We would like to extend our gratitude to the valuable reviews and contributions by the three anonymous referees and Matteo Ottaviani to our manuscript. We hereby provide detailed responses to their comments.

Blue – Author Response

Green – Manuscript text

Matteo Ottaviani

Thank you Matteo Ottaviani for your review.

As requested I compiled a list of items in relation to the review of your manuscript: "Quality control of automated hyperspectral remote sensing measurements from a seaborne platform" submitted to Ocean Science and available at: www.ocean-sci-discuss.net/8/613/2011/. The quality of the paper is rather affected by a not rigorous use of the English language. I would suggest to undergo a full revision of the manuscript assisted, if need be, by a native speaker, making sure that both the grammar and the punctuation correctly flow. Just to give a few examples:

-"underway measurements were made" does not make sense

-"complimentary" is not the same as "complementary"

-"Quantitative colour of seawater interpretation" is mesmerizing

We have asked native speakers to proofread our revised manuscript as suggested.

Notwithstanding these revisions, I list in the following a list of more technical/ methodological issues that would be in my opinion convenient to resolve.

- It is a standard remote sensing technique to threshold the sunglint signal via the Cox-Munk distribution. Comments should be included on how the proposed masks perform compared to this method.

In our work we aim to develop a methodology that limits dependence on models rather more on empirical work based on physical measurements and then use the findings to hopefully help refine available models. We have therefore added comments to include Cox and Munk work

'It is therefore important that in future studies the contribution of whitecaps and foam be specifically investigated with respect to sunglint flagging, possibly utilising novel glint measurement apparatus (Ottaviani et al., 2008a; Ottaviani et al., 2008c) integrated with sunglint and wave models e.g. (Cox and Munk, 1954; Munk, 2009).'

Here we are agreeing that our method still needs to be refined but with the help of other works we can further improve it.

- Sunglint data are often presented as "erroneous" or "invalid"; I think this is imprecise since instrument accuracy is not determined by the scene. "Sunglint-contaminated", "Sunglint-affected" can be a more correct terminology.

Revised and changed.

- I'd be curious to see how the results would change if no normalization were used in the analysis.

In this study the results are not taking into account normalization. We state in the manuscript the technique of normalization was only implemented to show the distinct features of the spectra while maintaining the spectral shapes.

'Data normalisation was applied maintaining the spectral shape while simplifying comparison of spectra, dividing each L_W and R_{RS} measurement by the maximum value for each measurement. However, for determining the new flag, the actual computed spectral measurements L_W and R_{RS} were used.'

Additionally, normalisation herein implemented is consistent to the methodology used in meteorological flagging (Wernand, 2002). In ocean colour, Mueller et al (2003) protocol, the normalisation utilises a modelled solar irradiation which we replace here with the maximum value for each measurement. We also use different notation so as to avoid any ambiguity of terms.

- Page 616, line 23: The angular dependence of L_sfc and L_sky is not clear. How about the illumination conditions, theta_sun? What is PHI?

According to Mobley (1999) and Fougnie et al. (1999) optimal radiometric optical measurements should be collected for L_sfc and L_sky ~ (40 - 45) ° to the nadir and zenith and phi ~ (90-135) ° to minimise the effects of sunglint. Phi is the relative azimuthal angle of the sensor to the sun. We assume that illumination conditions are sufficient for optical measurements meeting the two conditions

1. E_S (λ = 480 nm) > 20 mW/m^2 · nm sets a threshold for which significant E_S (λ) can be measured, 2. E_S (λ = 470 nm) / E_S (λ = 680 nm) < 1 will mask spectra affected by dawn/dusk radiation, (Wernand, 2002). As for theta_sun since we use RAMSES-ACC hyperspectral cosine irradiance meter (TriOS, Germany) to account for the variable ship motion with respect to the zenith.

- Page 617, line 21-22: what are the assumptions behind Eq. 1? Is it sufficient to define either the "remote sensing reflectance" or the "water-leaving" radiance? Why does the reflection coefficient depend on cloud cover?

It is assumed the reflected light detectable using an optical sensor, partly influenced by the air-sea refraction of light, is a fraction of the incident light at the sea surface. Yes we believe it is sufficient to define R_RS or L_W as it is common practice in recent works e.g. IOCCG, 2006 and Mueller et al, 2003. It has been reported that cloud scattered radiance is enhanced and spectrally dissimilar with respects to clear-sky radiance. This variability has been noted to enhance the level of reflected skylight changing its wavelength dependence (Mueller et al, 2003). Therefore cloud cover is assumed to influence the reflection coefficient.

- Page 618, lines 5-10: this is a long periphrasis to state that PHI_derived is the relative azimuth. Are all these angles measured due North, or is PHI_sensor measured with respect to the vessel's heading?

Revised.

'where Φ_{sun} is the sun's azimuthal position computed using SPA, Φ_{ship} ship's azimuthal heading from the DGPS due North, and the sensor azimuthal heading with respect to ship heading Φ_{sensor} = 85°.'

- Page 618, lines 5-10: if "water absorbs light in the NIR", what contribution does the "and/or scattering influenced by optically active water constituents" give?

Revised.

'The empirical sunglint flag was developed on the premise that open seawater is assumed to absorb all light in the NIR. Thus any light signal measured by an optical sensor would be sea surface reflectance or atmospheric scattered radiance. However, in turbid waters scattering influenced by optically active seawater constituents also contributes to this light signal measurable by an optical sensor (Jamet et al., 2011).'

- Page 618, line 17: At the end of the paper the reference to Mueller et al. (2003) is made, but since major findings are associated to the evaluate of the "goodness" of this angular range, its significance should be explained/discussed.

We agree and we have revised the statement as follows;

'Automated and unmanned above-water optical measurements based on recommend optical sensor setup for example the reported 90° ≤ Φ ≤ 135° (Mueller et al., 2003) cannot be completely achieved and does not guarantee valid measurements, an observation noted by e.g. Aas (2010). In this study it has been shown, with caveats that valid measurements can be obtained for 0° ≤ Φ ≤ 360° from an unmanned seaborne platform. There is some probability that the assumed valid measurements can be erroneous, as it is widely accepted that the validity and accuracy of R_{RS} or L_W is also influenced by the air-sea interface reflection coefficient $\rho_{air\text{-}sea}$.'

- Page 619, line 3: This step nr. 3 is actually not a step, and the paragraph indentation is weird. First indentation: to what end must the test be repeated? Second indentation: what defines a "classical" band ratioing and "dominant" bands? This reflect in the flowchart in Fig. 3 which seems overly complicated. At the end, I was confused as to what was actually visually inspected and what part of the process is automated.

Revised. The repeat task has been explained further as follows

'first evaluation test used the NIR mean spectra values for in set Nns and Ns to obtain threshold values specific to each set, this test was then repeated using the minimum spectra values in the NIR for set Nns and Ns so as to obtain unique feature for each set'

Concerning the terms and bands we revised the phrase as;

'involved band ratioing based on characteristic spectral bands both in the VIS and NIR'.

To explain further what was visually inspected we have revised the paragraph,

'simplified activity diagram illustrates the steps that were manually executed', 'visually (manually) inspected and classified into; Nns – image set without sunglint or Ns – sunglint-affected image set',

Page 620, line 13-17: the whole paragraph is redundant.

Revised.

'A spectra assessment for the sample sets indicated that R_{RS} (NIR) and L_W (NIR) were significantly enhanced for the set Ns compared to set Nns. The enhanced spectra in set Ns was assumed to be a result of sunglint, whitecaps/foam, and to some extent known or unknown optically active seawater constituents, identified during the manual sea surface image analysis.'

- Page 621, line 1: The sentence is ambiguous. It is the wind-roughened surface that generates sunglint, not viceversa.

Revised. That is correct 'wind-roughened surface that generates sunglint'

Page 621, line 5-ff: the whole paragraph on "band rating" should be revised, to explicitly state what the relevant ratio are, and for what purpose. "Band ratio" is hardly a "method" per se; " 940 nm is a band of water absorption and not only of precipitable water. I lost the point after line 19. Are the threshold setup to then let them vary? What does "iteratively" imply in this context?

We agree and have revised the whole paragraph. You are also right about 940nm as being also the band of water absorption.

'Spectral band ratioing, a conventional approach in remote sensing algorithms known to reveal unique attributes of two or more spectral bands, was implemented to identify the differences in Nns and Ns. Characteristic spectral bands were used based on other studies and their physical properties; …..'

We have revised the procedure to explain the threshold value generating task.

'The threshold values were obtained by repetitive testing of possible threshold values i.e. after statistically computing the min, max and mean for spectra in the NIR. The computed statistics provided a starting point and were then iteratively varied to obtain the best threshold values aimed at; i.) masking/eliminating as many measurements in the sunglint-affected set Ns, and ii.) unmasking/keeping as many measurements in the sunglint free set Nns. The performance test summarized in Table 1, revealed the best sunglint flag conditions'

-Page 622, line 17-19: the whole sentence appears more than needed in the manuscript.

We agree but the idea was to emphasise its importance. However, as suggested it has been deleted.

-Page 623, line 9: The pitch, roll, and yaw need not be italicized and put in quote. A quantitative estimation of the uncertainty seems due, after what discussed in this section.

We had used italics and quote to highlight the role played by pitch, roll, and yaw effect. We agree that this is not necessary. We also agree that a quantitative estimation of the uncertainty is important. However, we believe it would be better to provide it in our future works given a large comprehensive dataset (DGPS data with recordings of pitch, roll, yaw) allowing us to investigate the pitch, roll and yaw effect on optical sensor setup changes in PHI and THETA. We have also additionally warned the reader

'However, it is important to highlight that the validity of R_{RS} and L_W is dependent on the accuracy of air-sea interface reflection coefficient estimate.'

- Page 629, Table 1. The caption is not clear. The number of images does not seem to be consistent with the numbers reported in the first paragraph of Section 3. The statistical sample is good enough to draw conclusions but two decimal digits are an overkill.

We have revised the statistics as suggested and checked the images with those reported.

Reponses to Referee 1

Garaba, Wernand and Zielinski : Quality control of automated hyperspectral …

This paper is concerned with improvements in processing automatically collected, near-surface measurements of ocean colour. The data presented emphasise the fact that rigorous quality control is required before useful data can be extracted from measurements of this type. Innovative aspects of the work include the use of a dual-field security camera to capture sea surface and sky images coincident with spectral acquisition and the development of a sun-glint ‘flag’ that makes use of intrinsic features of the acquired spectra. The data presented is certainly interesting, simply as a quantitative illustration of the difficulties encountered in near-surface marine radiometry. There appear to be a number of weaknesses, however, in the presentation of the work. The authors may care to respond to the following comments:
Thank you for the detailed review.

1. The use of camera images to divide the data set according to its degree of susceptibility to sun glint is a novel feature of the paper – it should be mentioned in the abstract.
We agree with this comment and this has been included in the abstract. However, the use of the camera system has been used in other field campaign and to avoid any conflict we avoid here to claim being the pioneers of the methodology. However, we have emphasized the use of the camera system to obtain sea-surface images.

2. Page 6 lines 5-7: how was the sea surface image assessment conducted: by visual inspection, or more objectively? This is important because the classification carried out at this stage affects the analysis of the rest of the paper. A related point is the question of how such a clean binary classification of the images was produced – were there no intermediate cases?
The images were manually assessed and therefore to mitigate the problem of subjectivity in selection it was decided to have only two sets.
‘Sea surface images were retrieved, matching the unmasked spectra validated with the meteorological flagging (Wernand, 2002), visually (manually) inspected and classified into; *Nns* – image set without sunglint or *Ns* – sunglint-affected image set,’
Two people performed this visual inspection with an additional referee. We understand that it is subjective but using an image analysis algorithm would still require the human interaction to classify the cases. We added more information in the manuscript to point this out to the reader.
The image inspection was performed by two investigators with an additional referee.

3. Page 6 lines 18-25: It appears that the presence of sun glint results in greatly enhanced L_w values: ‘at least 10 times higher’ according to the text. In that case, why was this information removed by normalisation before the development of a new sun glint flag was undertaken?
As explained in the response to Matteo Ottaviani we applied normalisation for the purposes of visual comparison whereby the scale range is restricted to 0-1 which therefore means the actual difference is identifiable but not to scale. In determining the new sun glint flag as summarised in Table 2 we used the actual values not the normalised value. We have also added this information to the manuscript.

4. Page 6 line 26 (and associated Figures): nL_w (normalised water leaving radiance) is a standard term in ocean colour radiometry. It is used to described water-leaving radiances that have been adjusted to compensate for atmospheric effects and solar angle. In this paper, nL_w appears to be used to designated spectra that have been normalised relative to their own maximum value. This is confusing, and an alternative term should be employed.
We agree that it can be confusing and we have changed it accordingly in the manuscript and figures. We also added the fact that we used the computed actual measurements of L_W and R_{RS} to determine the new flag.

5. Page 7 line 10: this reads as if the absorption of chromophoric dissolved organic material was confined to discrete bands rather than occurring as a continuum from the UV to the red end of the spectrum.
We agree and to avoid any confusion we have corrected it as follows, ‘chromophoric dissolved organic material absorption from UV to visible, here investigated at ($\lambda \approx 400$ nm) and ($\lambda \approx 460$ nm)’

6. Page 7 line s 19 to 23: The adjustment of thresholds to optimise their discriminatory performance for one set of data may mean that they perform sub-optimally for a different data set. The obvious way to guard against this is to test classification algorithms using data sets which are independent of the one for which they were developed. The absence of validation using independent data is a significant methodological weakness of this paper.
We agree that it is important to have independent data sets to further validate our proposed flagging method. However, we have to take into account the sensor setup (Instrument capability, Camera system, correction algorithms applied) which contributes to differences in test datasets we could use. In this study our aim was to present a method that can be adapted in upcoming application. Using the upcoming generated measurements from these works it will then be possible, due to the close uniformity of methodology, to assess the applicability of our method. We have added it here 'Optical measurements collected as proposed in this study, e.g. the planned permanent installation of this system on R/V Heincke, will provide test datasets for further validation of the meteorological and sunglint flagging for different water bodies.'

7. Page 8 lines 24-25: These lines may over-state the case. The fact that spectra free of sun glint were collected over a very wide range of angles does not necessarily mean that they were 'valid'. The validity would depend on whether the correction procedure embedded in equation (1) was effective over such a wide range of angles. The high NIR reflectance values in Figure 7 may indicate that some surface-reflected skylight is present in the 'valid' spectra.
We agree and have rephrased it to' According to the available measurements, valid spectra can be collected at $0° \leq \Phi \leq 360°$. However, the validity of the spectra is dependent on the accuracy of air-sea interface reflection coefficient estimate.

8. The legend for Figure 1 implies that for some stations, in situ measurements were made in support of the above-surface measurements. Why does the paper make no use of the in situ data for validation purposes?
The in-situ measurements collected were very limited and could not be used to aid the sunglint flag validation. However, to avoid confusion we have corrected the legend to Figure 1 accordingly. 'The blue line represents the track were optical measurements were collected, and the red line is for the return cruise to Bremerhaven without measurements.'

9. Page 9 lines 24-27: This criticism of the radiometric setup recommended by Mueller et al., and sold as a commercial product by at least one instrument manufacturer, is important. If the authors think their data backs it up, it could be given more prominence – maybe in the abstract?
Corrected the following statement was added to the abstract. 'It is confirmed that valid optical measurements can be performed $0° \leq \Phi \leq 360°$ although $90° \leq \Phi \leq 135°$ is recommended.'

Additional comments

1. Page 3 line 13: I don't think the 'Scotland Sea' appears on any charts – perhaps The Minch would be more appropriate?
Corrected.

2. Page 4 line 7: something has gone wrong with the units – there should be no Sr associated with Es measurements.
Corrected.

3. Page 5 line 17: the statement 'retain/unmask at least all spectra in Nns' implies that spectra in addition to those classified as Nns should be retained – presumably it needs to be re-phrased?
Rephrased.

4. Page 5 line 23 and page 7 line 4: 'rationing' should probably read 'ratioing'.
Corrected.

5. Page 9 line 17: would local surface cooling significantly affect reflectances in the 700nm-800nm wavelength range?
From the works of Marmirino and Smith (2005) and others referenced herein, they suggest that in the infrared spectrum thermal conditions contribute to the measurable light signal. We therefore pointed it out in our manuscript to show that the overall effect of foam and whitecaps is still to be fully investigated but more works suggest whitecaps and foam drive an enhanced signal in the NIR.

6. Page 10 line 3: ‘eluded’ should probably read ‘avoided’
Corrected.

7. Fig. 1 legend: ‘…track where…’
Corrected.

Interactive comment on “Quality control of automated hyperspectral remote sensing measurements from a seaborne platform” by S. P. Garaba et al.
Anonymous Referee #3

General Comments:
The article described four data quality flags used for automated and unmanned hyperspectral sensor that measures above-water spectrum. However, the emphasis was given to the sun glint mask, while the rest three meteorological flags based on solar irradiance were just briefly introduced. The measurement of sea surface and sky radiance spectrum is accompanied by simultaneous snapshots of sea surface and sky by a dual camera system, based on which subjective analysis was performed to classify the radiance data into two categories: data affected by sun-glint and data unaffected by sunglint. After that, numerous test has been performed using various spectral-minimum or spectral-mean remote sensing reflectance or water-leaving radiance values and band ratios as the threshold to objectively classify the data. The previous subjective analysis results are then used as truth to evaluate the performance of the objective analysis. Among all the test performed, the authors found that the best results came from using spectrum-mean water leaving radiance of 2 mWmˆ-2nmˆ-1 or spectrum-minimum remote sensing reflectance of 0.01 Srˆ-1 as the threshold. The authors also concluded that valid optical measurements can be performed at any solar-sensor azimuth angle.
The title of the manuscript seems to be a little inappropriate, since it is focused on the sun glint mask. The three meteorological flags were just briefly introduced, to which no effort has been given and the authors seemed to just follow the previous studies. I will think an appropriate title will be something like "Sun-glint masking for an automated seaborn hyper-spectral remote sensing platform". The work is interesting and unique in the way the authors used cameras and performed subjective analysis to form the basis or truth for the various objective sun-glint masking tests. The resulting thresholds can be of certain value for similar studies. However, I think some specific aspects need to be addressed satisfactorily before this article can be published in OS.
We thank you for the comments and contributions which we followed to improve our manuscript. As suggested the title of the manuscript is now
‘Empirical sunglint masking for an automated and unmanned seaborne hyperspectral remote sensing platform.’

Specific Comments:
1. for the subjective analysis, no detail information has been given. It is a crucial process since it is a basis for all the objective analysis. Without detailed info on the criteria to classify the with-sun-glint and without-sun-glint data, it is impossible to confirm the reader the result from this test as truth data is justified.
We have revised the methods section so as to make it clear how the sunglint flag is generated and we also added theory behind the sunglint flag.
‘The empirical sunglint flag was developed on the premise that open seawater is assumed to absorb all light in the NIR. Thus any light signal measured by an optical sensor would be sea surface reflectance or atmospheric scattered radiance. However, in turbid waters multiple scattering influenced by known as well as unknown optically active seawater constituents also contributes to this light signal measurable by an optical sensor (Doxaran et al., 2007; Jamet et al., 2011). In Fig. 3 a simplified activity diagram illustrates the steps that were manually executed and implemented in this sunglint flag investigation and evaluation; ’
Furthermore, the image analysis was done manually for all the snapshots and quality control was performed. We have to concede that even with more investigators; the subjective nature of the process will always give a probability of error.
‘The image inspection was performed by two investigators with an additional referee.’

2. Sometimes effect of sun-glint is not apparent by simply looking. There is a gray area of sea surface between sun-glint-affected and sun-glint-free. The data from this area is contaminated. Is these kind of data subjectively classified as with-sun-glint also? If so how is it identified?
In this paper we simply split the measurements into two sun-glint-affected and sun-glint-free. In future our aim is to automatically scan the images with a dedicated software e.g. MATLAB but this will also be subjective. To address the gray area measurements we assumed our flag conditions to help obtain the best possible 'useful' measurements although we accept the possibility of having some sunglint-affected data unmasked.

3. about the threshold for the sun-glint mask. It is a little bit confusing to me. Is the data affected by sun-glint have higher-than-normal radiance due to sun-glint from water leaving radiance, or lower-than-normal due to subtraction of very high sky-radiance from sun? It had better be clarified in the manuscript.
We have added a statement to better clarify this,
'For example, when using 'mean $(L_W)_{NIR}$ < 2 mW/ m² nm Sr', any measurement satisfying this condition will be classified as sunglint-free. '
It means the sunglint affected water leaving radiance has a higher than normal water leaving radiance. This is based on the fact;
'The empirical sunglint flag was developed on the premise that open seawater is assumed to absorb all light in the NIR. Thus any light signal measured by an optical sensor would be sea surface reflectance or atmospheric scattered radiance. However, in turbid waters scattering influenced by optically active seawater constituents also contributes to this light signal measurable by an optical sensor (Jamet et al., 2011).'

4. about data normalization in Fig.5 and Fig.6. I don't believe normalization by maximum is justifiable since it lacks physics meaning. Normalization is generally used to remove effects that you don't want to include. For L_w, you can normalize it by Es (extraterrestrial solar irradiance) to remove effect of difference in solar irradiance. But that's just R_rs. R_rs itself should not be further normalized.
We agree that in ocean colour radiometry normalisation uses E_S and is termed nL_W but here we used a different terminology since we perform normalisation simply to show the differences in spectra. 'Data normalisation was applied maintaining the spectral shape while simplifying comparison of spectral, dividing each L_W and R_{RS} measurement by the maximum value for each measurement.' As explained in response to Matteo Ottaviani's comments.

5. about the conclusion that valid optical measurements can be performed at any solar-sensor azimuth angle. As I just mentioned in point 2, there may be lots of data in "no sun-glint" group that is actually contaminated by sun-glint. Also what is the error introduced when the ship "pitch, row and yaw"?
We agree and have addressed this point earlier above. A quantitative estimation of the uncertainty is important. However, we believe it would be better to provide it in our future works given a large comprehensive dataset (DGPS data with recordings of pitch, roll, yaw) allowing us to investigate the pitch, roll and yaw effect on optical sensor setup changes in PHI and THETA. We have also additionally warned the reader 'However, it is important to highlight that the validity of R_{RS} and L_W is dependent on the accuracy of air-sea interface reflection coefficient estimate.'

Interactive comment on Ocean Sci. Discuss., 8, 613, 2011.
Review of "Quality control of automated hyperspectral remote sensing measurements from a seaborne platform", by S.P. Garaba and co-authors (ms # osd-8-613)

General appraisal
The paper by Garaba and co-authors presents a set of underway measurements of radiometric quantities from an unmanned platform equipped with radiometers. These radiometers are installed at the bow of a research vessel and they measure the sky radiance, the radiance reflected by the sea surface plus the water-leaving contribution, and the downward irradiance above the surface. This is the usual setup to determine the water-leaving radiance, Lw, from above-water measurements. The goal of the study is to evaluate how sunglint contaminated measurements can be detected and eliminated (because they are inappropriate for a proper derivation of Lw). A new flag is proposed for that purpose.
Thank you for the suggestions and comments to help improve our manuscript. Our instrument setup is the usual setup and it also presents a novel method with a camera system recording sea surface and sky conditions at the same interval as spectra measurements. It therefore allows us to provide an empirical sunglint flag which is subjective when taking into account the manual image analysis for sunglint by three investigators in the present study.

There are several issues with this work:
1) The problem of contamination by sun glint is totally mixed with the issue of non-zero water-leaving signal in the near infrared. Therefore, I don't see any way to confirm that data of good quality are used in this analysis (contrary to what is said page 623).
Indeed 'good quality' is subjective taking into account that the correction factor $\rho_{air\text{-}sea}$ is not constant as shown in other works e.g. Mobley, 1999; Ruddick et al. 2006. In investigating sunglint we have to take heed of the black pixel theory as proposed by Siegel et al. (2000) as shown in a review on sunglint correction algorithms (Kay et al. 2009) The black pixel agrees well with the hypothesis that seawater absorbs all light in the NIR which is not true for optically complex waters (our study site) as explained here
'The empirical sunglint flag was developed on the premise that open seawater is assumed to absorb all light in the NIR. Thus any light signal measured by an optical sensor would be sea surface reflectance or atmospheric scattered radiance. However, in turbid waters multiple scattering influenced by known as well as unknown optically active seawater constituents also contributes to this light signal measurable by an optical sensor (Doxaran et al., 2007; Jamet et al., 2011).'

2) There is no objective qualification of the quality of this flag. The fact that sunglint contaminated measurements are identified does not mean that the rest of the data set is of good quality. The conclusion that good measurements can be taken for any azimuth angle is actually not supported at all (and it is in strong contradiction with admitted protocols. Why not, but this should be supported by a much stronger evidence).
We have changed the title of the manuscript so as suit the objective of the sunglint flag. In out introduction we also justify the purpose of this study i.e. quality control,
However, automated and unmanned optical measurements from any of these platforms are likely to be contaminated due to meteorological conditions (e.g. rainfall, cloud cover, humidity, and dusk/dawn conditions), sunglint, and sensor setup (Gordon and Jacobs, 1977; Wernand, 2002; Zhang and Wang, 2010). Therefore quality control to mitigate these disturbing factors is a crucial procedure in determining colour of seawater.
We obtain valid measurements but with a warning we tell the reader that they accuracy and validity of R_RS is dependent on the air-sea interface reflection coefficient $\rho_{air\text{-}sea}$ as well as other sources of error not fully understood at present.
'In this study it has been shown, with caveats, that valid measurements can be obtained for $0° \leq \Phi \leq 360°$ from an unmanned seaborne platform. There is some probability that the assumed valid measurements can be erroneous, as it is widely accepted that the validity and accuracy of R_{RS} or L_W is also influenced by the air-sea interface reflection coefficient $\rho_{air\text{-}sea}$. More work is needed to quantify such a probability among other sources of error e.g. whitecaps, foam, optically active seawater constituents contributing to sea surface reflected light signal.'

3) Overall the writing is approximate (language and manuscript organization). It is difficult to extract important information from the reading. I don't think the material presented in this manuscript is good enough for a publication in OSD. This is, at least for the moment and with the information provided,

rather like a technical note. Objectives should be reformulated and the method qualified in a more objective way (for instance through comparison with independent estimates of the water leaving radiance).
We have intensively revised the manuscript with the help of native speakers and believe the new version along with changes as suggested above by M.Ottaviani and the other referees make our manuscript a useful methodology article. Equipped with a novel security camera system, hyperspectral instruments, and empirical approach to investigate sunglint. Given that it is an empirical approach it is complicated to compare with an independent data set as changes will be necessary to the set conditions.
Several works concede that as much as we want generic approaches to remote sensing, 'optically complex' or turbid waters pose a major challenge to research (Dierssen, 2010; Siegel et al. , 2000). Using an independent data set will produce a different set of conditions valid for the sampled time-space period. It is thus key to our research that we postulate that it is a methodological article and thus output cannot be similar.

A few detailed comments
• The use of "mask" as a verb is possible but I'm not sure it is really appropriate here.
We have used the term mask to be consistent with ocean colour terminology.

• Page 614, lines 5-20: it is confusing to merge here the sunglint issue and the black pixel assumption. These two issues should be considered separately.
We address this point above on point 1). It is difficult to separate sunglint from the black pixel assumption because; in the open waters the sea is expected to absorb all light and thus any signal received by optical sensors will be from sea surface reflectance (sunglint) and atmospheric aerosol radiance (AAR). This would be an ideal scenario which is unfortunately not completely true due to the fact that even in the open waters as well as turbid waters we have known and unknown IOPs also contributing to this reflected signal along with AAR. However, it is assumed that for above-water measurements assume that AAR is insignificant leaving known and unknown IOPs influencing the reflected signal along with sunglint.

• Page 614, line 7: "suggest" is not the right term here. Sunglint is caused by reflection at the surface.
We agree and have changed it. 'It is caused by Fresnel reflection from a number of 'dancing facets…'

• Page 616, line 3: "compliment" should be "complement"
Corrected.

• Table 2: The sign in the equation for Flag 2 should be "<" instead of ">"
Corrected.

• Fig. 3 and elsewhere: replace "with no sunglint" by "without sunglint"
Corrected.

• Figs 5 and 6: "normLw" on vertical axes should be replaced by "Normalized Lw" (same for "normRrs")
"Normalized Lw" is 'as a standard term in ocean colour radiometry used to described water-leaving radiances that have been adjusted to compensate for atmospheric effects and solar angle.' Therefore we decided to use 'normL_W' to avoid confusion as we applied normalisation to spectra based on its own maximum value for purposes of better visual comparison of differences in spectra.

• All figures but figure 7 (and images) could be white and black.
We applied colour figures for readability of the manuscript.