Reply to Anonymous Referee #1

We wish to thank Referee #1 for the very detailed and pertinent comments that helped to greatly improve the manuscript. We addressed all the general and specific comments as detailed below with a line-by-line response provided in italic.

In the revised version of the manuscript we will clearly identifying its objectives in line with the CMEMS special issue and the operational oceanography requirements. We will provide more details on the processing chain implementation to clarify several issues pointed out in this review. We will restructure section 2 on data and methods. We will evaluate the effects of the time window on the match-up analyses. We will improve the quality of Figure 1. We will add the validation of Kd using BGCArgo float data with a new figure in the body of the manuscript. We will add new figures in the supplementary materials.

This is the review of the manuscript "The Mediterranean Ocean Colour Level-3 operational multi-sensor processing" by Volpe et al. The paper is mostly a description of the near-real-time/delayed-time processing of ocean color data in the framework of CMEMS. Results are mostly based on a validation analysis. Overall, the text reads more like a project report than a scientific paper, even if intended for a special issue on the European Copernicus Marine Service. It is actually incomplete or confusing in its description of various aspects of the processing (as shown by the list of secondary comments). So the text should be thoroughly reorganized to aim at clear descriptions (starting with the actual objectives of the paper) and the 'scientific' part should be reinforced before being considered for publication. These points are further developed below.

As demonstrated by the list of secondary comments, the description of the processing is often unclear and the pertinence of some processing steps is insufficiently supported. I understand that the authors ran into objective difficulties in documenting the processing chain with some elements that are fairly technical and apparently not fully described in literature (bow-tie effect, removing outliers, bias correction, smoothing. . .). I am wondering if some elements are described in more details in CMEMS (or other) reports / ATBDs that could be cited while simplifying the description of technical elements. An alternative is to use this paper as an opportunity to justify the choices made in the various processing steps and focus the work essentially on these. After the 'technical' part, the 'scientific' part is restricted to 2 pages (out of 13 pages of text) and could easily be reinforced with more discussion (currently there is very little description and discussion of the results, no comparison with published validation results, . . .). Based on the objectives of the paper, the authors should choose which part ('technical' or 'validation) to strengthen.

In the new version of the paper we will highlight the nature of the work by clearly identifying its objectives in line with the CMEMS special issue and the operational oceanography requirements. We will provide more details on the processing chain implementation to clarify several issues pointed out in this review. We will evaluate the effects of the time window on the match-up analyses. We will add the validation of Kd using BGCArgo float data. We will add new figures in the supplementary materials.

Besides the various points listed below, a more general lack of information can be noted when it comes to the comparison of processing chains in the validation analysis, CMEMS processing versus OC-CCI. The paper says that the CMEMS product used for validation is the near-real time (NRT) output. For me, this would mean that the data used in the validation analysis are those obtained in NRT mode, preserved as the processing went (i.e., computed with preliminary ancillary data, calibration at the date of acquisition, with climatology computed with very little data, etc. . .). In

that case validation results reflect the quality of the data as downloaded NRT by users. Otherwise they are DT data, or even fully reprocessed data if they result from a consistent processing all the way through the time period (in terms of calibration, climatology computed with multiple years. . .). This (and implications) must be made clear in the manuscript (actually, how the validation results evolve as data are brought from NRT to DT and to reprocessed data is an interesting point).

The Reviewer is right. In the revised manuscript we will make clear that the matchup analysis refers to DT data. It would be very interesting to investigate on the quality drift that data may have between NRT and DT. However, since the CMEMS processing chain downloads the Level-2 from space agencies and this work involves the Level-1 to Level-2 processing step, this would require a reprocessing from Level-1 (using the NRT configuration) of all the days involved in the matchup analysis. The current chain is not meant for that.

Besides the mode (NRT, DT) actually associated with the CMEMS data used in the analysis, more discussion should be given comparing those with the CCI data. The manuscript forgets to mention other large differences between the CMEMS and CCI processing, including the atmospheric correction for certain sensors. In the extraction step, the grids of the products are also different (1-km versus 4-km if I am not mistaken). The study should clearly identify all possible sources of differences between the CMEMS and the CCI stream.

We will briefly describe the OC-CCI method with respect to the inter-sensor bias correction and mention the different atmospheric correction for certain sensors. The different spatial resolution was already mentioned in the original submission (page 5, line 7). However, we will rephrase the paragraph to make it more clear.

I have an issue with the field data. They are introduced in the study to explain the MedOC4 algorithm and the validation analysis. In such a context, a word on the uncertainties associated with the data would be necessary. Not clear to me is the distinction between the data serving for the development of MedOC4 and the data used for validation with match-ups (points of the former going into the latter are not fully independent validation points). As they are used for validation, BOUSSOLE and AERONET-OC data should be described a bit better. While I'm not familiar with the BOUSSOLE data distribution, there is a clear data policy for use of the AERONET-OC data (offer of authorship if I'm not mistaken) and I'm wondering if this has been respected (there is not even an acknowledgment in the manuscript).

Figure 1 showing the Cal/Val dataset will be redrawn to better show their distribution, and from which it will appear clear that the calibration data are not being used for validation. The reviewer is right; the section acknowledgement was missing. We will fill it acknowledging AERONET-OC, BOUSSOLE, BGC-Argo and NASA OBPG as those providing high quality data and which we are grateful to. The AERONET-OC PI declined our offer of co- authorship .

Below are detailed comments, with requests for clarification/corrections and suggestions for improving the text. I'd recommend numbering the sections and sub-sections.

We will number all sections and sub-sections.

Page 1

l.10: I'd suggest: "multi-sensor processing applied to the Mediterranean Sea by the Ocean Colour Thematic Assembly Centre of the Copernicus. . .": The abstract should be readable by readers who don't know about CMEMS, TAC, . . .

OK

l.11: “A basin-scale. . .”

OK

l.12: “to fine-tune”

OK

l.14: “than those”

OK

l.15: “The Mediteranean. . .”: information associated with this sentence should be relocated in the beginning of the abstract.

OK

l.21: “CMEMS delivers. . .” rather than ‘includes’

OK

l.27: “users who”

OK

l.28: please define all acronyms at first use.

OK

Page 2

l.2: “near-real time (NRT) and delayed time (DT) modes.”

OK

l.3: in OC jargon, monthly data computed from daily data are still termed L3.

We will add the terms “Within CMEMS” at the beginning of the sentence to stress that this is the current terminology adopted by the Copernicus service.

l.15: is there a reference for this approach?

We will remove the statement from the introduction and add sentence framing this approach in the conclusions.

l.17-18: heavy sentence about 2 important benefits; should be reworded.

We will rephrase and split the sentence.

l.20: “are derived”

OK

l.23: “is foreseen”: so it is still not the case, which is at variance with the abstract.

This sentence refers to OLCI, which is still not included in the multi-sensor product. It will probably be so by 2019. We do not see any discrepancy with what is stated into the Abstract, where OLCI is not mentioned.

l.26: “DT data”

OK

l.26: “precise”: what does this indicate exactly? l.27: “both to be accurate”: does it mean DT and REP? then the sentence should be restructured. What does ‘consistent’ mean here?

The sentence will be rephrased into “As such, DT data are expected to be as accurate as timeliness allows. The accuracy of REP data need to be stable in time as these data are consistently processed with a single software version.”

l.27: “For the sake of timeliness. . .”

OK

Page 3

l.4: “This work. . .”: this paragraph describes the structure of the work, but the primary objective of the whole study is not clear. Page 2, line 28, “one of the aims” is mentioned, but how does it fit here, and what are the other aims?

We will add a sentence explicitly stating that the overall objective of the work is to provide Copernicus users with a comprehensive description of the method currently applied in the OCTAC context of CMEMS to produce the multi-sensor ocean colour product over the Mediterranean Sea. Propagating the REP configuration to the DT processing mode is part of the method.

l.14: “relies on an in situ. . .”

OK

l.18: odd sentence; it means “absorption due to CDOM, absorption due to algal and non-algal particles, absorption due to TSM, and both AOPs and IOPs.” absorption is part of the IOPs, so which are the others measured? l.20: are the in-situ IOPs used in this study?

We will rephrase the sentence to avoid confusion and to more clearly specify what are the in situ data that were actually used in this study.

l.30: are the authors sure about the choice of acronym . . . a large part of the scientific community is sufficiently well-versed in latin languages to understand the meaning of the word.

We will change to “Multi-level data processing is achieved using the Software for the Elaboration of Radiometer Data Acquisitions (SERDA)”.

Page 4

l.2: “normalised by”

OK

l.3: Kl and Ku should be defined.

OK

l.9: “using the primary sub-surface quantities, it is then”

OK

l.10: “such as the Q-factor”

OK

l.18: “Chl”: are Chl data from BOUSSOLE used in this work?

Actually they are not. We will remove Chl from the sentence.

Page 5

l.6: “using the OC-CCI. . .”

OK

l.7: this sentence reads: “OC-CCI... at 1-km ... rather than at 4-km for OC-CCI”: unclear.

We will rephrase the entire paragraph.

l.14: “Single-sensor pre-processing for NRT/DT modes” I presume.

*It will be changed to “*NRT/DT single-sensor pre-processing*”*

l.15: “quality-checked”: what does it entail?

We will add a sentence to briefly explain the kind of data quality checks that are operationally performed.

l.15: I assume that the atmospheric correction applied in these cases is l2gen. This should be mentioned together with an appropriate reference.

We will add a sentence in the previous paragraph to mention that the NRT/DT data are downloaded from the OBPG at NASA that uses the l2gen in its default parameterisation for the L1 to L2 processing. Later in the paragraph we will add similar details for the REP processing.

l.22: “that reflects”. . . “stripes originate”

OK

l.32: “the dimension . . .”: it happens also for the other sensors, doesn’t it?

Yes. We will rephrase the paragraph to make it clear.

Page 6

l.2: “these missing values”: what is the benefit of this step?

We will specify that the benefit is particularly evident in view of the sensor merging.

l.6: and what about SeaWiFS and MERIS?

To generalize we removed the reference to MODIS-AQUA and VIIRS. Please see also response to comment P7.L7

l.6: “space agencies”: in that case it is only NASA.

OK

l.7: “atmospheric correction failure”: this is a strong step; I would assume that Rrs associated with that flag is just not usable as the AC failed according to the software. What is the criterion used by the software to consider the AC a ‘failure’? what is then the status of Rrs? The authors should also support this decision with robust evidence that the resulting Rrs is actually valid.

We will add details on the processing of pixels identified through the bowtie removal flag. We will clarify that the atmospheric correction failure flag is not applied to VIIRS because it overlaps with the bowtie removal flag for almost all water pixels. We tested the 645 granules (each of 3200x3232 pixels) acquired over the Mediterranean Sea in 100 days (10 April 2018 to 18 July 2018) and found that only in 31 pixels the atmospheric correction failure flag was raised for pixels not affected by bow tie deletion or any of the other OBPG standard flags.

l.8: “to avoid”

OK

l.8: “salt and pepper”: please define what this refers to.

OK

l.9: “removing all isolated pixels...”: in general, this part needs clearer explanation. How is an “isolated pixel” defined?

We will add a sentence to specify the meaning of the isolated pixel.

l.13: please provide characteristics of the map (extent, resolution).

OK

l.17: “spectra” from different missions?

Yes. We actually wrote “to merge single-sensor Rrs spectra into a single spectrum”. Adding from “different missions” would be redundant.

l.21: “differ by”

OK

l.25: “In general”: may be removed.

OK

l.30: “in theory”: and in practice? Might be removed.

OK

l.31: “apply algorithms to derive geophysical products”

OK

Page 7

l.3-5: not clear what the idea is here.

We will remove this sentence.

l.7: “Differences between MODIS and VIIRS”; and what about SeaWiFS and MERIS? In general some aspects of the manuscript associated with NRT/DT seem focused exclusively on MODIS/VIIRS (that has some logic since they are still active) but how SeaWiFS and MERIS are handled should also be described as the related products are used in the validation analysis.

We will add a sentence in section 2.2 (satellite data processing chain) to clarify that the current NRT/DT processing only involves MODIS-AQUA and VIIRS platforms, and that for the sole scope of the product validation we used the Multi-sensor production chain described in that section (2.2) to process the entire satellite data archive from 1997, thus including SeaWiFS and MERIS data.

l.11: “normalized”: does the BRDF correction make use of the OCI Chla value? In that, it is not consistent with the MedOC4 values. This should be acknowledged.

We will add a sentence acknowledging the issue in the result section when commenting the Rrs matchup analysis, focusing on the trade-offs between accuracy and timeliness in the operational oceanography framework. Furthermore, this issue will be also the topic of a new paragraph discussing the the accuracy of Rrs in section 4.

l.17: “we tested”: by doing what? l.19: “inter-calibration”: it is certainly a factor but it is not the only one that could apply. I would still argue that geometry can play a significant role in the differences. Operating the atmospheric correction with different bands might also have an impact through the AC code (eliciting different responses by the AC code and its assumptions/simplifications).

At pixel scale, the cosine of the scattering angle was compared with the ratio between MODIS-AQUA and VIIRS Rrs for each band. As it will be shown by the new figure in Supplementary Material, no specific pattern was evident through this analysis. We will acknowledge that there might be some other factors linked to the AC influencing the difference between sensors.

l.22: it does not seem that the bias correction operated by CCI is described in that reference.

OK, reference will be corrected. This paragraph will be partially rewritten to address the series of comments below about the implemented bias-correction method.

> l.25: “climatological”: what are the periods used for each sensor? is it the same? l.28: “two steps”: but then 3 are listed. l.29: “temporary” ? Page 8: l.1: aren’t the same equations used by CCI? l.5: “weight w” l.8: “and time”: not necessary. l.8: “as for OC-CCI”: this should be written before. I think CCI also operates some type of spatial averaging.

l.12: “deemed insufficient”: on the basis of what?

One of the output of Figure 6 is that the performances of SeaWiFS at 670 nm were the worst as compared with the other bands and sensors. This justifies the choice of not using this sensor band as reference for bias adjusting the other sensors. This sentence will be rephrased.

l.13: “40%”: what is the reason for so many missing values at 670 nm? Negative values? But earlier in the text, it is written that the quality check removes all incomplete spectra (if Rrs is <0 at any one band). So at that point, working with incomplete spectra for match-ups is inconsistent with the processing chain.

In the “Flagging & Mosaicking” paragraph, we will correct “one negative value within the spectrum (excluding the NIR bands) is enough for the entire spectrum to be rejected” replacing NIR with 670nm.

l.22: which “method”?

We will replace “method” with “merging”.

l.24: “are only the result of the merging procedure”

OK

l.26-29: I have to say that I don’t understand how the approach works. Volpe et al. (2018) not being published, the explanation should be clearer. For instance, is the climatology field computed as in the previous section? And then it is not clear how the smoothing operates.

We will rephrase the sentence into “To prevent the occurrence of such horizontal discontinuities, here we apply the smoothing procedure described in Volpe et al. (2018) and based on the use of the climatology field, described below”. Volpe et al. (2018) is now published. We will also add a few sentences to better explain how the smoothing operates.

Page 9

l.2-8: The examples are not so obvious to me. I don’t even see the Rhone plume on any map (should be a pattern leaving the coast. . .).

We will add labels on Figure 3e to easy the reading.

l.12: “In both cases”: this is true for the cases shown but is it a general result? In any case, there is no reason to expect that the bias-corrected merged data would be closer to the field data than a simple average.

The validation results of Figure 6 show that this is a general result.

l.15: “The climatology field is obtained. . .”: climatology is also mentioned in the 2 previous sections (bias correction, merging), so it is not clear what this SeaWiFS Chla climatology is for, nor why it is computed in a different way.

We will make a clear difference between the daily climatology bias map used in the context of the bias correction and the field climatology used for the sensor merging. We will also remove the incorrect bit “using the MedOC4 regional algorithm for CHL (Volpe et al., 2007)”. We will add a sentence to mention that the next version of the processing chain will include a climatology field

computed following the method adopted in the bias correction section.

l.20: “this has been estimated”: how? (or where?)

We will add the appropriate reference.

l.22: “To overcome. . .”: but this type of filtering was already introduced for the Level-2 data. Do outliers appear again later in the processing ? Speaking of ‘biases’ here is not appropriate.

Since it is mentioned earlier in the manuscript, we will remove this part from this context.

l.32:“Even though the latter now show performance comparable to that of empirical algorithms. . .”

OK

Page 10

l.1: “in discussing the characteristics that data must have to be used. . ..”

OK

l.2: “pointed out that . . ..”: the sentence should be re-written.

We will rephrase the sentence

l.2: “theoretical”: to be replaced by ‘semi-analytical’ (what is a theoretical algorithm?)

OK

l.10: “weather conditions”

OK

l.15: “To identify. . . the identification. . .”: sentence to be re-written.

We will rephrase the sentence.

l.17: “average spectra”: the covariance matrix is also needed for such an approach.

We will add the necessary details to the sentence.

l.21: “user survey”: refers to the CCI user consultation?

Yes, we will change it.

l.6-24: in general I think this paragraph may be too long as the approach has already been well described in literature. Regardless of length I don’t find it clear for a reader without a prior knowledge of the method.

We hope that after the suggested changes the paragraph will read better.

Page 11

l.12: “unable to perform”

We will add the results of validation analysis performed using the BGC-Argo float L3 data (Organelli et al., 2017, Earth Syst. Sci. Data).

l.15: “Inherent Optical Properties”: I was therefore expecting validation results for the IOPs while there are none.

We will remove the IOPs from this sentence. A reference to the Pitarch et 2016 al for the assessment of QAA based bbp in the the Mediterranean Sea will be added in the methods section at 2.2.1 where the use of QAA is first described.

l.18: “in most instances”: some example stats would be appropriate.

We will rephrase this sentence to show that the overall validation results are not significantly affected by the temporal window. We will add a new figure in Supplementary Material to show the statistics behaviour with a variable temporal window.

Page 12:

l.2: “used the NRT”: meaning data processed as they were in NRT conditions (with preliminary ancillary data, calibration known at the time of acquisition, preliminary climatology, etc. . .)?

We will correct NRT into NRT/DT. As mentioned earlier, we will add a couple of sentences in Section 2.2 (Satellite Data Processing Chain) where we better explain the differences between NRT and DT data and that for the sole scope of the product validation we used the NRT/DT production chain described in that section (2.2) to process the entire satellite data archive from 1997, thus including SeaWiFS and MERIS data.

l.12: “common spectral behaviour”: is this a point that could be discussed?

We will add some discussion on this point.

l.18: “significantly”: was any statistical test performed?

No, we did not perform any statistical tests. However, “do not significantly differ” here referred to the fact that the standard deviation bars in Figure 6 do overlap to one another. We will rephrase the statement by replacing “significantly” with “substantially”.

Page 13

l.4: “significant”; what does it mean here?

We will change “significant” to “relevant”.

l.5: “In the NIR”: in the red?

OK

Table 4: why are there less matchups at 670 nm (I understood that only complete spectra were kept)? The fact that the Multi and CCI products have a different number of match-ups should be discussed.

In the revised version of the manuscript we will expand on this.

l.12: “NRT”: or DT?

We will correct to NRT/DT.

Figure 1: not so easy to distinguish validation from development points.

We will make it better

Figure 3: remind that the climatology fields are from a daily climatology.

OK

Figure 5: “Pope and Fry 1972”?

Corrected to Mueller (2000).

Table 5: writes ‘REP’ versus ‘Multi’ while Table writes ‘CCI’ versus ‘Multi’; coherence is needed to avoid confusion. While Table 4 shows different numbers of match-ups for the 2 products, here the numbers are the same. Is there an explanation for this?

A thorough check throughout the manuscript will ensure consistency of the use of REP to identify the dataset and CCIv3 to identify the source processing chain for Rrs.