We thank Reviewer #3 for the useful comments. The point by point response is provided below in blue font, and the proposed text modifications in red.

This paper quantifies the skill of the OASIM radiative transfer model in the Mediterranean Sea region. Although there is no (cited) comparison of the skill of other atmospheric optical models, the OASIM performance is impressive, and would be both good news and useful information for the groups using, or considering using, OASIM.
Another interesting feature of the manuscript is the use of the Bio-Argo array for model assessment. As this will no doubt become the primary means of assessing large-scale marine atmospheric models in the future, the technique developed in the manuscript will be interesting to others.
While there are lots of excellent figures, the manuscript itself is quite short and does not make as much of the analysis as it could – leaving some room for additions, which would help to make the manuscript more useful and citable.

A couple of ideas for strengthening the manuscript:
1. The OASIM is assess primarily based on Ed(0-), the downwelling planar irradiance. It is mentioned that the BOUSSOLE observations have above water observations. Analysis of skill of Ed(0+) would help to distinguish between OASIM errors in the atmosphere and in the transmission across the sea-surface interface. This distinction would be useful for both improving OASIM, and also for those using OASIM.

As suggested by the reviewer, we show in Fig. R1 the multispectral downward planar irradiance (**Ed**(**λ**,**0**+)) simulated by OASIM (blue lines) and measured at BOUSSOLE (red lines) for the month of March. Comparing Fig. R1 with Fig. 6 proposed in the submitted version for **Ed**(**λ**,**0**-), we show that differences, related to the atmospheric parameters controlling seas surface reflection, have a low impact on the results. In fact otherwise we would observe a much higher deterioration in model performance when computing **Ed**(**λ**,**0**-) from **Ed**(**λ**,**0**+). In order to provide a general overview of the differences from Ed(**λ**,**0**+) from model and from BOUSSOLE we reported the skill in Table R1 in analogy to the Table 1 that we proposed in the manuscript. The differences in the skill (e.g. RMSD and BIAS percentual scores) between computation of **Ed**(**λ**,**0**-) from **Ed**(**λ**,**0**+) indicates that the model is, in most of the cases, slightly better in computing **Ed**(**λ**,**0**+) than **Ed**(**λ**,**0**-) but the differences are in any case of second order so is the impact of surface pressure and wind.
We propose to include the following text to the section 3.4 "Summary of the OASIM model skills in the Mediterranean Sea":

> Similar summary analysis performed for **Ed**(**λ**,**0**-) was also performed for **Ed**(**λ**,**0**+), to estimate the impact on reflection processes at sea atmosphere interface on irradiance (not shown). These processes are regulated by atmospheric parameters shown in Fig. 4. Percentual skill metrics indicate that RMSD is only marginally affected, with differences lower than <1%, while BIAS for **Ed**(**λ**,**0**+) shows differences lower than 5% with respect to **Ed**(**λ**,**0**-).

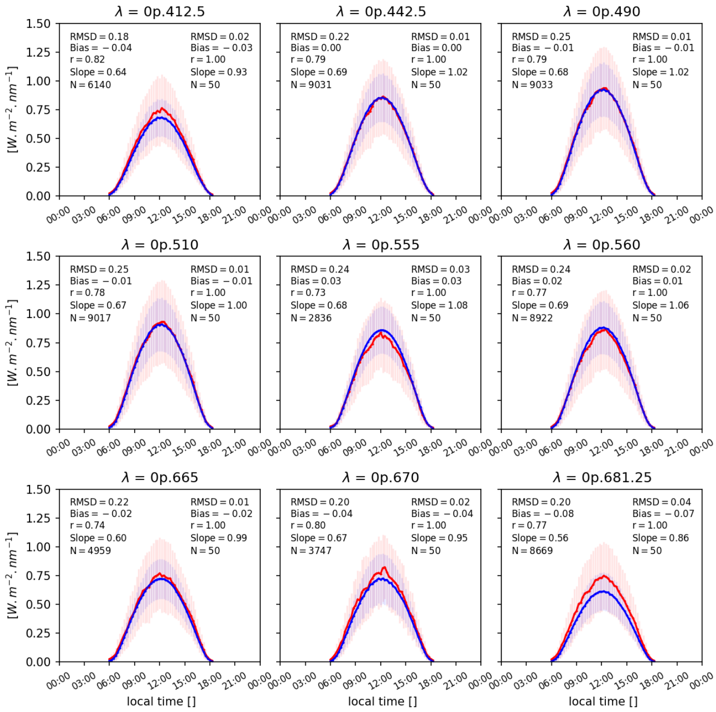

Figure R1. Multispectral downward planar irradiance **Ed(λ,0+)** simulated by OASIM (blue lines) and measured at BOUSSOLE (red lines). The wavelengths considered are those measured by the BOUSSOLE sensors for the average March data derived from the time series. For each panel, the reported statistics (RMSD, Bias, r, and regression slope) are related both to the high-frequency signal (with a temporal resolution of 15'; top left) and to the average day in the considered month (top right). The vertical bars indicate the variance in the monthly averaged values of the average day.

Table R1. Summary of the model skill compared to the available data from the BOUSSOLE buoy (from 2004 to 2012) and BGC-Argo floats (from 2012 to 2017) for the irradiance (**Ed(λ,0+)**) at the different wavelengths (WL) and for DPAR. RMSD, bias, and Y-int are expressed in W m-2 nm-1, while all the other indicators (regression R and slope) are dimensionless, where N is the number of match-ups between the model and observations. For the BOUSSOLE comparison, the green numbers are derived by filtering out the day-to-day variability (i.e., the intra-monthly variability). Given the large number of samples, all statistics are significant (p-value < 0.05). For the RMSD and BIAS, the percentage values normalized by average data are reported in parentheses.

BOUSSOLE vs OASIM-ECMWF [2004-2012]						
WL	**RMSD**	**BIAS**	**R**	**SLOPE**	**Y-int**	**N**
412.5	0.15 (34.1%)	-0.04 (-9.5%)	0.83	0.66	0.08	55239
	0.04 (10.0%)	-0.04 (-9.6%)	0.99	0.88	0.00	
442.5	0.18 (33.6%)	0.00 (0.6%)	0.84	0.77	0.09	111010
	0.04 (7.2%)	0.00 (0.5%)	0.99	1.00	-0.01	
490	0.20 (34.4%)	0.00 (-0.1%)	0.84	0.76	0.10	112186
	0.04 (7.4%)	0.00 (-0.2%)	0.99	1.00	-0.02	
510	0.20 (34.6%)	-0.01 (-2.0%)	0.83	0.74	0.10	112071
	0.04 (7.2%)	-0.01 (-2.1%)	0.99	0.98	-0.02	
555	0.20 (33.4%)	0.03 (5.1%)	0.85	0.83	0.10	55309
	0.05 (8.6%)	0.03 (5.0%)	0.99	1.05	-0.03	
560	0.20 (35.5%)	0.01 (2.3%)	0.83	0.76	0.11	106660
	0.04 (7.9%)	0.01 (2.3%)	0.99	1.02	-0.02	
665	0.18 (34.1%)	-0.02 (-3.0%)	0.84	0.75	0.09	76247
	0.04 (7.1%)	-0.02 (-3.1%)	0.99	0.99	-0.03	
670	0.17 (39.6%)	-0.04 (-9.3%)	0.79	0.63	0.08	32733
	0.04 (10.1%)	-0.04 (-9.5%)	0.98	0.92	-0.02	
681.25	0.17 (36.4%)	-0.08 (-16.4%)	0.81	0.62	0.07	110418
	0.05 (10.3%)	-0.07 (-16.6%)	0.99	0.85	-0.02	

2. Following on from 1., the article distinguishes between errors in cloud-free and cloudy days. I wonder whether a somewhat similar, but more diagnostically-useful distinction might be the fractions of direct and diffuse radiation. As per point 1 this might be more useful for OASIM developers and also for those attempting to apply these results outside the Mediterranean where lower sun angles might change the balance of direct vs diffuse for the same cloud.

Thanks for the comment. We introduced the following indicator:

$$\mathbf{IND} = \mathbf{E_{dir}(\lambda,0^-)} / (\mathbf{E_{dir}(\lambda,0^-)} + \mathbf{E_{dif}(\lambda,0^-)}) * 100 \quad \text{(R1)}$$

IND varies in the interval [0,100]. **IND** is 0 when **Ed(λ,0-)** = **Edif(λ,0-)** and it is 100 when **Ed(λ,0-)** = **Edir(λ,0-)**. **IND** =50 indicates perfect balance: **Edir(λ,0-)** = **Edif(λ,0-)**.

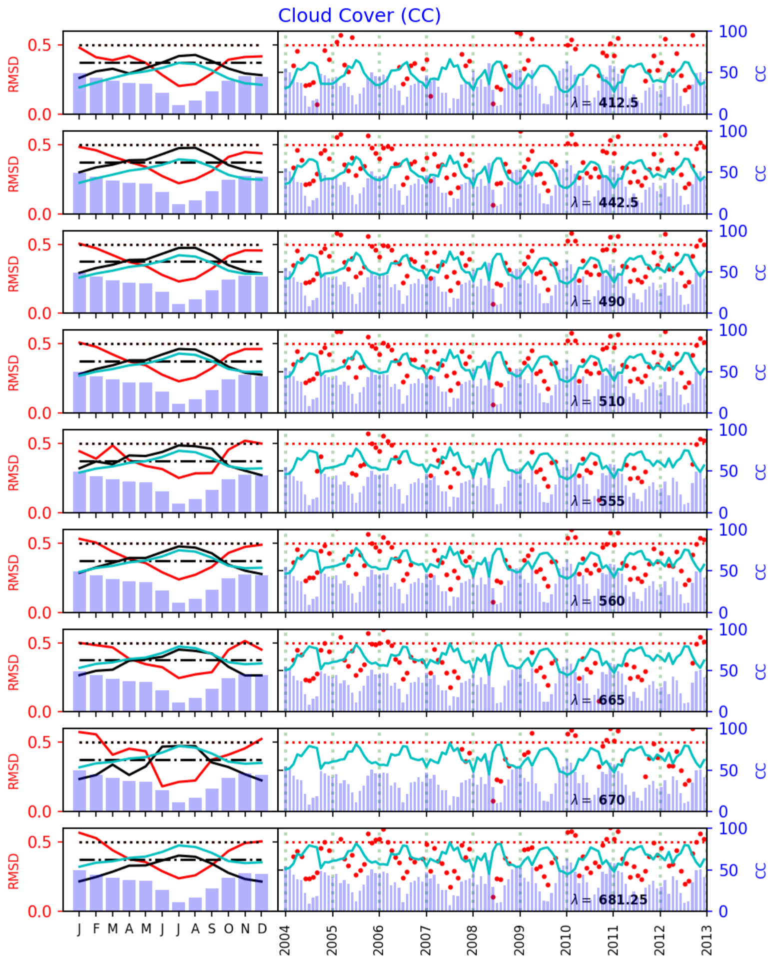

Figure R2. Comparison of the OASIM downward irradiance (Ed(λ,0-) at the nine wavelengths) to the BOUSSOLE data from 2004 to 2012 in terms of the RMSD and regression slope and their relationship with the ECMWF ERA-Interim cloud cover (CC). The left section of each panel shows the monthly climatology of the RMSD (normalized by its averaged value; red lines and labels) and regression slope (normalized by its averaged value; black line), superimposed on the monthly climatology of the cloud cover (in %, blue bars and labels). Regression slope thresholds at 1 (dotted black line) and 0.75 (dot-dashed black line) are also shown. The right section of each panel shows the monthly means of the time series of the RMSD (red dots, with a 0.5 value; red dotted line), superimposed on the monthly means of the time series of the cloud cover (blue bars and labels). The IND parameter defined in equation 3 is also reported (cyan lines), and in all panels varies in the range [0,100]

We complemented Figure R2 (corresponding to Figure 14 in the manuscript) including the indicator **IND** in cyan defined above. The indicator shows higher values during summer, when the direct component is dominant, corresponding to better model performances. On the contrary, during winter lower values of **IND** are found, with the diffuse component dominating, corresponding to the worser model performances. Similar results are observed considering the data corresponding to the Mediterranean Sea region as reported in Fig. R3.

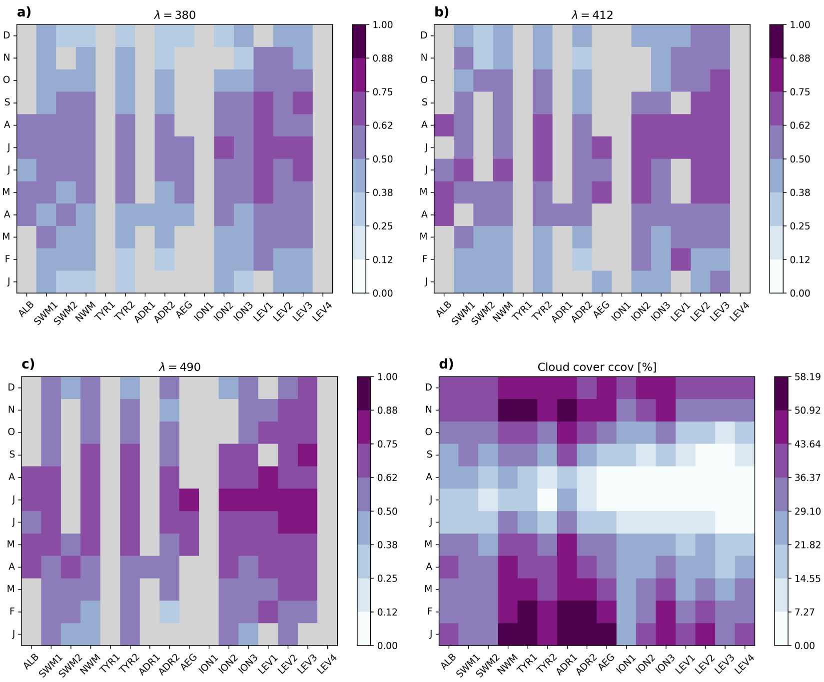

Figure R3. Evaluation of the IND indicator normalized to 1, for the three wavelengths measured by BGC Argo floats panel a) **λ**=380 nm, panel b) **λ** = 412 nm, panel c) **λ** = 490 nm. Panel d) shows the cloud cover.

We propose to add the following text to the discussion section:

Beside cloud cover, we introduced a further diagnostic (IND) based on the fraction of direct and diffuse irradiance components.

This diagnostic is defined as:

$$\mathbf{IND = Edir(\lambda,0\text{-})/(Edir(\lambda,0\text{-})+Edif(\lambda,0\text{-}))}*100 \quad (3)$$

where IND is a-dimensional and varies in the interval [0,100]. Recalling that **Ed(λ,0-)** = **Edir(λ,0-)** + **Edif(λ,0-)**, **I** is 0 when **Ed(λ,0-)** = **Edif(λ,0-)** and it is 100 when **Ed(λ,0-)** = **Edir(λ,0-)**. **I** =50 indicates perfect balance: **Edir(λ,0-)** = **Edif(λ,0-)**. As show in Fig. 14, IND provides similar interpretation of cloud cover in fact the model skill is higher when IND is higher and vice-versa. This diagnostic indicator could be used to generalize results in regions outside the Mediterranean Sea where lower sun angles are found implying a different balance of the direct versus the diffuse component. In these situations the effect of clouds in increasing RMSD and bias could be even higher. It is worth to mention that in the present case, since IND covariates with cloud cover, it is difficult to separate the role of clouds from direct versus diffuse irradiance ratio. Nonetheless, IND at 412 nm is lower than all the other wavelengths and this could explain, at least in part, the lower skill observed at 412 nm.

3. The extrapolation of the Bio-Argo data to the surface sounds like a critical step. Can you give more details? For example, do you assume an exponential decay with depth?

Thanks for pointing this out, we propose to rephrase with the following text:

Before comparing model values to observations, the irradiance profiles obtained from floats were extrapolated to the surface with an exponential fitting procedure based on the curve_fit tool of the python package scipy. Further, we required profiles to have at least one measurement in the first 1.5 m depth from sea surface and to have at least 4 measurements in the first 10 m. In addition, any sub-basin (as defined in Fig.2) and month containing fewer than 5 profiles was discarded

Minor comments.

1. L23 "daily variability" – this is somewhat ambiguous. Do you mean between days or within days?
   We mean between days variability we substitute "when the daily variability is filtered out." with the following: "when the variability between days is filtered out."

2. L24 replace 'high' with a number. We propose to rephrase with "Both BOUSSOLE and BGC-Argo indicate that bias is up to 20 % for the irradiance at 380 nm, 412 nm, and for wavelenghts higher than 670 nm, whereas it decreases to less than 5% at the other wavelengths."

3. 'cloud dynamics and seasonality" – the former is a process, the later a timescale. They don't quite fit together within one phrase. We agree, we removed "seasonality".

4. L30 remove "notably" Ok, removed.

5. L47 -51. Confusing. Tried to say too much in one sentence. We propose to rephrase in order to make the sentence more clear. This meets the requirements and high data quality standard expected for remote system calibration of ocean colour spaceborne sensors (Antoine et al., 2020) and for the Copernicus biogeochemical operational model system for the Mediterranean Sea (MedBFM; Lazzari et al., 2010, 2012, 2016; Cossarini et al., 2015; Teruzzi et al., 2014, 2018, 2019; Salon et al., 2019). This system is used to produce analysis, forecasts and reanalysis of the biogeochemical state, recently upgraded to assimilate BGC-Argo float data (Cossarini et al., 2019).

6. L83 'while' – this is not a 'while' we propose to change the sentence as follows: "In OASIM, gaseous absorption by ozone, oxygen, carbon dioxide and water vapor is resolved before cloud transmittance determination, and aerosol effects are ignored in the presence of clouds."

7. L84 – remove 'the' before 'aerosol'. Ok, we agree.

8. L107 "resolve the diurnal variability" – this is within a day right? Yes, it is.

9. L108 'properly' is a subjective, rather than objective, adjective. We remove this adjective.

10. L115 remove 'totally'. Ok, removed.

11. L134. At this point I didn't know how you were defining regions. We rephrased the text as indicated above specifying that the sub-basins are shown in Fig.2.

12. L166. Would it be more accurate to say "W m-2 per waveband"? Yes, we propose the updated the text with "The OASIM outputs for the irradiance are expressed in W m-2 per waveband"

13. L167 “to W m-2”. We converted the data and model to W m-2 nm-1 as shown in the figures.

14. L202 – 15’ and 1-degree – most people would write 15 minutes and 1o!. Yes, we updated to suggested standards.

15. L257 – stick with bias = model – observation. Don’t say a negative bias, but a bias of -20%. Otherwise it can get confusing. Yes, we agree, we changed the text accordingly.

16. L264, L265 – subscript of d different to elsewhere. Ok, we corrected.

17. L294 ‘float cluster” is new wording that is unnecessary. We modified with BGC-Argo float.

18. L354. More details on the wavelength discretisation. For example, Does 412 sit in a 400 to 425 nm band? The bands are centred on 400nm and 425nm so 412 nm is at the interface of two bands. We propose to add the following sentence to the text at L 354: “Especially due to the fact that in the present simulations the 412 nm wavelength is at the interface between the band centred at 400 nm and the one centred at 425 nm.”.

19. L373 “The OASIM model . . . this information” – Conclusions need tighter sentence than this. We agree to remove this sentence.

20. L386-387 – this manuscript could be more helpful for motivating this sentiment as per main point 2. We propose to add the following sentence to the conclusions: L385:” The improvement of the model skill at BOUSSOLE, when variability between days is filtered out, indicates that spatial and temporal resolution in resolving clouds distributions is probably the most important parameter affecting skill. Nonetheless, the analysis of the relative contribution of Edir and Edif indicates that skill is correlated to their ratio, suggesting that improving the physical description of the radiative processes should be considered. To this end, novel atmospheric models …”

21. Fig 4. Wind speed at what height? Wind speed is at 10 meters, we updated the caption of Fig. 4 including also this information.

22. Fig 12-Fig 13 – It took me a moment to work out what ‘B’ and ‘F’ meant, especially since you them both BioArgos and floats. In figure BOUSSOLE and BioArgo would be quicker for interpretation. Yes, we agree with the reviewer to write the name in full extensions.