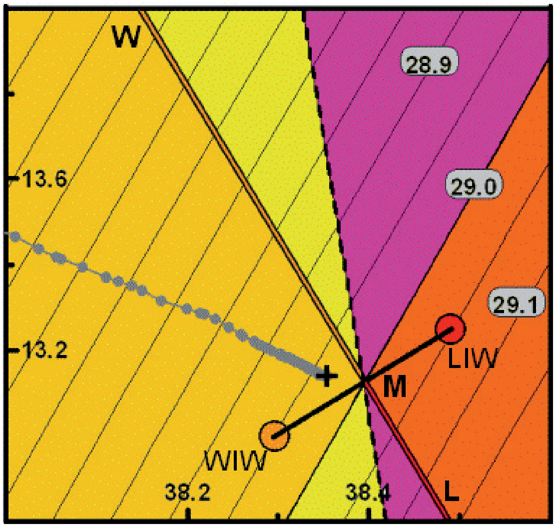


Fig. 3 - As compared to the WIW (bright orange) and LIW (bright red) centroids, all samples in the orange (red) sector are computed by Naranjo *et al*. (2015) as mainly composed of WIW (LIW), just because of the lengths $\theta_c^2+S_c^2$ as made explicit with Fig. 2 and whatever the normalisation of σ_c^2 is. This normalisation specifies the exact position of the dashed line and extends these orange and red sectors with, respectively, the yellow and pink ones. Profile #5 already plotted in Fig. 1 is plotted (gray dots, θ-S-σ_{max} set as +) as a representative of all profiles in the Strait, in particular those shown by Naranjo *et al*. (2015).

mainly, as are most if not all of the Naranjo *et al*. (2015) samples in the Strait when they do not postulate the absence of WIW.

Profile #5 is also plotted in Fig. 4 that represents the major results inferred from Fig. 3 in the general case considered by Naranjo *et al*. (2015) west of Camarinal when they postulate the absence of WIW, hence the occurrence of LIW, TDW and WMDW only. This figure is simplified by the fact that, even though TDW and WMDW are defined by θ-S centroids that lead to a slight realistic $\Delta\sigma_\theta$ ~0.002 kg×m^{-3} between them, both are considered herein as having σ_θ ~29.11 kg×m^{-3} so that, LIW being associated with σ_θ ~29.06 kg×m^{-3}, it can be dealt with a common intermediate isopycnal at σ_θ ~29.085 kg×m^{-3}. When considering the lengths between any sample and the LIW and TDW (or WMDW) centroids, all of them will be shorter for LIW when both i) $\sigma_\theta < 29.085$ kg×m^{-3} and ii) the samples are on the LIW side of the LIW-TDW (or WMDW) perpendicular bisector. All samples of the whole #5 being in the red sector in Fig. 4 will thus be, for sure,

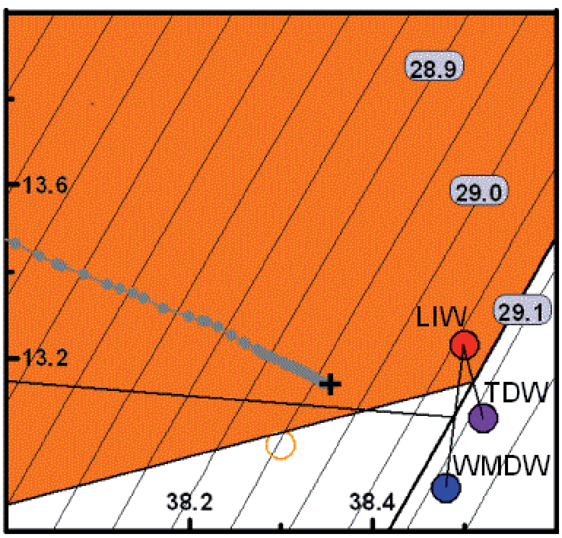


Fig. 4 - The four MWs centroids are plotted from the Naranjo *et al*. (2015) values, hence on inferred σ_θ ~29.113 (WMDW), ~29.111 (TDW) and ~29.06 (LIW) while WIW (~28.95 kg×m^{-3}) is generally not considered. For simplicity of the discussion, TDW and WMDW are expected to have the same σ_θ (29.11 kg×m^{-3}) so that the thick isopycnal is the 29.085 [(29.11 + 29.06)/2] kg/m^{-3} one while the black lines represent either the LIW-TDW and LIW-WMDW segments or their perpendicular bisectors on the lowest σ_θ side. All samples in the red sector, whatever the normalisations of θ_c, S_c, and σ_c are, will be computed as mainly composed of LIW.

computed as mainly composed of LIW; but sectors that would correspond to the light-colours sectors in Fig. 3 (not represented in Fig. 4) would enlarge the number of samples computed as mainly composed of LIW. This clearly explains why all Naranjo *et al*. (2015) samples west of Camarinal are computed as mainly composed of LIW when they postulate the absence of WIW.

2.2. The θ-S diagram approach

Considering my own understanding of the Strait functioning (e.g. Millot, 2014a) and the fact that vertical profiles in the Strait are relatively straight on a θ-S diagram, I have always hypothesised (Millot, 2008, 2009; Millot and Garcia-Lafuente, 2011) that they result from MWs mixing individually with one or the other of the AWs. I am now able to formalise this hypothesis with simple Microsoft Excel-based calculations [as done in Millot (2013, 2014b)] that, in no way, should be taken for a proper simulation of mixing processes between the AWs and the MO. These computations and Figs. 5 and 6 consider a hypothetically homogeneous MO but could apply to any individualised MW. In Fig. 5, a relatively thick and homogeneous surface layer (either SAW or NACW) lies above a homogeneous bottom layer of limited thickness (the MO), each layer being characterised at t=0 by specific values V of either θ or S (the θ-S pairs), and the MO being relatively static. The mixing at depth i and time t is simulated by a running mean over three depth levels such as $V_{i,t}=(V_{i-1,t-1}+V_{i+1,t-1})/2$ and, if the bottom is at depth j, the $V_{j,t}$ value that cannot be computed is replaced by $V_{j-1,t}$.

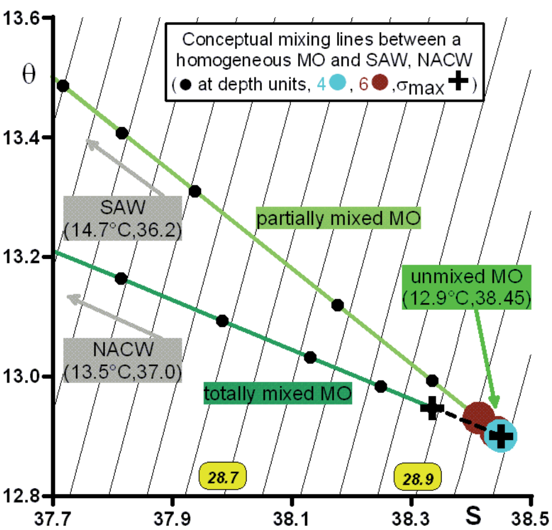
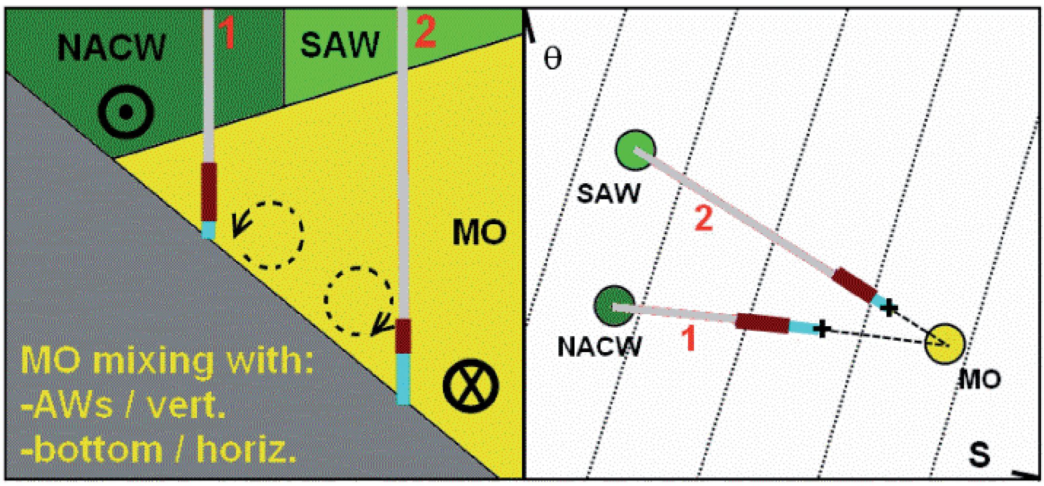


Fig. 5 - Conceptual mixing lines between a homogeneous/unmixed MO (12.9 °C, 38.45) and either SAW (14.7 °C, 36.2) or NACW (13.5 °C, 37.0) schematising a mixing of the MO either partial or total. In both cases, a single black cross ends the two mixing lines and specifies the associated σ_{max} value. Four superimposed cyan and two (apparent, in fact six in total) close brown dots aim at providing information about the structure with depth of the partially mixed MO while the dashed segment identifies the linear fit to the θ-S diagram in the totally mixed MO case.

Such mixing lines across the MO west from Camarinal allow checking whether it is, there, homogeneous or not. If these conditions (a homogeneous MO of limited thickness mixing with homogeneous AWs) represent "Stage A", then direct consequences representing "Stage B" are: i) all mixing lines on a θ-S diagram converge towards the original MO point (θ-S original pair), ii) a total mixing of the MO results in a mixing line ending by a single θ-S pair different from the original one, iii) any partial mixing of the MO results in a mixing line reaching the MO original point, and mixing points accumulating towards this original point; note that, in this

iii) case, comparing the number of points accumulated "over or nearly over" the MO original one (cyan dots) with the number of points "nearby" (brown dots) allows characterising the thickness and stratification of the MO upper levels. With "A implies B", and just because then "non-B implies non-A", the non-observation of any of the features i) to iii) over the whole MO at a given location west from Camarinal implies that the MO there is not homogeneous, i.e. is heterogeneous.

Note that such a static MO, as soon as totally mixed (as with NACW in Fig. 5), would no more be possibly identified with any specific θ-S pair, and that a mixing line ending with a unique θ-S pair can also result from a cast not deep enough. Also, in the actual case of a set of MWs mixing with a set of AWs, most mixing lines cross each others: four MWs mixing with two AWs lead to six crossing points at intermediate locations where any objective identification of the involved waters would be impossible. In any event, while mixing with the AWs on the vertical, a totally mixed MO (or any of the MWs in case they are juxtaposed side by side) continuously flows with relatively large velocities (several m×s^{-1}) so that it simultaneously encounters over the bottom an important turbulent mixing that will always tend to re-homogenise it (Fig. 6). Since the MO flows along a sloping bottom (the African slope in Fig. 6a; a similar diagram could have been drawn for the European slope even though NACW is less clearly occurring there), furthermore against the AWs, hence with a sloping AWs-MO interface and possibly different AWs, the actual ends (in cyan and brown) of the two mixing lines in Fig. 6b indicate θ-S-σ_θ sets having markedly different characteristics. Note that the ratios between the number of samples in cyan (very homogeneous) and brown (relatively homogeneous) quantified by the lengths of the corresponding segments aim at schematising the fact that, in such a case, the MO mixes more intensively with NACW than with SAW: as a result of the homogenisation on the bottom,



a b

Fig. 6 - Schematisation of the deeper part of CTD vertical profiles #1 (across NACW and the MO) and #2 (across SAW and the MO) in the south of the Strait, and the associated mixing lines representative of a total mixing of the MO (as with NACW in Fig. 5) and cyan and brown colours (as for SAW in Fig. 5); also schematised by the lengths of the cyan and brown segments is the fact that, in this example, the deepest MO mixes less intensively with SAW than with NACW, so that re-homogenisation on the bottom leads to a MO thicker and more homogeneous in the deep Strait. Note that SAW spreads in fact over the whole Strait and that NACW actually flows as do intermediate waters in the Sea (Millot and Taupier-Letage, 2005), being constrained by the Coriolis effect along the African slope (e.g. Millot, 2014a).

the MO in cyan as sampled by profile #2 is thicker (on the left) and more homogeneous (on the right) than the one sampled by profile #1.

In the simplest case of an initially homogeneous MO mixing with the same AW along its course in the Strait, each of the two profiles in Fig. 6 indicates that, once the MO has been totally mixed, its re-homogenisation continuously occurs along the same mixing line so that, even though continuously characterised by a homogeneous bottom layer, it can no more be identified with a specific *θ-S* pair. In case such an initially homogeneous MO, or a MO that has become homogeneous, mixes with different AWs that are lying in a direction parallel to the MO streamlines, the two profiles in Fig. 6 indicate that it will be automatically split in two homogeneous components juxtaposed side by side. In these two simple cases, fits to the mixing lines indicate the original MO point. In a more complex case, the MO, or any of the MWs it is initially composed of, mixes with different AWs along its course in the Strait: the *θ-S* pair characterising any homogeneous component (e.g. the cyan dots in Figs. 6 and 7) will thus be continually shifted from its original location (the yellow MO dot in Fig. 6) along different and varying mixing lines: inferring the single original (at Camarinal) *θ-S* pair from any *θ-S* pair measured anywhere downstream will be possible only if mixing lines have been specified with a small sampling interval upstream.

This being hypothesised, note that the hundreds of profiles in the Strait I have analysed (from both the mid 1980s GIBEX and the MO-2009 experiment) display a similar structure in their deepest part illustrated in Fig. 7 by #5 that is very homogeneous over ~20 m (the cyan dots) and still relatively homogeneous over ~70 m (the brown dots), a linear fit being possibly inferred from this second set of samples (impossible from the first one). Would the centroids defined by Naranjo *et al*. (2015) be correct, the MW sampled by #5 would be somehow in between TDW and WMDW. Now, the yo-yo time series collected in 1985 at Camarinal (Fig. 1) and never analysed before my own papers to come indicate a totally different distribution (in a *θ-S* diagram as in Fig. 7) of the MWs that should be i) validated by similar data sets (never collected there since then to my knowledge) and ii) collected contemporaneously with any sample in the Strait itself, especially when considering the long-term changes in both the MWs (Millot *et al*., 2006) and the AWs (Millot, 2007). It is clear from this example that, in addition to data collected at Camarinal, i) data in the Strait should be collected downstream with an as small as possible spacing (to make the cyan dots as close as possible, first to the coloured crosses and then to each others), ii) would the 1985 data had been shifted also in *S* by ~0.04, the violet or blue crosses would have been located roughly in between the violet and blue dots and indicated by the fit (my papers to come clearly account for #5 in blue), iii) the linear fit approach can be less conclusive than in this example; for instance, in case the fit indicates the densest MWs, it will be hardly difficult with such a slope of the fit to differentiate the violet from the blue, which will need taking into account the intensity of the mixing, the same sets in cyan being obtained with either a relatively intense mixing of the blue MW or a relatively reduced mixing of the violet MW. Note that, depending on the AW involved in the mixing, the slope of the fits can be larger (Millot, 2008) or lower (my papers to come) by several tens (*sic*) of degrees (with such a $\Delta\theta/\Delta S$ ratio), which will lead to very different conditions making easier or harder the MWs identification.

Finally, note that the crosses in Fig. 7, that represent raw data (actual samples) collected near Camarinal during a single day in the mid 1980s, form relatively well identified groups of homogeneous σ_θ values that are relatively heterogeneous in both θ and S. Would the identification

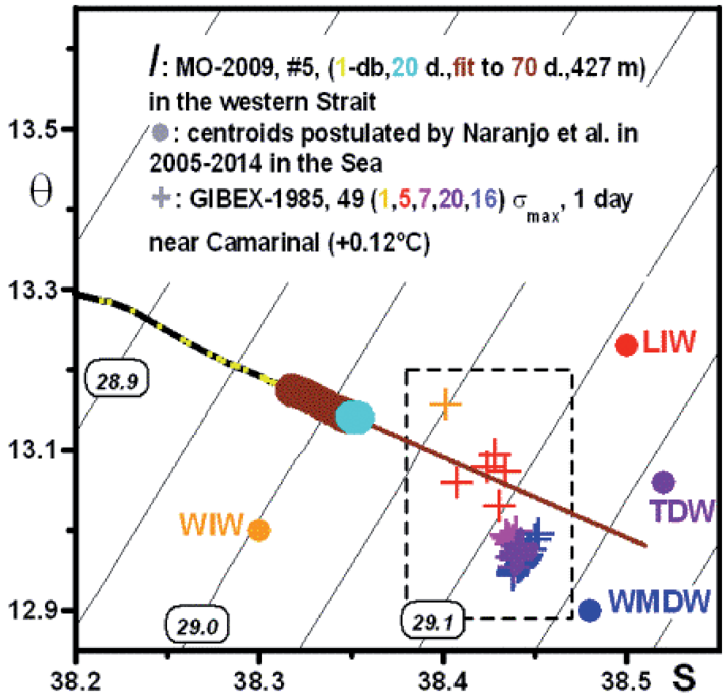


Fig. 7 - The profile (1-db data) is #5 from the MO-2009 experiment (6°15' W, 427 m, Fig.1); the 20 densest samples are plotted in cyan over the 70 densest ones in brown (50 are still visible), the linear fit in brown being inferred from the 70 densest samples. The coloured dots are the MWs centroids defined by Naranjo *et al*. (2015) from contemporaneous (2005-2014) data. The similarly-coloured crosses in the dash rectangle are the 49 θ-S-$\sigma_\theta(z)$ maximum (in σ_θ values) sets inferred from a yo-yo time series of 49 vertical profiles collected within ~1 day at ~5°43' W (Fig. 1) during GIBEX (MEDAR Group, 2002) in 1985. These sets are identified with each of the MWs occurring successively in the study area by just considering their distribution with respect to σ_θ values that form well identified groups (of 1, 5, 7, 20, and 16 elements), even in the largest σ_θ range. These sets have been arbitrarily shifted by +0.12 °C (see Fig. 1 legend; not shifted in S to avoid having the crosses over the dots and keep this didactic figure clear) to conveniently illustrate: i) the differences between the distribution of actual (directly inferred from raw data) MWs' characteristics in the Strait entrance and centroids arbitrarily (without any objective data analysis) defined, and ii) how linear fits to the lower part of CTD vertical profiles can be efficiently used. Note that the pink and violet groups of sets in these 1985 data are probably linked to two different components of TDW occurring at the time, as expected by Millot (2009).

I make (details in paper to come) between these groups and the MWs be correct, then major features could be inferred. First, when considering the relatively large spreading in both θ and S of the blue crosses (as compared in particular to the TDW-WMDW segment) that are associated with σ_θ in the relatively narrow range of 29.0955 - 29.0975 kg×m^{-3}, hence being probably identified with the densest MW (WMDW), one can wonder what is the significance of any centroid concept. In addition, the mean σ_θ of the violet group being 29.0925 kg×m^{-3}, the difference with the mean σ_θ of the blue group (29.0965 kg×m^{-3}) is 0.004 kg×m^{-3}, that is twice the difference I inferred from the σ_θ values I computed for the Naranjo *et al*. (2015) TDW and WMDW centroids (see the legend of Fig. 4): identification of the MWs at Camarinal with σ_θ could be easier than with θ and S [and consistent with Millot (2013, 2014b)]. In any case, the identification expected from raw actual parameters (the crosses) is dramatically different from the identification assumed up to now (something like the centroids) and illustrates, at least, our misunderstanding of the functioning of the Sea and the characteristics of the MO.

3. Discussion

The major aim of this discussion is to compare objective methods with expert approaches, in particular in places such as the Strait of Gibraltar where mixing processes are so intense that they have prevented, up to recent years, from addressing hydrological characteristics of the MO. Since my own expert approach can be fully validated only after the publication of a huge amount of data analyses in papers to come, I can only consider herein the objective method of Naranjo *et al*. (2015) who claim doing clustering, and I found interesting, since I was previously unaware of such a method, to comment on it.

In the simplest case, a clustering method is an objective method that aims at defining water masses characteristics (as centroids) from a given set of data and consists in grouping 2-variable samples, as θ-S sets, according to a Euclidean distance based on the normalised (by either the range or the standard deviation) θ and S differences between them, hence using the Pythagoras theorem in a θ-S diagram. The procedure hence consists in making/naming each sample a single cluster, measuring all distances between two clusters and combining, in every iteration, the closest two clusters into one new cluster, hence iteratively lowering the number of clusters until it reaches a designated level that is the number of expected water masses; the θ-S characteristics of the remaining clusters are the so-called centroids that are thus specified as a result of the method.

Kim *et al*. (1991) deal with two seasons and two depths, they normalise with the ranges and come with a given set of water masses per season and depth. A less simple, but still comprehensible grouping consists in considering 3-variable samples, adding the depth to θ and S generally normalised by the standard deviation, the method hence specifying the depth at which this or that water mass is preferentially found, eventually over time (Warn-Varnas *et al*., 2005). Such a 3-variable Euclidean distance can be modulated by a weighting factor considering the geographical separation between clusters (Hur *et al*., 1999). Therefore, all clustering methods dealing with a given set of oceanographic data are only based on a set of samples, they do not make any a priori on the θ-S characteristics (the so-called centroids), eventually the depth and/or the geographical distribution and/or the seasonality, of a set of water masses, especially since these methods aim at providing them. Among all articles available to me out of the oceanography domain, the PhD thesis of Yan (2005) specifies seven critical steps in clustering analysis (their p. 17), of which only two are addressed herein.

3.1. Objects used in clustering should represent the cluster structure (if any) in the data

"Objects" are CTD profiles and the "cluster structure" is understood as the result *a posteriori* obtained, if it does exist, from classical clustering computations that would be similar to the centroids a priori defined by Naranjo *et al*. (2015).

Note that all profiles upstream from (east of) Camarinal are sinuous on a θ-S diagram, indicating superimposed MWs just slightly mixing together; samples there are located amidst the polygon defined in Fig. 1 by the MWs centroids that could coherently represent the cluster structure, allowing the identification of this or that MW along the profile, hence over depth. On the contrary, profiles downstream from (west of) Camarinal are almost straight mixing lines that have their MWs end moving more and more (downstream, hence with longitude) away from that polygon. The representativity of the centroids and the reliability of any objective method with such straight mixing lines can thus be questioned.

Also note that the AWs centroids are far away from the MWs ones (Figs. 1 and 5), essentially giving two very different groups of centroids; whether such a dichotomy is compatible with the use of a clustering method can be questioned. In addition and in the AWs ranges, neither the SAW nor the NACW actual centroids are directly inferred in the mixing and centroids should have been chosen at the base of these AWs, which does not allow an accurate definition. Furthermore, since many profiles in the Strait (in all available data sets) cross each others, both the necessity and the scientific arguments that lead to differentiate SAW from NACW in the Naranjo *et al*. (2015) computations can be questioned. Considering the densest MWs in the Sea as composed of SAW and NACW, even in small quantities, is also strange.

3.2. Variables selected for clustering should provide sufficient and relevant information for the discovery of the correct cluster structure

A clustering method with N variables consists in minimising a Euclidean distance in an N-dimension space. With only two variables such as θ and S close to the surface, Kim *et al*. (1991) were able to specify groups of values having specific characteristics and distribution, even identifying a new water mass. One can think that when dealing with only θ and S along a given transect upstream from Camarinal, classical clustering computations could have allowed specifying both these MWs' θ-S characteristics and then the groups of clusters essentially associated with them. Plotting these groups over depth and latitude could have given sound and realistic distribution of this or that MW.

With three variables in a 3-dimension space, basic results can be easily imagined: for instance, Hur *et al*. (1999) consider θ, S, and the depth d, so that the method provides an objective way to specify the distribution over depth, hence along the z axis, of the various clusters specified as points in the 2-dimension/x-y θ-S plane/diagram; this is exactly what was achieved by Warn-Varnas *et al*. (2005). In the case of the data upstream from Camarinal, a problem might have occurred since the MWs, in particular, might not be distributed only with depth but also with latitude, as now said by Naranjo *et al*. (2015) to support my own results (e.g. Millot, 2014a) that have been previously refuted by the team from the University of Malaga [all their papers up to Garcia-Lafuente *et al*. (2015)].

Whatever the case, the selection of θ, S, and σ_θ provides a much less obvious information since, even though any objective method ignores the θ-S-σ_θ relationship, one is tempted to think that a given σ_θ is an isopycnal in the 2-dimension θ-S/x-y plane, which is hardly compatible with σ_θ plotted over z: while the physical sense of a cloud of d-points on the z-axis corresponding to a given θ-S cluster is easily imagined (if the water masses are distributed only with depth and not with latitude), the interest provided by a single σ_θ-point on the z-axis is much "less obvious". One can just note that there is a roughly linear relationship between σ_θ and d since (*in situ*) density increases with depth.

To summarise, the set of centroids representing the MWs, be it forming a quadrilateral (Fig. 1) or a triangle (Fig. 4), can be soundly used only to analyse samples i) included in this polygon and ii) do not evidencing an intense mixing, especially when such mixing can hypothetically involve only one MW and one AW (my understanding). Where such conditions occur, that is only east of Camarinal, a classical squared distance ($D_c=\theta_c^2+S_c^2$) could provide sound results; but the squared distance $D_c=\theta_c^2+S_c^2+\sigma_c^2$ will obviously provide biased results. West of Camarinal, where samples rapidly exit from the polygon, even a classical θ-S-d set of variables (*a fortiori* the inadequate θ-S-

σ_θ set) will provide unreliable results: with the Naranjo *et al*. (2015)'s set of profiles (similar to #5) having a relatively low $\theta(S)$ slope, if WIW is considered (Figs. 2 and 3), the MO is computed as essentially composed of WIW; if WIW is not considered (Fig. 4), the MO is computed as essentially composed of LIW, whatever the amounts of TDW and WMDW could be.

It is thus clear that Naranjo *et al*. (2015) did not perform a clustering analysis. They just computed Euclidean distances between samples and a set of *a priori* fixed benchmarks or centroids claimed to represent water masses. Their computations just consist in quantifying distances estimated visually as lengths in a θ-S diagram, allowing simply to say "this sample is mainly composed of the nearest water mass". But even a sound clustering method dealing with sound Euclidean distances between samples/clusters and iteratively sorting them to come with a set of water masses cannot be used where intense mixing occurs. Obviously, the characteristics of any water mass cannot be specified by any objective method in regions where these characteristics are dramatically evolving.

4. Conclusion

The clustering method is a fully objective tool to analyse a set of hydrological samples on the basis of a Euclidean distance involving parameters that are the normalised differences between pairs of samples for variables such as θ and S, and essentially aims at specifying in fine water masses characteristics named centroids on a θ-S diagram. Contrary to what they claim, Naranjo *et al*. (2015) do not perform any clustering analysis since they fix *a priori* such centroids and then compute distances between a given sample and the centroids, directly linking the relative distances to the relative amounts of this or that water mass in the sample. In addition, and while some validated clustering analyses have considered a third true variable (such as the depth or the geographical location), Naranjo *et al*. (2015) consider the θ-S-σ_θ set of parameters that is biased since these three parameters are not independent, which is, to say the least, strange. Whatever the singularity of their "method" and the strangeness of the distance they use, they conclude that a sample is mainly composed of the nearest centroid everywhere, hence whatever the shape of the θ-S diagram could be, either sinuous (as in the Sea east of Camarinal) or straight (in the Strait west of Camarinal).

Their computations provide results that compare not too badly with those inferred from an expert subjective analysis (as done when visually analysing a θ-S diagram) in regions where mixing processes are relatively moderate, hence only in the Sea where, essentially, the MWs are superimposed and lead to θ-S diagrams that are sinuous within the MWs centroids polygon. But west of Camarinal, intense AWs-MWs mixing processes lead to θ-S diagrams that are almost straight mixing lines having their densest values rapidly exciting the MWs centroids polygon, then being always closer (for the AWs encountered during the mentioned experiments, hence for the associated mixing lines location and slope) to their WIW centroid. But mixing lines having a steeper $\theta(S)$ slope and involving the upper part of SAW can be closer to their LIW centroid (e.g. Millot, 2008). I demonstrate herein why, with their data set and if they consider WIW, Naranjo *et al*. (2015) compute a MO essentially composed of WIW and why, if they do not consider WIW, they compute a MO essentially composed of LIW, whatever the actual amounts of TDW and WMDW could be. I think that it is because LIW is "more famous" than WIW, a MW the team has ignored in all their previous papers [despite its mention in Gascard and Richez (1985); in my papers to come I show 2009 (contemporaneous) data evidencing a WIW amount even larger than the LIW amount], that they postulate the absence of WIW; note that, when doing this, they modify their hypotheses about the MO original composition in order to get what they think are less inconsistent results.

In any event, any objective method cannot specify the characteristics of any water mass where such characteristics are dramatically evolving, especially when different AWs and different MWs must be considered, which can lead to mixing lines crossing each others. I thus disagree with their overall comment [p. 46 of Naranjo *et al*. (2015)]: "Should we have displaced any of the centroids of the MWs by a tiny distance, the algorithm would have possibly returned a different prevailing cluster. The reasonable conclusion is that the MWs are hardly distinguishable once the MO has passed the Camarinal sills and that the sensible option is to speak of a unique "Mediterranean water"." I claim that displacing any of the centroids, which can be done visually from all figures herein, and/or using any other distance, or any sound objective method, in a region where almost straight θ-S diagrams indicate a complete mixing of the original MWs would have given similar inconsistent results.

The Naranjo *et al*. (2015) analysis does not demonstrate that the MO becomes homogeneous downstream from Camarinal, which thus remains a postulate without still any support while my previous and forthcoming papers clearly account for its continuous heterogeneity all along the Strait, from the Sea to the Ocean.

Acknowledgements. This research did not receive any specific grant from funding agencies in public, commercial, or not-for-profit sectors. I have very much appreciated the open-mindedness of Josep Pelegri from ICM Barcelona who kindly provided me, without any restriction, with all the CTD data from the MO-2009 experiment. I am grateful to the referees for having provided me with constructive and positive comments.

REFERENCES

García-Lafuente J., Naranjo C., Sánchez-Leal R., Sammartino S., Bellanco M.J., Sánchez-Garrido J.C. and Soto-Navarro J.; 2015: *On the origin of the seasonal and interannual T-S variability of the inflow through the Strait of Gibraltar*. Deep Sea Res. Part I, **101**, 38-53, doi: 10.1016/j.dsr.2015.03.005.

Gascard J.C. and Richez C.; 1985: *Water masses and circulation in the western Alboran Sea and in the Straits of Gibraltar*. Prog. Oceanogr., **15**, 157-216.

Hur H.B., Jacobs G.A. and Teague W.J.; 1999: *Monthly variations of water masses in the Yellow and East China Seas, November 6, 1998*. J. Oceanogr., **55**, 171-184, doi: 10.1023/A:1007885828278.

Kim K., Kim K.R., Rhee T.S., Rho H.K., Limeburner R. and Beardsley R.C.; 1991: *Identification of water masses in the Yellow Sea and the East China Sea by cluster analysis*. Elsevier Oceanogr. Series, **54**, 253-267, doi: 10.1016/S0422-9894(08)70100-4.

MEDAR Group; 2002: *The MEDATLAS/2002 database: Mediterranean and Black Sea database of temperature salinity and bio-chemical parameters and climatological atlas*. IFREMER Edition, Brest, France, 4 CDroms.

Millot C.; 2007: *Interannual salinification of the Mediterranean inflow*. Geophys. Res. Lett., **34**, L21069, doi: 10.1029/2007/GL031179.

Millot C.; 2008: *Short-term variability of the Mediterranean in- and out-flows*. Geophys. Res. Lett., **35**, L15603, doi: 10.1029/2008/GL033762.

Millot C.; 2009: *Another description of the Mediterranean Sea outflow*. Progr. Oceanogr., **82**, 101-124, doi: 10.1016/j.pocean.2009.04.016.

Millot C.; 2013: *Levantine intermediate water characteristics: an astounding general misunderstanding!* Sci. Marina, **77**, 217-232, doi: 10.3989/scimar.03518.13A.

Millot C.; 2014a: *The Mediterranean Sea in- and out-flows' heterogeneities*. Progr. Oceanogr., **120**, 254-278, doi: 10.1016/j.pocean.2013.09.007.

Millot C.; 2014b: *Levantine intermediate water characteristics: an astounding general misunderstanding! (addendum)*. Sci. Marina, **78**, 165-171, doi: 10.3989/scimar.04045.30H.

Millot C. and Taupier-Letage I.; 2005: *Circulation in the Mediterranean Sea*. In: The Handbook of Environmental Chemistry, Water Pollution, vol. 5K, pp. 29-66, doi: 10.1007/b107143.

Millot C. and Garcia-Lafuente J.; 2011: *The seasonal and fortnightly variability of the Mediterranean outflow*. Ocean Sci., **7**, 1-8, doi: 10.5194/os-7-1-2011.

Millot C., Candela J., Fuda J.L. and Tber Y.; 2006: *Large warming and salinification of the Mediterranean outflow due to changes in its composition*. Deep-Sea Res., **53**, 656-666, doi: 10.1016/j.dsr.2005.12.017.

Naranjo C., Sammartino S., Garcia-Lafuente J., Bellanco M. and Taupier-Letage I.; 2015: *Mediterranean waters along and across the Strait of Gibraltar, characterization and zonal modification*. Deep-Sea Res. Part I, **105**, 41-52, doi: 10.1016/j.dsr.2015.08.003.

Warn-Varnas A., Gangopadhyay A., Hawkins J.A. and Robinson A.R.; 2005: *Wilkinson Basin area water masses: a revisit with EOFs*. Cont. Shelf Res., **25**, 277-296, doi: 10.1016/j.csr.2004.09.005.

Yan M.; 2005: *Methods of determining the number of clusters in a data set and a new clustering criterion*. PH.D. Thesis in Statistics, Faculty of the Virginia Polytechnic Institute & State University Blacksburg, VA, USA, 107 pp., <pdfs.semanticscholar.org/8492/87862e01639378d2301fe5489378df4adf59.pdf>.

Corresponding author: Claude Millot
Les Katikias, 83150 Bandol, France
E-mail: ailesetiles@gmail.com