On the interplay between ocean color data quality and data quantity: Impacts of quality control flags

Chuanmin Hu[1], Brian Barnes[1], Lian Feng[1], Menghua Wang[2], Lide Jiang[2]

[1]University of South Florida; Email: huc@usf.edu

[2]NOAA/NESDIS Center for Satellite Applications and Research, 5830 University Research Ct., College Park, Maryland 20740, USA

Abstract

Nearly all calibration/validation activities on satellite ocean color missions have been focused on data quality in the past in order to produce data products of the highest quality (i.e., science quality) for climate-related research. Little attention has been paid to data quantity, however, and how data quality control during data processing impacts data quality and data quantity has rarely been reported. In this letter, we attempt to fill this knowledge gap using measurements from the Visible Infrared Imaging Radiometer Suite (VIIRS) onboard the Suomi National Polar-orbting Partnership (SNPP) as an example. This is because the same Level-1B data are processed independently using different quality control methods by NASA and NOAA, respectively, allowing for an in-depth evaluation. Results indicate that stray light and sun glint are the two primary quality control flags affecting data quantity, where the criteria for flagging pixels "contaminated" by stray light and sun glint may be relaxed in the NASA ocean color data processing in order to increase data quantity without compromising data quality.

Keywords:

Remote sensing, ocean color, data quality, data quantity, quality control, Level-2 flags stray light, sun glint, calibration, validation, uncertainty

1. Introduction

In order to establish a seamless climate data record from satellite ocean color measurements, significant amount of effort has been dedicated to calibration of the satellite-measured top-of-atmosphere (TOA) radiance and validation of the derived data products, in addition to development and improvement of algorithms and data processing methods. There is a wealth of literature on these efforts, from vicarious calibration, atmospheric correction, and field-based validation (e.g., SeaWiFS technical report series; Gordon and Wang, 1994; Gordon, 1997; Wang and Bailey, 2001; Stumpf et al., 2003; McClain et al., 2004; Bailey and Werdell, 2006; Franz et al., 2007; Wang and Shi, 2007; McClain, 2009; Cannizzaro et al., 2013; Moore et al., 2015; Barnes et al., 2019; many others).

However, nearly all these previous efforts have primarily focused on producing data products of the highest quality (i.e., science quality as opposed to provisional quality during the initial phase of data processing after satellite launch), where various quality control flags are used to mark pixels of "low quality". In contrast, little attention has been paid on data quantity. On the other hand, once data quality is verified, the ultimate advantage of remote sensing is in its data quantity (i.e., spatial and temporal coverage frequency), otherwise ship-based measurements are always better in data quality than those from remote sensing because they can provide more direct measurements with lower data uncertainties.

In 2016, Feng and Hu (2016) demonstrated that after discarding low-quality ocean color data as marked by the various quality control flags, daily percentage valid observations (DPVOs) from the Moderate Resolution Imaging Spectroradiometer (MODIS) on both the Aqua and Terra over global oceans were only ~5% for an average 1-km pixel. This is actually a surprise considering that MODIS had near daily global coverage and cloud-free fraction over global oceans is between 25–30% (King et al., 2013). Assuming daily coverage from MODIS, the reduction from 25–30% to ~5% must be due to factors other than clouds. However, it is not trivial to determine what factors contributed to such a data reduction, as it requires a complete data reprocessing after varying individual quality control criteria.

Such a difficulty may be circumvented by evaluating data products from two independent ocean color data processing systems using the same satellite data, where the quality control criteria are different from the data processing. One example is from measurements of the Visible Infrared Imaging Radiometer Suite (VIIRS) onboard the Suomi National Polar-orbting Partnership (SNPP), where data have been processed independently by the two groups: the NASA Ocean Biology Processing Group (OBPG) and the NOAA STAR Ocean Color Team. The data processing software and methods on data quality control are all different, thus providing an excellent opportunity to evaluate the impacts of quality controls on data quality and data quantity, which is also the main objective of this letter. From such comparisons, we hope to provide suggestions on future ocean color data processing in order to achieve an optimal trade between data quality and data quantity.

2. Data and Methods

VIIRS-SNPP global Level-3 data between 2012 and 2018 have been obtained from two sources. One is from NASA GSFC (https://oceancolor.gsfc.nasa.gov), and the other is from NOAA STAR Ocean Color Team (Wang et al., 2017). The former were produced using the quality control methods specified in the NASA software SeaDAS, which are also named as "NASA L2GEN" in this study. The latter were produced by using different processing and quality control methods, which are named as "NOAA MSL12" in this study.

In addition to these global Level-3 data (4.6 km x 4.6 km bin), Level-2 daily data at original resolution covering the entire Gulf of Mexico for the year of 2013 were also obtained from the above two sources in order to examine the various Level-2 quality control flags (i.e., l2_flags). The description of the l2_flags can be found in Table 1. The global Level-3 data were generated from the Level-2 files after applying the various flags, defined in the "L3 Mask" field of Table 1: if a pixel is associated with any flag under "L3 Mask", the pixel is not used in the Level-3 composite.

Furthermore, VIIRS Level-2 data corresponding to *in situ* measurements from 53 cruise surveys between 2012 and 2017 (Barnes et al., 2019) were obtained from both data sources in order to examine data quantity and data quality when different flags are applied. These spectral remote sensing reflectance ($R_{rs}(\lambda)$, sr^{-1}) data were collected from the Gulf of Mexico and other North America waters from 53 cruise surveys ranging from 1 day to 35 days, mostly from coastal waters (Barnes et al., 2019). The *in situ* – satellite matching pairs were selected by applying the flags under "Validation" in Table 1, together with a 3x3 spatial homogeneity test.

Data quantity is measured by the daily percentage valid observation (DPVO) (Feng and Hu, 2016a) at each 1-km location, calculated as:

$$DPVO = N_v / (4.6 * 4.6 * N_{day}) \times 100\%, \qquad (1)$$

where N_v is the number of valid retrievals in the Level-3 grid from that month, 4.6 (km) is the resolution of the grid, and N_{day} is number of days during the month (28 – 31 days for different months).

DPVO for each climatological month is calculated as the arithmetic average of monthly means, which is then used to calculate the climatological mean for the data period.

3. Results

3.1. VIIRS DPVOs

Figures 1a & 1b show the VIIRS average DPVOs for the period of 2012 – 2018, derived from both MSL12 and L2GEN ocean color data processing, respectively. Although their spatial patterns are similar, on average MSL12 nearly doubles the data quantity of L2GEN (9.8% versus 5.1%). The increases are not uniform across the global ocean but show regional differences, where for most ocean regions the increases are from 50–100%. In some regions, however, there is a slight decrease from L2GEN to MSL12. These regions mainly include the equatorial Atlantic off the West Africa and the Arabian Sea.

3.2. Causes of different DPVOs between MSL12 and L2GEN

Because the Level-1 data used to generate Level-2 and Level-3 data products are identical in the MSL12 and L2GEN processing, the main reason causing the different DPVOs must be their different selections of flags when generating the Level-3 global composites. This is because that although the details in their vicarious calibration and atmospheric correction methods may be slightly different, the approaches are essentially the same. Table 1 shows that of all flags used in generating the Level-3 global composites, the only differences are their ways to flag clouds, stray light, and high sun glint. Therefore, in order to diagnose what caused the significant difference in the DPVOs between MSL12 and L2GEN, the proportions of pixels marked as clouds, stray light (or stray light/cloud shadow), and high sun glint for the GoM data are calculated and presented in Fig. 2. For each of the three flags, the definition is different between L2GEN and MSL12:

- "CLDICE" of L2GEN versus "CLOUD" of MSL12 – the former is defined as Rayleigh- and glint-corrected reflectance at the VIIRS near-infrared (NIR) band 862 nm ($R_{rc}(862)$, dimensionless) > 0.027 (Robinson et al., 2003), while the latter is defined at the VIIRS shortwave infrared (SWIR) band 1238 nm as $R_{rc}(1238) > 0.0235$ (Wang and Shi, 2006).
- "STRAYLIGHT" of L2GEN versus "CLDSHDSTL" of MSL12 – the former is a simple dilation from the identified "CLDICE" pixels, while the latter is based on spectral

analysis of individual cloud-adjacent pixels (Jiang and Wang, 2013). Because of the data aggregation scheme applied along the scan direction (zone 1 for near nadir-view, zone 3 for near-edge view, and zone 2 for intermediate pixels), the dilation is 9 x 7, 13 x7, and 25 x7 for zones 1, 2, and 3, respectively.

- "HIGLINT" of L2GEN versus "HIGLINT" of MSL12 – while the glint strength is calculated using NCEP wind and Cox and Munk model (1954) with detailed provided in Wang and Bailey (2001), the threshold to define "HIGLINT" pixels is different: 0.005 sr^{-1} for the former but 0.01 sr^{-1} for the latter.

From Fig. 2, it is clear that the primary reasons why MSL12 leads to higher DPVOs than L2GEN are their different ways to flag stray light and high sun glint pixels as opposed to cloud masking. More pixels are flagged as stray light and high sun glint from the L2GEN processing, leading to lower DPVOs (14%) than from the MSL12 processing (21%) for this particular case (Gulf of Mexico, 2013). The same reasons could be extended to the difference in the global DPVOs (Fig. 1).

3.3. Data quality of the "additional" pixels from MSL12

While VIIRS DPVOs from the MSL12 ocean coolor data processing are higher than from the L2GEN processing because of primarily of their difference in flagging stray light and high sun glint pixels, the question is whether the pixels flagged by L2GEN but not by MSL12 (i.e., "additional pixels") have similar data quality as those pixels not flagged by either L2GEN or MSL12 (i.e., "common pixels"). To partially answer this question, the compiled dataset in Barnes et al. (2019) from 53 cruise surveys was used to evaluate the data quality of "common" pixels and "additional" pixels separately. Results are shown in Fig. 3.

Of all *in situ* data points, 90 found "matching" VIIRS data from the MSL12 processing but only 57 found "matching" VIIRS data from the L2GEN processing, consistent with their differences in the global DPVOs. Forty nine (49) of these matching pairs are common from the both processing, whose evaluation statistics are listed in Table 2 (under "Common"). Forty one (41) of these matching pairs are unique from the MSL12 processing, whose evaluation statistics are also listed in Table 2 (under "Unique"). Note that the MSL12-unique data in Barnes et al. (2019) were derived from using the "STRAYLIGHT" flag instead of "CLDSDSTL" flag, thus resulting in fewer data points in Barnes et al. (2019).

From visual inspection of Fig. 3, the data quality of the MSL12-unique $R_{rs}(\lambda)$ data does not appear to be different from the "common" data for all five wavelengths. They show similar data spread around the 1:1 lines over most of the data range. This is confirmed by the statistical measures listed in Table 2. From all these statistical measures, there is no consistent pattern showing that the quality of one dataset (either "common" or "unique") is consistently better than the other. Therefore, statistically, the quality of the "common" data and MSL12-unique data can be regarded as similar. There are also 8 matching pairs that are unique to the L2GEN processing, but the number is too small to lead to statistically meaningful results.

4. Discussion

Several findings from this simple exercise are noteworthy.

First, despite the fact that VIIRS has a much wider swath (3060 km) than MODIS swath (2330 km), VIIRS global DPVOs from L2GEN are comparable to MODIS DPVOs from L2GEN (Feng and Hu, 2016), all around 5%. This is mainly because that most of the increased swath is "truncated" by the large-sensor-zenith-angle flag (> 60^{o}) in the global composites. Some of the increased swath, however, does remain because VIIRS altitude (829 km) is higher than MODIS (705 km), leading to 18% more coverage even after 60^{o} truncation. This increased coverage is compensated by the more conservative stray light dilation than MODIS (7 x 5), resulting in similar global DPVOs between VIIRS and MODIS from the same L2GEN ocean color data processing.

Second, global DPVOs from the MSL12 processing nearly double those from the L2GEN processing. Although further evaluation of the data quality for the "additional" data (i.e., those unique to MSL12 processing) at global scale is still required because the limited validation is mostly restricted to North America coastal waters, the results do support the proposal by Hu et al. (2019) that some of the Level-2 flags should be revisited in order to improve coverage without compromising data quality. Indeed, Hu et al. (2019) showed that a simple dilation change from 7x5 to 3x3 pixels could increase MODIS global DPVOs by ~40% (i.e., from the original 5% to 7%). More importantly, it appears that a 3x3 dilation can also work on VIIRS data (Hu et al., 2019). Therefore, in future data reprocessing, after some additional evaluations at global scale, L2GEN could adopt the new 3x3 dilation for flagging stray light.

Third, the second primary reason (after stray light masking) leading to the large differences in DPVOs is the use of different "HIGLINT" threshold: 0.005 sr^{-1} for L2GEN but 0.01 sr^{-1} for MSL12. The latter threshold was actually recommended by Wang and Baily (2001) for the Sea-viewing Wide Field-of-view Sensor (SeaWiFS) ocean color data processing as it allows for more pixels to be used in global composites. Clearly, although some validations have been done over clear waters (Wang and Baily, 2001), more *in situ* validation of the VIIRS pixels with sun glint reflectance between 0.005 and 0.01 sr^{-1} will be helpful for quality assurance of these "additional" data from the MSL12 processing. Unfortunately, of all data collected from the 53 cruise surveys used in this study, no matching pair were found in this category. Future study is therefore required to search for such matching pairs at global scale.

Fourth, the ~10% global DPVOs from the MSL12 processing is actually very similar to MODIS sea surface temperature (SST) DPVOs shown in Feng and Hu et al. (2016a), suggesting this is possibly a current upper limit for ocean color retrievals. It is still far from the 25-30% cloud-free probability because many pixels are still classified as stray light ("CLDSDSTL") and high sun glint (> 0.01 sr^{-1}) in addition to data truncation after 60° sensor-zenith angle. While there is perhaps no hope to "recover" such stray light contaminated pixels, alternative atmospheric correction scheme such as POLYMER (Steinmetz et al., 2011; Zhang et al., 2018) may "recover" most of the sun glint contaminated pixels, thus may further increase data quantity.

Fifth, although DPVOs from the MSL12 processing are mostly higher than those from the L2GEN processing, it is not always the case. For the *in situ* validation, there are some matching pairs that are unique to the L2GEN processing (Fig. 3). For the global distributions, in several reasons MSL12 showed lower DPVOs than L2GEN, for example in the eastern equatorial Atlantic. This is mostly due to the different treatment of absorbing aerosols (e.g., dusts).

Finally, the global DPVOs appear very low regardless of the processing schemes. For example, a 5% DPVO indicates that there is one valid retrieval every 20 days. In practice, ocean color data are rarely used at 1-km daily resolution in climate-related studies. Rather, they are binned both spatially and temporally (e.g., 4.6-km or 9-km bin and monthly bin). After spatial and/or temporal binning, the data gaps are significantly reduced. However, this does not rule out the need to improve DPVOs for various applications. For example, for daily snapshot images, increases in DPVOs would improve observations of algae blooms and other ocean features. For long-term time series studies, increases in DPVOs would reduce data product uncertainties.

Conclusion

The most restricting factors in addition to clouds on ocean color data quality and data quantity are stray light and sun glint, as shown here using VIIRS observations. Different treatments of stray light and sun glint, from the NASA L2GEN and NOAA MSL12 ocean color data processing, respectively, can result in significantly different data quantity in the global data composites. Relaxing some of the flagging criteria may lead to significantly increased data quantity without compromising data quality, as shown from the field-based validation using data collected from North America waters. Once validated with more extensive global datasets, the relaxed quality control flags may be adopted in data reprocessing using L2GEN (or other ocean color data processing) in order to improve global data coverage.

Acknowledgements

This study was supported by the U.S. NASA (NNX16AQ71G, NNX14AM63G, NNX15AB13A, and 80NSSC18K0340) and by the Joint Polar Satellite System (JPSS) funding for the NOAA ocean color calibration and validation (Cal/Val) project (NA15OAR4320064). The authors wish to thank both NASA and NOAA for providing the processed satellite data used in this study. The in situ data were collected mainly by the USF Optical Oceanography Lab and Florida Fish and Wildlife Research Institute from targeted cruise surveys or cruises of opportunity, including Gulf of Mexico Research Initiative (GoMRI) DEEPEND cruises (DP01-DP05). Some of the in situ data used in this study were collected during the NOAA dedicated VIIRS ocean color calibration and validation (Cal/Val) cruises (NF-14-09, NF-15-13, NF-16-08) and the GOMECC-3 cruise, supported by the Joint Polar Satellite System (JPSS) program, NOAA Office of Marine and Aviation Operations, and GOMECC. We thank many individuals on these cruises for their assistance in collecting field data. GoMRI data are publicly available through the Gulf of Mexico Research Initiative Information & Data Cooperative (GRIIDC) at https://data.gulfresearchinitiative.org (DOIs: 10.7266/N7ZC818X, 10.7266/N7NC5ZK6, and 10.7266/N7ZK5F24). The views, opinions, and findings contained in this paper are those of the authors and should not be construed as an official NOAA or U.S. Government position, policy, or decision.

References

Bailey, S.W., & Werdell, P.J. (2006). A multi-sensor approach for the on-orbit validation of ocean color satellite data products. *Remote Sens. Environ.*, 102, 12-23.

Cannizzaro, J., C. Hu, K. L. Carder, C. R. Kelble, N. Melo, E. M. Johns, G. A. Vargo, and C. A. Heil (2013). On the accuracy of SeaWiFS ocean color data products on the West Florida Shelf. J. Coastal Res., 29(6), 1257--1272. DOI: http://dx.doi.org/10.2112/JCOASTRES-D-12-00223.1.

Cox, C., and W. H. Munk (1954), The measurement of the roughness of the sea surface from photographs of the sun's glitter, J. Opt. Soc. Am., 44, 838–850.

Feng, L., and C. Hu (2016). Comparison of Valid Ocean Observations Between MODIS Terra and Aqua Over the Global Oceans. IEEE Trans. Geosci. Remote Sens. 54:1575-1585.

Franz, B. A., S. W. Bailey, P. J. Werdell, and C. R. McClain (2007). Sensor-independent approach to the vicarious calibration of satellite ocean color radiometry. Appol. Opt. 46:5068-5082.

Gordon, H. R., and M. Wang (1994), Retrieval of water-leaving radiance and aerosol optical thickness over the oceans with SeaWiFS: A preliminary algorithm, Appl. Opt., 33, 443–452.

Gordon, H.R. (1997). Atmospheric correction of ocean color imagery in the Earth Observing System era. *J. Geophys. Res., 102*, 17081-17106.

Hu, C., Feng, L., Lee, Z., Franz, B. A., Bailey, S. W., Werdell, P. J., & Proctor, C. W. (2019). Improving satellite global chlorophyll a data products through algorithm refinement and data recovery. Journal of Geophysical Research: Oceans, 124. https://doi.org/10.1029/2019JC014941.

Jiang, L., and M. Wang (2013). Identification of pixels with stray light and cloud shadow contaminations in the satellite ocean color data processing. Appl. Opt., 52: 6757-6770.

King, M. D., S. Platnick, W. P. Menzel, S. A. Ackerman, and P. A. Hubanks (2013). Spatial and temporal distribution of clouds observed by MODIS onboard the Terra and Aqua satellites. IEEE Trans. Geosci. Remote Sens., vol. 51, no. 7, pp. 3826–3852.

McClain, C. R., Feldman, G. C., & Hooker S. B. (2004). An overview of the SeaWiFS project and strategies for producing a climate research quality global ocean bio-optical time series. *Deep-Sea Research II, 51*, 5-42.

McClain, C. R. (2009). A decade of satellite ocean color observations. Annu Rev Mar Sci 1:19–42.

Moore, T.S., J.W. Campbell, H. Feng. (2015). Characterizing the uncertainties in spectral remote sensing reflectance for SeaWiFS and MODIS-Aqua based on global in situ matchup data sets. Remote Sens. Environ., 159: 14–27.

Steinmetz, F., Deschamps, P.-Y., Ramon, D., 2011. Atmospheric correction in presence of sun glint: application to MERIS. Opt. Express 19, 9783–9800.

Stumpf, R. P., R. Arnone, R. W. Gould et al. (2003). A partially-coupled ocean-atmosphere model for retrieval of water-leaving radiance from SeaWiFS in coastal waters, vol. 22, NASA Goddard Space Flight Center, Greenbelt, Maryland, 2003.

Wang, M., and S. W. Bailey (2001). Correction of sun glint contamination on the SeaWiFS ocean and atmosphere products. *Appl. Opt.* 40:4790-4798.

Wang, M., Shi,W., 2006. Cloud masking for ocean color data processing in the coastal regions. IEEE Trans. Geosci. Remote Sens. 44 (11), 3196–3205.

Wang, M., & Shi, W. (2007). The NIR-SWIR combined atmospheric correction approach for MODIS ocean color data processing. *Optics Express*, *15*, 15722-15733

Wang, M., Liu, X., Jiang, L., Son, S., 2017. Visible Infrared Imaging Radiometer Suite (VIIRS) Ocean Color Produts Algorithm Theoretical Basis Document Version 1.0.

Zhang, M., C .Hu, J. Cannizzaro, D. English, B. B. Barnes, P. Carlson, and L. Yarbro (2018). Comparison of two atmospheric correction approaches applied to MODIS T measurements over North American waters. Remote Sens. Environ., 216:442-455.

Tables and Captions

Table 1. Level-2 processing quality control flags (l2_flags) defined in a 32-bit long integer (http://oceancolor.gsfc.nasa.gov/atbd/ocl2flags/; Wang et al., 2017). Adapted from Barnes et al. (2019).

Bit position	*L3 Mask, L2GEN	**Validation, L2GEN	*L3 Mask, MSL12	**Validation, MSL12	Name (L2GEN)	Name (MSL12)	L2GEN Description [MSL12 description]
0	x	x	x	x	ATMFAIL	ATMFAIL	Atmospheric correction failure
1	x	x	x	x	LAND	LAND	Pixel is over land
2					PRODWARN	PRODWARN	Warning from ≥ 1 product algorithms
3	**X**	**X**	**X**	**X**	HIGLINT	HIGLINT	Sunglint: reflectance exceeds threshold
4	x	x	x	x	HILT	HILT	Radiance very high or saturated
5	x	x	x	x	HISATZEN	HISATZEN	Sensor zenith angle exceeds threshold
6					COASTZ	COASTZ	Pixel is in shallow water
7					Spare	LANDADJ	[Probable land-adjacent contamination]
8	**X**	**X**			STRAYLIGHT	STRAYLIGHT	Probable stray light contamination
9	**X**	**X**	**X**	**X**	CLDICE	CLOUD	Probable cloud or ice contamination
10	x		x		COCCOLITH	COCCOLITH	Coccolithophores detected
11					TURBIDW	TURBIDW	Turbid water
12	x	x	x	x	HISOLZEN	HISOLZEN	Solar zenith angle exceeds threshold
13					Spare	HITAU	[High Aerosol Optical Thickness]
14	x	x	x	x	LOWLW	LOWLW	Very low water-leaving radiance
15	x		x		CHLFAIL	CHLFAIL	Chlorophyll algorithm failure
16	x	x	x	x	NAVWARN	NAVWARN	Navigation quality is suspect
17	x		x		ABSAER	ABSAER	Absorbing Aerosols determined
18			**X**	**X**	Spare	CLDSHDSTL	[Cloud straylight or shadow]
19	x	x	x	x	MAXAERITER	MAXAERITER	NIR iteration limit reached
20					MODGLINT	MODGLINT	Moderate sun glint
21					CHLWARN	CHLWARN	Chlorophyll out-of-bounds
22	x	x	x	x	ATMWARN	ATMWARN	Atmospheric correction is suspect
23					Spare	ALGICE	[Sea ice identified by nLw]
24					SEAICE	SEAICE	Pixel is over sea ice
25	x	x	x	x	NAVFAIL	NAVFAIL	Navigation failure
26					FILTER	FILTER	Insufficient data for smoothing filter
27					Spare	ALTCLD	[Cloud detected]
28					BOWTIEDEL	FOG	VIIRS deleted overlapping pixels [Fog]
29					HIPOL	FROMSWIR	High polarization [SWIR atm. corr. used]
30					PRODFAIL	PRODFAIL	Failure in any product
31					SPARE	OCEAN	[Pixel is over ocean]

* The L3 mask is used for generation of global composite data products.
** These flags are applied when performing *in situ* – satellite comparisons.
The highlighted flags have different definitions between L2GEN and MSL12, therefore analyzed in this study.

Table 2. Matchup statistics of VIIRS data for collected in the Gulf of Mexico and other North America waters (station locations shown in Fig. 3 of Barnes et al., 2019). The graphical forms are presented in Fig. 3 of this paper. UPD: unbiased percentage difference; MRD: mean relative difference; RMSE: root-mean-square-error; ; MR: mean ratio; ; RMA β_1: reduced major axis (i.e., "Model II") regression slope; RMA β_0: regression intercept.

412 nm	COMMON	UNIQUE
UPD (%)	27	73
MRD (%)	32	-1
RMSE (sr^{-1})	0.0014	0.0013
MAPD (%)	45	56
MR	0.97	2.82
STDR	0.48	12.34
N	49	41
R^2	0.86	0.88
RMA β_0	0.0003	-0.0003
RMA β_1	1.06	1.22

443 nm	COMMON	UNIQUE
UPD (%)	20	32
MRD (%)	18	10
RMSE (sr^{-1})	0.0014	0.0014
MAPD (%)	27	33
MR	0.94	1.09
STDR	0.24	0.53
N	49	41
R^2	0.80	0.89
RMA β_0	0.0000	-0.0005
RMA β_1	1.09	1.29

486 nm	COMMON	UNIQUE
UPD (%)	18	22
MRD (%)	13	8
RMSE (sr^{-1})	0.0019	0.0017
MAPD (%)	22	23
MR	0.94	1.00
STDR	0.21	0.34
N	49	41
R^2	0.74	0.94
RMA β_0	-0.0006	-0.0009
RMA β_1	1.22	1.29

551 nm	COMMON	UNIQUE
UPD (%)	21	19
MRD (%)	15	8
RMSE (sr^{-1})	0.0021	0.0018
MAPD (%)	27	20
MR	0.95	0.99
STDR	0.25	0.30
N	49	41
R^2	0.79	0.96
RMA β_0	0.0002	-0.0007
RMA β_1	1.06	1.21

671 nm	COMMON	UNIQUE
UPD (%)	40	8
MRD (%)	43	35
RMSE (sr^{-1})	0.0007	0.0009
MAPD (%)	61	67
MR	1.03	0.96
STDR	0.93	0.62
N	49	41
R^2	0.72	0.76
RMA β_0	0.0002	-0.0003
RMA β_1	0.91	1.35

Figure captions

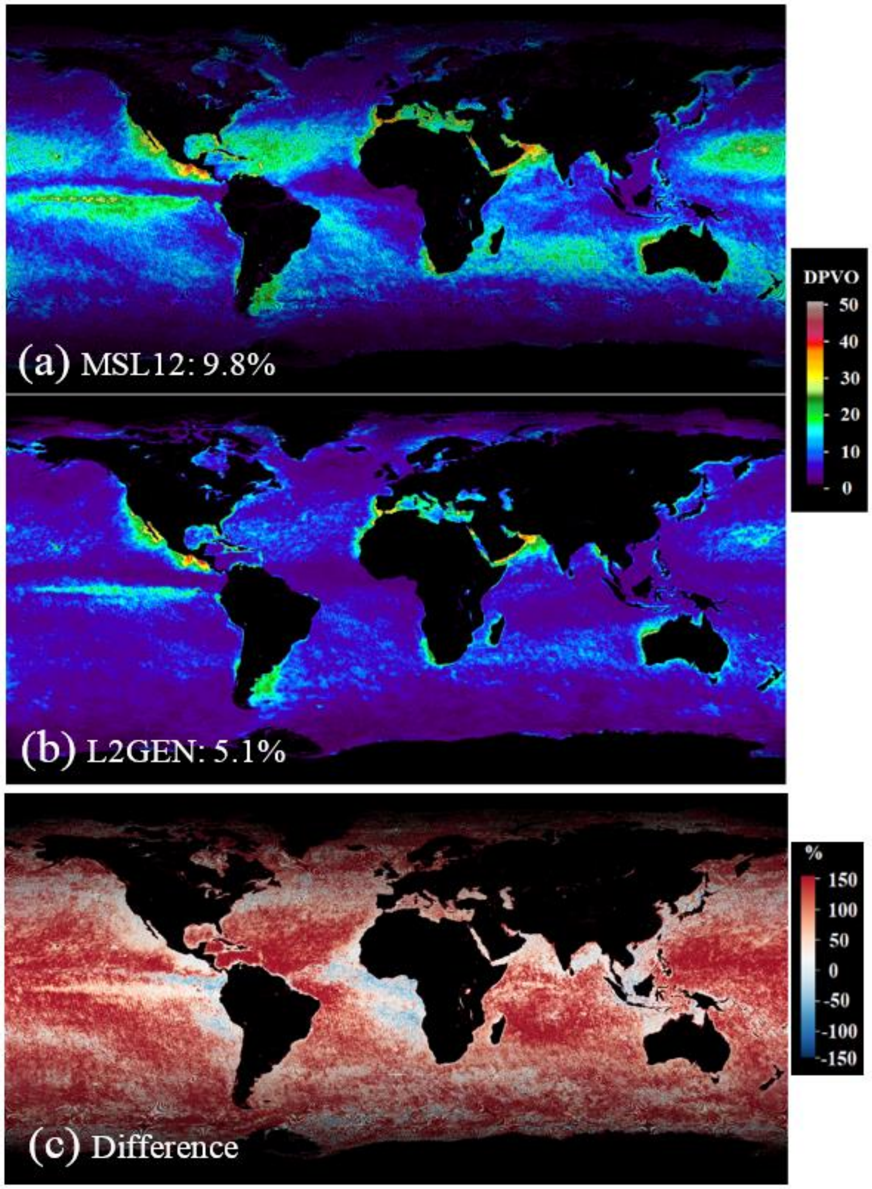

Fig. 1. VIIRS average DPVOs from (a) NOAA MSL12 and (b) NASA L2GEN processing for the period of 2013 – 2018. The global means are 9.8% and 5.1%, respectively. (c) Relative differences of DPVOs of between MSL12 and L2GEN, defined as (MSL12 – L2GEN)/L2GEN × 100%.

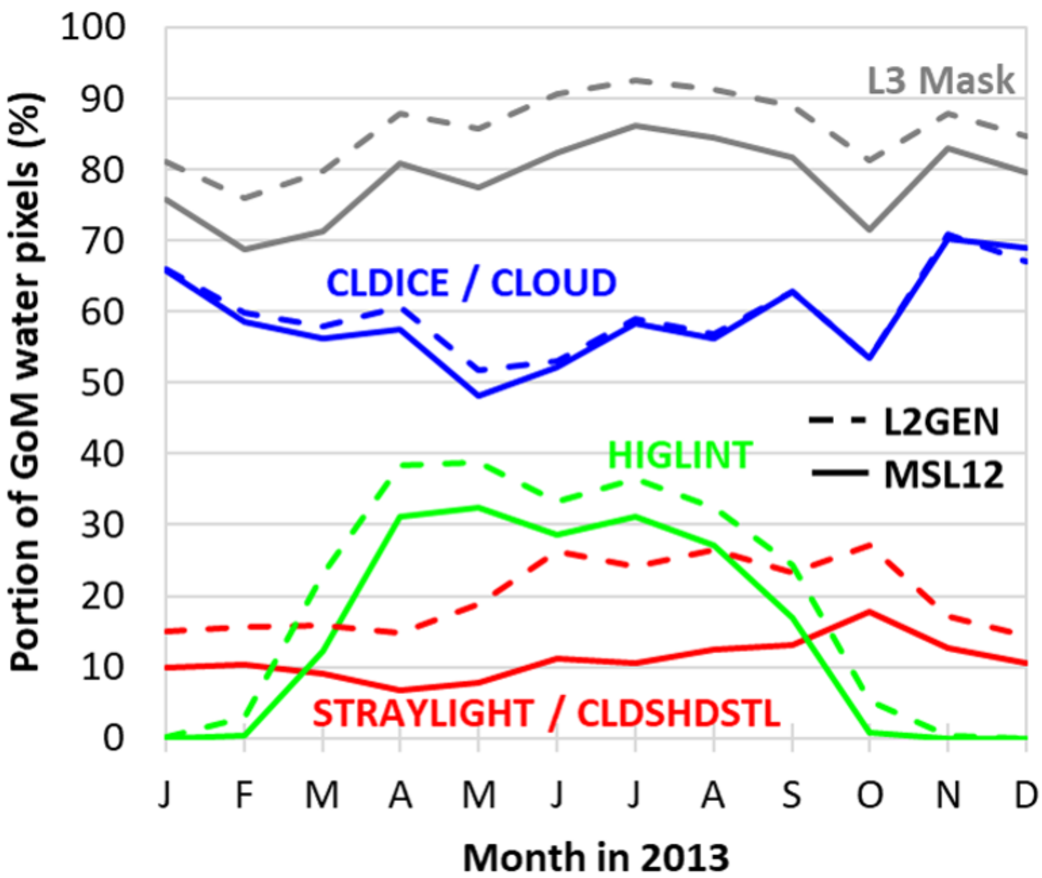

Fig. 2. Comparison between MSL12 and L2GEN for their flagged pixels (in percentage of the total pixels covering the Gulf of Mexico). Note that L2GEN flags way more pixels as STRAYLIGHT and HIGLINT than MSL12. These are the primary reasons that MSL12 leads to much higher DPVOs (21% overall, 25% in winter and 17% in summer) than L2GEN (14% overall, 18% in winter, and 10% in summer) for the Gulf of Mexico for 2013.

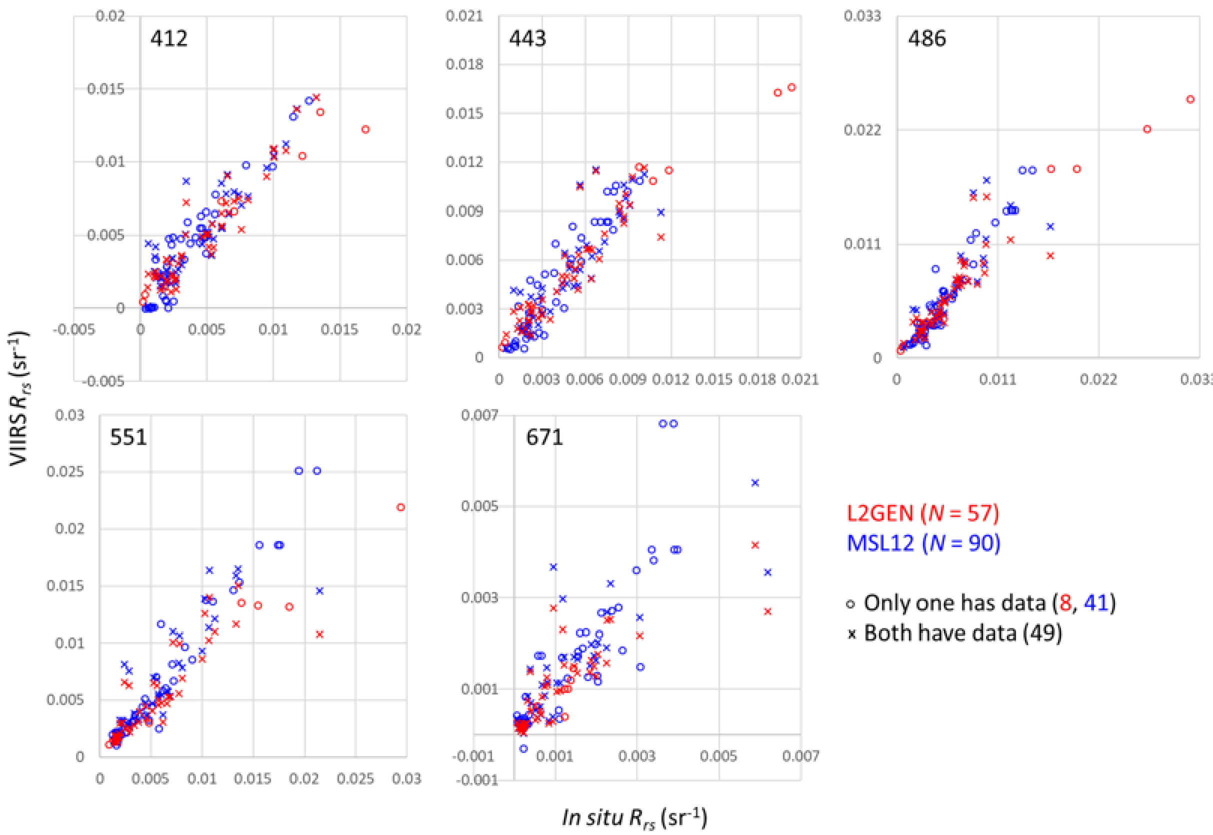

Fig. 3. Comparison between VIIRS-derived and *in situ* measured R_{rs} for the 5 VIIRS bands. While 57 matching pairs were found for L2GEN, 90 matching pairs were found for MSL12, 41 of which were unique to MSL12 due to its relaxed STRAYLIGHT and HIGLINT flagging. Matchup statistics are listed in Table 2.