os-2015-31 Authors' replies to the comments from G. Zibordi (Referee #1) on "Estimation of upward radiances and reflectances at the surface of the sea from above-surface measurements"
by Ø. Kleiv et al.

The comments from our reviewer are in bold italics, and our replies are in normal font.

I.
The manuscript presents and discusses the application of a new above-water method for the determination of in-situ water leaving radiances (L_w) or equivalent quantity, to support the validation of satellite ocean color radiometry data.
It is not entirely correct that the purpose of the described method is to support the validation of satellite radiometry data. We state in the Abstract that the method can serve as "a first check of satellite products in coastal waters". We are fully aware that the accuracy of the method does not fulfil the requirements set by e.g. NASA for validation of radiometric quantities, and unfortunately the term "validation of satellite products" can be misunderstood. By "a first check" we mean to see if the difference between the satellite product and the result of the field method is within the range of the estimated error for the method. It has been our experience that radiometric values produced by MERIS could differ from field values by a factor of more than 2 in the Oslofjord-Skagerrak area, and accordingly " a first check" by a simple method can then be useful.

We also have to point out that the possibility of checking satellite products is only half of our motivation for the present study. Section 1 describes the need for monitoring Norwegian coastal waters by satellite remote sensing as well as by ship-mounted sensors on ships of opportunity. It should be evident from our paper that the possibility of improving the latter method has been another motivation. We have now emphasized this point in Sect. 1.

II
The manuscript addresses the problem with a limited data set (22 series of measurements) performed in challenging illumination conditions. Results indicate that the methodology can lead to large uncertainties, which the authors still consider adequate for the validation of satellite ocean color radiometry data.
We agree with our reviewer that we have a limited data set and challenging illumination conditions. However, these conditions are quite normal at northern latitudes, while the perfect situation with a completely clear sky is very rare. Based on observations by the Norwegian Meteorological Institute from the last 10 years, the average number of days with a cloudiness of 0-1 okta at 12:00 UTC in Oslo during the 61 days of May and June is 4.5. If a cloudiness of 0 is required, the average number of days is reduced to 0.3. This means that on an average we would need 3-4 years of observations in May and June to obtain 1 day with a completely clear sky at noon. The average cloudiness at 12:00 UTC in these months, based on the 10-years series, is 5.4 oktas. Thus Table 1 demonstrates that on 3 of the 4 field days the cloudiness was less than the average conditions. This information has been added to Sect. 2.1, where Table 1 is presented.

We have described a method where the estimated reflected radiance may have relative rms errors equal to or less than 15 % in the spectral range 351-754 nm (Table 4), and where the errors of the water-leaving radiance at 560 nm may be equal to or less than 24 % (Table 5). As mention in Point I above, we do not claim anywhere that these uncertainties are adequate for validation of satellite radiometry data, but we have stated that we find the uncertainties acceptable "for a first check of satellite products in coastal waters". We have now changed "coastal waters" in the last line of the Abstract to "the Inner Oslofjord", to indicate that results may be different at other latitudes with less cloudiness and in other areas with different sea waters.

III

While considering of general interest the proposed method, I believe that the work is limited in terms of supporting data and analysis (see the extensive comments below). Additionally, the manuscript conclusions are often based on personal feelings and not science/technical requirements. I cannot recommend the manuscript for publication in Ocean Science.

We do not agree that any of our conclusions are based on personal feelings. Our only personal feeling in this context is a positive surprise that the assumption of similarity in the spectral shape of radiance reflectance is able to produce estimates with deviations from the measured field values less than or equal to 13 % throughout the spectrum. To our knowledge no other published method is able to obtain similar results under the same cloudiness conditions in such a simple way.

However, we appreciate that our reviewer has been willing to spend time on commenting the different details. This gives us the chance to clear up misunderstandings and improve the text.

Main concerns

1. In situ data set. The data applied for the analysis were collected under challenging observation conditions (i.e., mostly cloudy sky). Validation activities for satellite ocean color radiometry data must rely on in situ data collected during clear sky conditions. Because of this some of the proposed parametrizations may not apply to the clear sky that necessarily characterizes operational validation measurements.

Our results were obtained for a cloudiness range of 1-8 oktas. Unfortunately the term "clear sky" is defined in different ways by different organizations:

- NASA describes a clear sky as a cloud cover less than 10 %.
- CIE describes a clear sky as less than 30 % cloud cover.
- WMO defines a cloudiness of 0-2 oktas as Fine, 3-5 oktas as Partly Cloudy, 6-7 oktas as Cloudy and 8 oktas as Overcast. Cloudiness 0 is described as a "completely clear sky".

We agree with our referee that results obtained under perfect, cloud-free conditions may be different from the results presented in our paper. However, a perfect method requiring perfect conditions can seldom be applied in our areas, as was explained in Point II above.

2. Proposed measurement method. The method relies on the nadir-viewing geometry that may be largely affected by sun-glint during clear sky conditions (i.e., the observation conditions relevant for the validation of satellite radiometry data). This cannot be simply ignored because sun-glint perturbations in above-water radiometry, which statistically depend on sensor and illumination-viewing geometries, may become the source of unpredictable uncertainties.

We agree that at latitudes where the sun is close to zenith, sun glints from nadir may be a severe problem for recordings of nadir radiance. However, along the Norwegian coast the latitudes vary from 58° to 71°N, implying that the range of the minimum solar zenith angle will be 35°-48°. It is our experience from the Oslofjord that recordings in air of radiance from nadir do not exhibit more variations than recordings at tilted angles, rather the contrary. As mentioned in Sect. 2.1, we are using a median filter to avoid signal spikes.

Relevant comments

a. Page 1053, line 17. When considering a nadir-viewing geometry for the above-water radiance sensor, the measurement problem is certainly simplified from the analytical point of view. But, as indicated in the concerns, it enormously increases the difficulties to produce accurate data due to sun-glint perturbations during ideal measurement conditions (i.e., clear sky).

Our reply is the same as in Point 2 above.

b. Page 1053, line 20. The text is ambiguous. The manuscript aims at proposing a method applicable to nadir-view observations and indicates the need to later address the application of the method to a non-nadir viewing geometry. This should be made clearer to avoid that future developments are intended for any above-water method that makes use of non-nadir observations.

We cannot find any ambiguity in the text. We say that "The next step will then be to relate these results to recordings by sensors tilted at an angle from the nadir....". The last paragraph of Section 4 has a similar comment. Still, in order to avoid any misunderstandings we have now added a line to page 1053 explaining that the next step is not described in this paper.

c. Page 1054, line 11. The meaning of "were checked at the start" referred to the performance of the radiometers, is too general.

We agree. Accordingly we have now explained in line 11-12 that the sensors were tested against the calibration device FieldCAL from TriOS at the beginning of each cruise. (The same sensors were also compared with similar sensors from other institutions in the EU project HighROC, in 2014 and 2015, with satisfactory results.)

d. Page 1054, line 23. The fixed depth measurements (i.e., 0.5, 1.0, 1.5, . . . m) are difficult to perform from a ship which is naturally rolling due to wave effects. This has an impact on the uncertainties of extrapolated subsurface values.

We agree that waves will influence the recordings, although they did not produce any observed rolling of the research vessel during our work. The recordings lasted for 60 seconds at each depth, and a median filter was later on used to cancel out wave effects and spikes. We have added a line explaining this.

e. Page 1054, line 26. As already anticipated in the concerns, data collected under cloudy conditions may not be suitable for the development and assessment of a method expected to support investigations in clear sky conditions.

We agree, but it is explained in the points above that our method is meant for use under realistic conditions in our coastal areas.

f. Page 1055, line 7. The averaging of measurements, when referred to the above- water downward irradiance, smooths the effects of ship roll and pitch. A well performing radiometer should not produce appreciable "electronic spikes".

We agree. The word "electronic" has been omitted.

g. Page 1055, line 22. Above-water radiance measurements which were performed at different time during cloudy conditions, are likely to be affected by illumination conditions changing over time. This may have largely affected the accuracy (or even the reliability) of the water-leaving radiance ratios given in the manuscript for the nadir to 40 degrees viewing geometries.

We agree. But Doxaran et al. (2004) did not normalize their L_{ua} values against E_{da}. We have moved this paragraph to Section 1.

h. Page 1055, line 24. It is not clear which is the "attempt" that did not lead to satisfactory results.

Some recordings of L_{ua} were made at tilted angles, but it was later discovered that the instrumental set up was not satisfactory. It may be better not to mention this attempt at all, and we have omitted this line.

i. Page 1056, line 4. The diffuse attenuation coefficient is only constant when the extrapolation water layer exhibits homogenous optical properties.

We agree. We have now added a line saying that the upper 3 meters were well mixed.

j. Page 1056, line 21. The indication that deploying the in water radiometer behind the stern of the ship may lead to a reduction of the signal up to 20% is a too general statement. An inappropriate deployment of radiometers may lead to errors much larger than 20%.

We agree. In the referred investigation by Korsbø and Aas the instrument was on the sunlit side of the ship. Their recordings necessarily included both the ship shading effect as well as the influence of reflected light from the superstructure of the ship. The text has been changed to explain this.

k. Page 1057, line 6. "Personal feelings" should not have much space in a methodology paper. From the manuscript it is clear that there was an attempt to avoid ship shading effects (not to determine their effects) by operating the in-water radiometers at some distance from the superstructure. This implicitly may mean that the authors also assume that the reflection effects of the superstructure are negligible. This is completely fine. But "assuming something negligible" and "thinking that something is negligible" are quite different statements.

We disagree that "We think..." implies personal feelings. When we write: "In our case we think that this reflectance is included in the determination of the ship shading effect", it is a conclusion based on the findings of Korsbø and Aas, referred to in the previous paragraph. We have rewritten this and the previous paragraph.

l. Page 1057, line 22. The quantity "Br" does not appear declared.

True. *Br* should have been in italics, and it should have been mentioned that *Br* is the product of *B* and *r* defined in the previous paragraph. The text has been changed accordingly.

m. Page 1059, line 8. Immersion factors for different series of TriOS radiometers characterized by different optical windows (i.e., different values of n_g) were documented by Zibordi and Dareki (2006). Those factors provide some sensitivity to equation 9 (actually, not essential for the manuscript).

The immersion factors found by Zibordi and Darecki for the Ramses ARC radiometers differ less than 0.01 from those found by Ohde and Siegel for the range 400-700 nm. We have added the reference to Zibordi and Darecki.

n. Page 1060, line 17. The quantity R_r has spectral dependence that varies with the illumination conditions mostly as a function of E_{da}. Because of this, the fit of data taken under very different illumination conditions may be affected by a large variance. This suggests that fits depending on specific illumination conditions (i.e., overcast and clear sky) may much better support the application of the proposed method.

We agree that data sampled under either a clear sky or overcast conditions might produce better results. However, we have pointed out that the completely clear sky is very rare in our coastal areas, and while the case of the overcast sky may be useful for the monitoring by ship-mounted sensors, it is no good for optical satellite remote sensing. Consequently we are still of the opinion that our proposed method, based on no-perfect conditions, will be more useful in our areas than methods requiring either a clear or overcast sky.

o. Page 1061, line 10. Probably it would be better to state that the spectral shape of R_w in the near-infrared is invariant, and not simply "constant".
Accepted. The word has been changed.

p. Page 1062, line 19. The values indicated as uncertainties for L_w, only refer to the uncertainties in the determination of the sub-surface values through the extrapolation process still ignoring uncertainties in depth values. A comprehensive determination of L_w uncertainties would require much more inputs. Because of this, the estimated uncertainty for L_{ua} is not at all supported by evidence.
Here we disagree. We have a set of $\ln(L_{\mathrm{uw}})$ attributed to different depths z. The depths are based on the length of a wire running over a meter wheel, and at each depth the recording periods (60 seconds) were chosen so as to average out the effects of waves (and eventual ship roll). The meter wheel is adjusted to zero when the radiance sensor passes the sea surface. We are thus confident that the accuracy of z is better than 10 cm, probably in most of our cases better than 5 cm. When we perform a correlation analysis between the two variables, the analysis does not distinguish between errors in depth or signal, but it minimizes the squared distances between $\ln(L_{\mathrm{uw}})$ and the correlation line. Accordingly we conclude that our estimates of the relative uncertainty of $L_{\mathrm{w}}(0)$ in Table 2 are valid.

As for the estimated uncertainty of L_{ua} in the central part of the spectrum there may be a misunderstanding. This estimate is not influenced by the uncertainty of L_{w}, but is based on the information from the TriOS company (quoted on page 1062, line 9), the magnitude of the signals in the different parts of the spectrum, and the quality of the field recordings, expressed by the difference between mean and median values (page 1055, line 8-12). We have rewritten the text at this point.

q. Page 1063, line 22. The statement that 13% (maximum) uncertainty is still satisfactory for a first check of satellite products is just a very personal judgment that may mislead the scientific community. Uncertainties of in situ data applicable for validation purposes should reflect mission objectives/requirements and not personal believes.
It is quite correct that it is a personal judgement that errors less than 13 % are satisfactory for results based on radiometric recordings made from ships during field cruises. This judgement is based on several years of experience. Because the reader is informed on page 1065, line 4, that NASA requires an accuracy of 5 % for ground truth measurements, the scientific community should not be mislead.

r. Page 1064, line 3. The low rms determined at 351 and 754 nm for R_r are simply due to the fact that the values of R_r at 351 and at 754 nm are hinge points in the fitting procedure.
The smaller rms deviations of R_{r} at 351 and 754 nm are not due to the procedure, but to the physical fact that at these wavelengths the water-leaving radiance in our waters becomes so small that the recorded value of L_{ua} comes close to L_{r}, or R_{ua} close to R_{r}. We have added this explanation.