Global land mask for satellite ocean color remote sensing

Karlis Mikelsons [a,b,*], Menghua Wang [a], Xiao-Long Wang [a,c], Lide Jiang [a,c]

[a] *NOAA National Environmental Satellite, Data, and Information Service, Center for Satellite Applications and Research, E/RA3, 5830 University Research Court, College Park, MD 20740, USA*
[b] *Global Science and Technology, Inc., Greenbelt, MD 20770, USA*
[c] *CIRA, Colorado State University, Fort Collins, CO 80523, USA*

ARTICLE INFO

ABSTRACT

We develop a methodology to derive a global medium resolution (250 m) land mask from several existing data sources. In particular, a number of improved land mask data sets have been developed from satellite measurements recently, though some artifacts and omissions still remain. We show how combining global land mask data from multiple independent data sources can decrease the frequency of artifacts, and improve the data consistency and quality. We use the ocean color product imagery derived from measurements of the Visible Infrared Imaging Radiometer Suite (VIIRS) onboard the Suomi National Polar-orbiting Partnership (SNPP) to evaluate and validate the new global land mask implemented in the Multi-Sensor Level-1 to Level-2 (MSL12) ocean color data processing system, and demonstrate the improvements in the derived global ocean color data coverage. Results show that when using the new proposed land mask the accuracy of global ocean color data coverage is significantly improved over coastal and inland waters. The new land mask more accurately represents the current global land coverage status, providing more complete and consistent global land/water coverage data set for ocean color remote sensing and for various other satellite Earth observing applications.

1. Introduction

Land mask (or water mask) is one of the most basic Earth observation data products, yet obtaining the accurate global data is surprisingly difficult. Part of the difficulty is due to an attempt to capture a changing environment in a static data set. Another difficulty is due to the fact that different applications require different water or land mask data. In the context of monitoring global waters from satellite Earth observations, the land mask data serve an important purpose to distinguish the areas where the different retrieval algorithms designed for land or water are applicable. In particular, for the satellite ocean color remote sensing, the land mask helps to screen out any invalid ocean color retrievals, which may be contaminated near coastal areas. The land mask also aids in visualization of satellite data sets such as those in the Ocean Color Viewer (OCView) (Mikelsons and Wang, 2018).

The surface of the Earth is always changing, due to natural causes and human activities. New land is formed as sediment is deposited in river estuaries, and many areas in coastal waters are reclaimed from the sea by industrial development. Other low-lying lands are flooded by gradually rising sea levels, and new water reservoirs are artificially

created to store water. Even within the diurnal time scale, many coastal areas are subject to high and low tides, which may also significantly impact water properties (Shi et al., 2011, 2013). Thus, it is not possible to account for all these changes with a single binary land-water data mask. Nevertheless, most of Earth's surface waters have remarkably stable coastlines, especially compared with much faster changes in water properties.

Most satellite ocean color (and other physical and optical environmental quantities) retrieval algorithms rely on some prior knowledge of global land and water extents. The satellite sensors designed for monitoring Earth's waters are typically medium spatial resolution (250–1000 m) spectrometers flying in a polar sun-synchronous orbit reaching global coverage in ~1–3 days, e.g., the Moderate Resolution Imaging Spectroradiometer (MODIS) on the Terra and Aqua satellites (Esaias et al., 1998; Salomonson et al., 1989), the Visible Infrared Imaging Radiometer Suite (VIIRS) onboard the Suomi National Polar-orbiting Partnership (SNPP) and NOAA-20 (Goldberg et al., 2013), the Ocean and Land Colour Instrument (OLCI) on the Sentinel-3A and Sentinel-3B (Donlon et al., 2012), the Second-Generation Global Imager (SGLI) on the Global Change Observation Mission-Climate (GCOM-C)

([Tanaka et al., 2009](#)), etc. The medium spatial resolution (250–1000 m) is adequate for studies of most of the global oceans. Retrieval of water properties from these observations requires a land or water mask of comparable resolution. Indeed, some of the land mask data sets have been derived from medium resolution Earth observing satellite data and thus have the same spatial resolution.

Historically, data from MODIS and the Shuttle Radar Topography Mission (SRTM) have been used to derive land and water mask data sets ([Carroll et al., 2017](#); [Carroll et al., 2009](#); [Li et al., 2013](#)). In particular, MOD44Wv5 data ([Carroll et al., 2009](#)) have been used for many years by several ocean color retrieval algorithms including the Multi-Sensor Level-1 to Level-2 (MSL12) ocean color data processing system, which was originally developed for satellite ocean color data processing from multiple satellite sensors ([Wang, 1999](#); [Wang and Franz, 2000](#); [Wang et al., 2002](#)). MSL12 has been used for routinely producing VIIRS global ocean color products ([Wang et al., 2013](#)). However, MOD44Wv5 dataset was derived using MODIS measurements from 2003 to 2007 ([Carroll et al., 2009](#)) which was more than a decade ago, and thus does not capture recent changes to land and water surface. Furthermore, the land mask data for the Antarctic region in MOD44Wv5 are based on bedrock elevation data, thus this data set often misclassifies large areas perennially covered by ice as water.

A newer version, MOD44Wv6 ([Carroll et al., 2017](#)), was derived in 2017 using MODIS and SRTM data from 2000 to 2015. However, this data set is somewhat heavily biased toward including more water pixels, even in areas with seasonal water presence. Furthermore, this land mask dataset does not contain any data for latitudes south of 60°S. Due to these significant differences between MOD44Wv5 and MOD44Wv6, and due to historical use of MOD44Wv5 in the ocean color retrievals, we have opted to include both of these data sets in our analysis.

In addition to the development of MOD44Wv6, machine learning techniques were used to derive several data sets characterizing the Global Surface Water (GSW) ([Pekel et al., 2016](#)), including seasonality, occurrence, recurrence, and change from the decades long record of Landsat data with higher spatial resolution (~30 m). In fact, [Pekel et al. (2016)](#) show detailed changes to the Earth surface environment in high spatial resolution. Thus, it is perhaps the most comprehensive study up to date. However, some artifacts are still present in this data set, especially in the Arctic polar region. Data for the Antarctic region are also missing from this data set.

Another land mask data set was developed from the Landsat data from the studies of the global forest cover (GFC) ([Hansen et al., 2013](#)). This data set is designed to avoid water contamination in land forest studies, so it is somewhat biased toward including/masking more water pixels. It also has little to no data over both the Arctic and Antarctic polar areas where vegetation is scarce or non-existent ([Hansen et al., 2013](#)).

Finally, yet another approach is taken by OpenStreetMap (OSM) (www.openstreetmap.org), which is a community mapping project. Here, coastlines and contours of bodies of water are derived from multiple data sources, including high spatial resolution satellite imagery. This approach yields good quality land mask data for the coastlines, but coverage and data quality are somewhat inconsistent for inland bodies of water. For example, the level of detail and degree of spatial resolution for inland bodies of water may change from one region to another. For that reason, in this study we have chosen to use this data set only to improve the quality of the derived water mask in the polar regions.

None of these existing land (or water) mask data sets are specifically designed for satellite-based global ocean color retrievals, and indeed, none of these data sets provide a consistent good quality water mask over the entire globe. The aim of this work is first to establish a methodology to combine these data sets into a new derived land/water mask by improving the land mask coverage and eliminating the artifacts. While the aim of this work is to derive an improved land/water mask that is suitable for satellite ocean color studies, the same methodology can be used to derive land/water mask data for studies of other land or water based measurements. We note that while some of the land/water mask data sources used in this study may be derived from the same satellite measurements, the differences in methodologies ensure that the land mask data sets can be treated as independent sources for the purposes of this study. We evaluate and validate the newly derived land mask using the satellite-measured clear sky true color and false color imageries (from VIIRS SNPP and NOAA-20, as well as OLCI-Sentinel-3A), which can be used as references to compare with the new land mask data visually and quantitatively. Finally, the improvements of the new land mask are demonstrated from the comparisons of global ocean color retrievals using the MSL12 ocean color data processing system among various land mask data over global oceans, and coastal and inland waters.

2. Methodology

2.1. Combining land mask data sets

In this section, we show how to combine the land mask data sets from multiple data sources, including high spatial resolution data sources, to derive an improved medium spatial resolution global land (or water) mask. As noted above, in general, different data sources contain different quantities characterizing distribution of global surface water. While most data sources only distinguish between the two possibilities of land or water, the GSW data set includes a number of parameters characterizing the spatial and temporal distribution of global surface water ([Pekel et al., 2016](#)). For high spatial resolution data sources, multiple pixels correspond to a single medium resolution pixel, and thus provide information about the fraction of area covered by water. In addition, each data source has areas where no data are present. In order to capture all of this information in a single variable, we introduce a land-water indicator Λ, which can take values from -1 to $+1$. Positive values indicate water, and negative vales indicate land. Zero value represents absence of data, and the absolute value of this indicator measures the degree of certainty that the particular pixel represents land or water. We derive this indicator for each of the data sources, and then combine these values to obtain the indicator for the combined land/water mask.

The MOD44Wv5 and MOD44Wv6 data represent the simplest connection with the indicator Λ. We choose the spatial resolution of the combined land mask to be the same as for these data sets. For each pixel, these data sets provide three possible values: land, water, or no data. Thus, we define the land-water indicator for MOD44Wv5 data as:

$$\Lambda^{Mv5} = \begin{cases} -1 & (land) \\ +1 & (water) \\ 0 & (no\ data) \end{cases} \tag{1}$$

We note that the MOD44Wv5 data set also includes flag data specifying the data source used for each pixel, but we treat all data as equally valid, except for the case of no data. We use an identical expression to derive the land-water indicator for the MOD44Wv6 data. However, here we also take into account the values of data flags: for the flag values of 2 and 8, which indicate that the data are inherited from MOD44Wv5 data, we interpret the data as missing so to avoid double counting the same data.

The GFC data, derived from high spatial resolution Landsat data, have around 30 m spatial resolution (more precisely, 4000 pixels per degree of longitude or latitude). To derive the medium spatial resolution land-water indicator from GFC data, we first count the number of high-spatial resolution pixels that fall within the particular medium resolution pixel, and represent land, water, or no data. For the high-spatial resolution pixels that have only a partial overlap with the medium spatial resolution pixel, we count the fraction proportional to the area of overlap with the medium resolution pixel. We record the total number of high-spatial resolution land, water, and no data pixels as N_L, N_W, and N_N, respectively, and calculate the normalized medium-spatial resolution

values of water and land pixel numbers as $n_W = N_W/(N_W + N_L + N_N)$, and $n_L = N_L/(N_W + N_L + N_N)$. In order to derive the medium spatial resolution land-water indicator from the fractional land and water pixel values, we further introduce a water fraction threshold f^{GFC}. This parameter determines the threshold value for the fraction of high-spatial resolution water pixels, above which the medium-spatial resolution pixel is interpreted as water. Thus, if n_W is larger than $f^{GFC}(n_W + n_L)$, then the data indicate water, and if n_W is less than $f^{GFC}(n_W + n_L)$, then the data indicate land. Thus, it is the difference $n_W - f^{GFC}(n_W + n_L)$ that determines if the pixel is to be interpreted as water or land. But it also indicates the degree of certainty of determination. As a result, instead of immediately assigning $+1$ (water) to pixels with n_W larger than $f^{GFC}(n_W + n_L)$, and -1 (land) otherwise, we use a hyperbolic tangent function to map the values of this difference onto a continuous interval ranging from -1 to $+1$ for the land-water indicator, which is then calculated as:

$$\Lambda^{GFC} = (n_W + n_L) \tanh\left[\frac{n_W - f^{GFC}(n_W + n_L)}{\Delta^{GFC}}\right] \quad (2)$$

here, Δ^{GFC} is the smoothing parameter, which is used to smooth out the transition around the threshold value f^{GSW}, so that for pixels where the data source has water to land pixel ratio close to the threshold value, the indicator would be close to zero. For example, if 50% threshold value is used ($f^{GFC} = 0.5$), and the numbers of water and land pixels are equal, then this data source does not provide any information, and the weighted indicator will be zero. The smoothing parameter determines how the difference of the water fraction above (or below) the threshold value is related to the certainty of pixel being classified as water (or land), respectively. For example, using the same threshold value $f^{GSW} = 0.5$ when the water fraction is $n_W = 0.6$, and smoothing parameter $\Delta^{GFC} = 0.05$ will yield $\Lambda^{GFC} = \tanh(2) \approx 0.964$, a fairly certain determination of a water pixel, whereas $n_W = 0.45$ will yield $\Lambda^{GFC} = \tanh(-1) \approx -0.762$, a somewhat less certain determination of land. We discuss choosing the adjustable parameters in the next section. Furthermore, Λ^{GFC} is also proportional to the total number of known pixels, i.e., $(n_W + n_L)$, so that it will be zero in the areas where this data source has missing data.

The GSW data set includes several quantities describing water presence and temporal change, such as occurrence, recurrence, seasonality, and change. We choose to use the seasonality dataset, as it describes the state of global surface waters on a yearly basis, and provides data for the most recent time period—the year 2018 (latest as of now). Other data provided in this study quantify the evolution of the global surface waters over decades of Landsat observations and are very useful for historical analysis, but are not necessarily indicative of the most recent state of the surface waters. As in the case of GFC data, these data are derived from Landsat data series, although using a different,

machine learning based approach. The spatial resolution is also 4000 pixels per degree of longitude or latitude, same as for GFC data. However, instead of counting an entire high-spatial resolution pixel as either land or water (for valid data points), we count the fraction of the months per year that water was detected at the particular high-spatial resolution pixel. Thus, for high-spatial resolution pixels with valid data, we have $N_W^{GFC} = m_W/12$, $N_L^{GFC} = 1 - m_W/12$, and $N_N^{GFC} = 0$. Here, m_W is the number of months for which water was detected at the particular high-spatial resolution pixel (i.e., the seasonality). For high-spatial resolution pixels with no data, $N_N^{GFC} = 1$, and $N_W^{GFC} = N_L^{GFC} = 0$. We then sum N_W^{GFC}, N_L^{GFC}, and N_N^{GFC} over all high-spatial resolution pixels within each medium-spatial resolution pixel, and derive the normalized numbers of values for land and water pixel numbers in a similar manner as described for the GFC data (see Fig. 1 for graphical interpretation). We use the same formula as for the GFC data (Eq. (2)) to derive the medium spatial resolution land-water indicator for the GSW data, Λ^{GSW}. However, in general, values of the water fraction threshold f^{GSW}, and the smoothing parameter Δ^{GSW} for the GSW data may differ from those used for the GFC data.

We also use the OSM data to improve the quality of water mask in the polar areas. Since these data are available in vector format, we first transform these data into a raster format, sampling in the same ($\sim$30 m) spatial resolution as the GSW and GFC data derived from Landsat data. This yields high-spatial resolution raster data with only two possible values assigned to each pixel (land or water). We then use the same procedure as described for the GFC dataset to derive the corresponding medium-spatial resolution land-water indicator for the OSM data (Λ^{OSM}).

The combined water/land indicator is the weighted average over all the data sources:

$$\Lambda = \frac{\Lambda^{GSW}w^{GSW} + \Lambda^{GFC}w^{GFC} + \Lambda^{Mv5}w^{Mv5} + \Lambda^{Mv6}w^{Mv6} + \Lambda^{OSM}w^{OSM}}{w^{GSW} + w^{GFC} + w^{Mv5} + w^{Mv6} + w^{OSM}} \quad (3)$$

The overall weights for each individual data source are chosen based on accuracy and age of each data source. Since OSM data are only used in polar areas, we set $w^{OSM} = 0$ elsewhere. In particular, we only use the OSM data for the latitudes south of 56°S, and north of 74°N, as well as for Greenland and the waters surrounding it. As the last step, we interpret all non-negative values of combined land-water indicator Λ as water, and the negative values as land. The entire algorithm is summarized in a schematic diagram in Fig. 2.

2.2. Parameter selection

The values of the adjustable input parameters are selected based on visual evaluation of the combined land mask and comparison with each
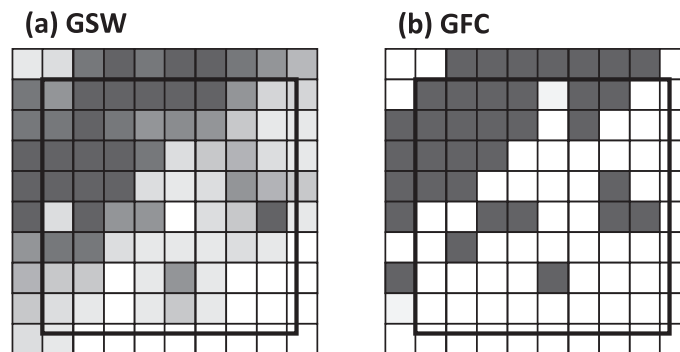

Fig. 1. Schematic illustration for deriving medium spatial resolution land-water fraction from the high-spatial resolution data: (a) GSW high-spatial resolution data of seasonality provide a fractional indicator (indicated by different shades of gray) of how many months per year water was detected in the area of each high-spatial resolution pixel and (b) GFC high-spatial resolution data only provide binary land–water value. Both GSW and GFC yield a fractional value at the medium-spatial resolution, based on the fraction of high-spatial resolution water pixels overlapping with the medium-spatial resolution pixel. The thick black line in both panels indicates the extent of a medium spatial resolution pixel. The two data sources do not necessarily have a perfect correlation.
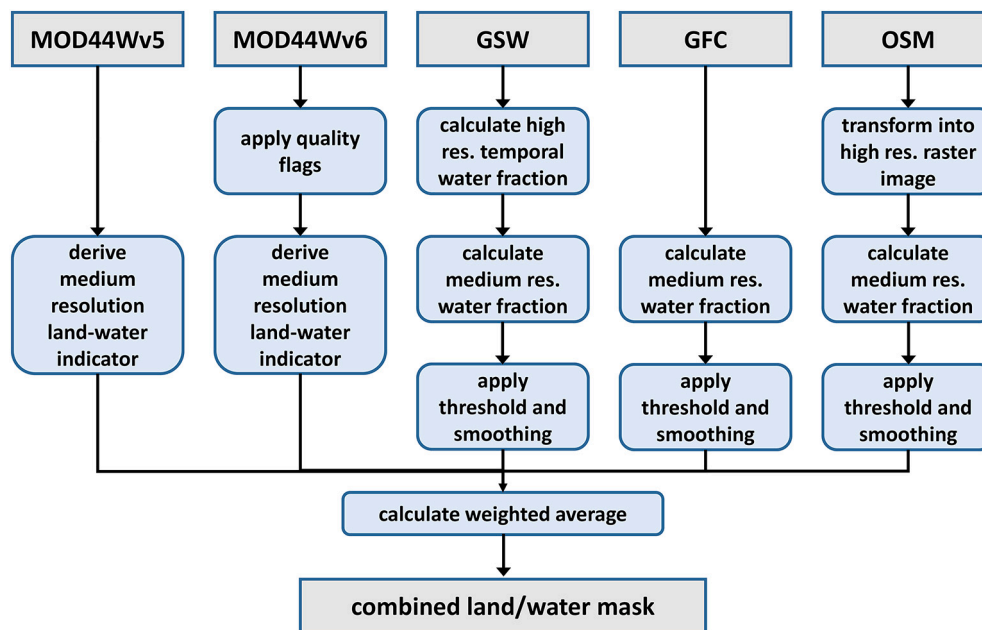
Fig. 2. Schematic diagram of the algorithm, showing the preprocessing of different data sources and combining into a single output land (or water) mask data set.

data source. The high spatial resolution data sources require the water fraction threshold and smoothing factor. Since land has typically much higher reflectance in the visible spectrum than that of water, a relatively small fraction of land within the medium resolution pixel can contaminate the ocean color observations. Therefore, for the purposes of the satellite ocean color observations, the water fraction threshold should generally be above ½. For evaluation, we chose three values of the water fraction threshold, i.e., $f = 0.7$, 0.8, and 0.9, and observed relatively small differences in the resultant combined land mask, with higher value of 0.9 producing slightly more ocean color retrievals. We choose the smoothing parameter $\Delta = 0.05$ for all three high-spatial resolution data sources, and observed very minor impact on the results with variation of Δ. The smoothing parameter Δ should generally be smaller than the differences between the water fraction threshold values f used in comparison, otherwise it will greatly reduce sensitivity to f. We note that f and Δ only have an effect on combined land mask for coastal areas, as well as for small and seasonal inland waters.

The overall weights for each data source are determined based on accuracy and age of each data source. The values of these weights should not be too dissimilar: setting the overall weight too high for a single data source will cause the combined data set be entirely determined by this data source, including all of its artifacts. Setting the overall weight too low for any data source will essentially render it irrelevant to the combined land mask data. With all of these considerations, the values of the adjustable parameters used in this work are summarized in Table 1.

It should be noted that while we found these values of parameters a good choice to aid the satellite ocean color retrievals, different application of land mask for other remote sensing products may require adjustments of these values.

Table 1
The values of adjustable parameters for the five data sources.

Land mask	Threshold (f)	Smoothing (Δ)	Weight (w)
MOD44Wv5	–	–	0.7
MOD44Wv6	–	–	0.8
GSW	0.9	0.05	1.0
GFC	0.9	0.05	0.9
OSM	0.9	0.05	0.9

3. Results

3.1. Comparison and evaluation using clear sky imagery

We use the true color and false color clear sky imageries from the years 2018 and 2019 to validate the combined land (or water) mask. The true color imagery is derived using the spectral bands corresponding to the human perception of the red, green, and blue colors (Mikelsons and Wang, 2018). In the false color imagery, however, the green channel is replaced by the near-infrared (NIR) band, at which radiance is strongly absorbed by water (Hale and Querry, 1973), but is reflected by land and especially strongly reflected by vegetation (Qi et al., 2020). The clear sky imagery is derived using series of daily images over a certain time period and selecting the statistically robust estimate for the darkest pixel from these series, thus avoiding clouds and seasonal ice, and minimizing the impact of cloud shadows. Since water appears darker in both true color and false color images, the clear sky imagery is also somewhat biased toward water. Thus, detection of (darker) water in the false color clear sky imagery only implies presence of water at some point within the time period used to derive this imagery. However, detection of (brighter) land in clear sky imagery almost certainly implies absence of water at the given location during the entire time period used to derive this imagery. The false color imagery used in this study is derived from VIIRS-SNPP daily false color imagery (Qi et al., 2020), while the true color clear sky imagery includes the data from observations of VIIRS SNPP and NOAA-20, as well as OLCI on the Sentinel-3A/3B.

We illustrate the comparison of the global combined land mask with the data sources in Fig. 3. The panels (a) and (b) in Fig. 3 show the MOD44Wv5 and MOD44Wv6 land masks, respectively. Land is displayed in gray and water is in white. While MOD44Wv5 land mask contains few inland water bodies in this area, MOD44Wv6 includes a large extent of inland water bodies, which are mostly seasonal (such as rice fields), and not necessarily relevant to ocean or water color studies. The panels (c) and (d) in Fig. 3 show the GSW seasonality and GFC data, respectively. These originally high-spatial resolution data sets have been resampled in the medium resolution with various degrees of water fraction shown in different shades of gray. In these data sets, the areas with no data are shown in pink. While the GSW seasonality data set (Fig. 3c) describes the spatial and temporal extent of inland waters in much greater detail than other data sources, it also shows some artifacts
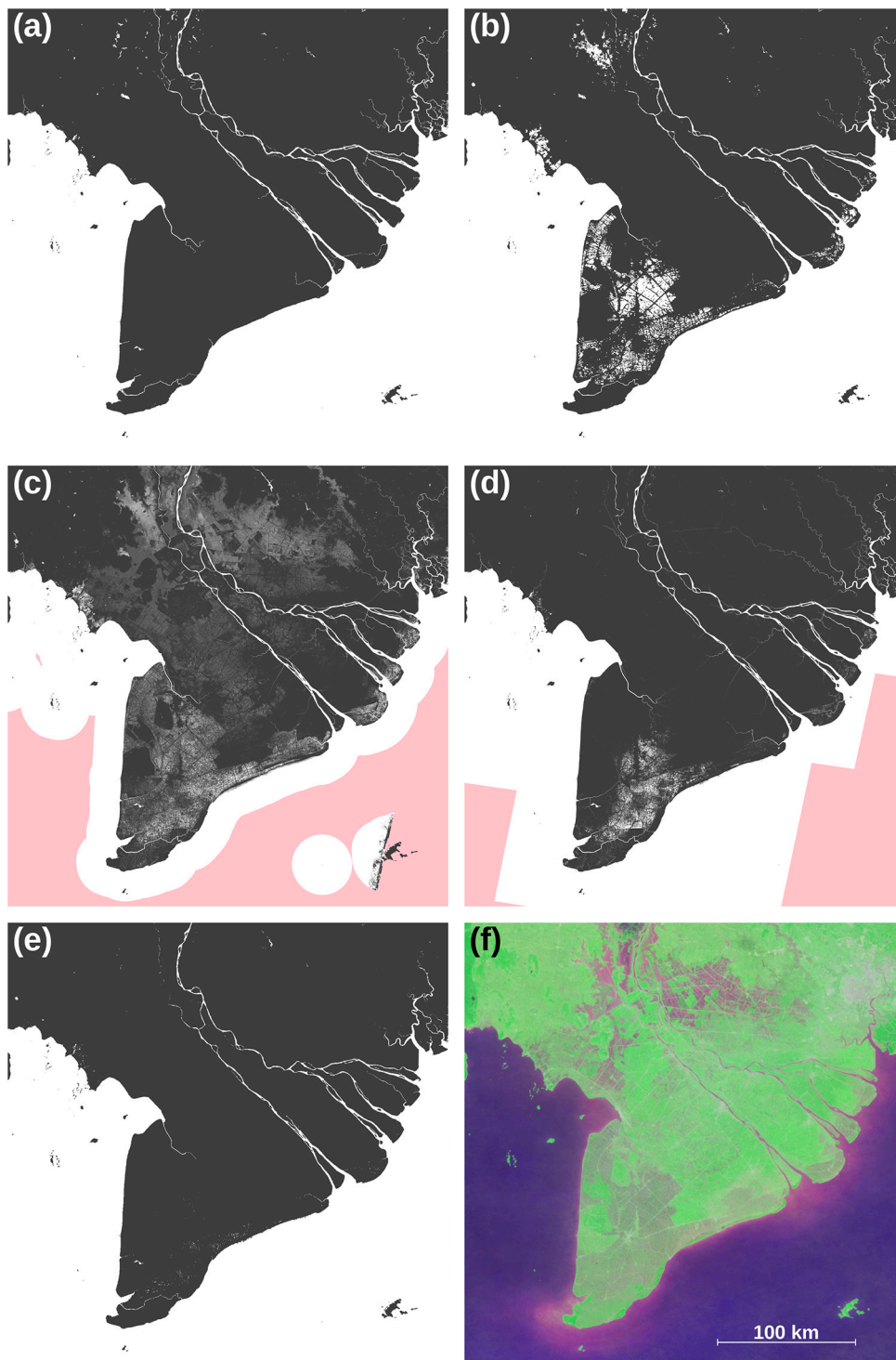
Fig. 3. Area of the Mekong River delta in (a) MOD44Wv5 land mask, (b) MOD44Wv6 land mask includes many seasonal bodies of water, such as rice fields, (c) GSW seasonality data indicate the number of months per year water was detected in the particular area (indicated by varied shades of gray), yet showing artifacts around the island in the lower right corner of the sample area, (d) GFC high-spatial resolution data also provide a water fraction in the medium-spatial resolution data, yet are also biased toward water, and completely miss the island in the lower right corner, (e) the combined water mask, and (f) false color clear sky imagery. Pink shades in panels (c and d) indicate gaps in the corresponding data sources. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

around the islands in the southeast part of the area. On the other hand, the GFC data (Fig. 3d) do not contain information on the seasonal variation, and thus show fewer inland waters, but also completely miss the islands in the southeast part of the area. Fig. 3e shows the combined land mask, which also excludes most seasonal inland water bodies, yet does not contain artifacts for the islands. The panel (f) in Fig. 3 shows the false color clear sky imagery of the same area from 2018. Here, the surface waters appear in purple, while land (mostly covered by vegetation) is bright green. As indicated above, the darker purple areas are only indicative of water presence during some time of 2018, while lighter green color indicates continuous land cover throughout this year.

We have used the clear sky true color and false color imagery for a global evaluation and validation of the new combined land mask. In addition, we have also compared them to all the land mask data sources used in this study, as well as examined the differences between the new combined land mask and each data source. In some places, mostly in the coastal areas of China, we found that the most recent GSW data set matches better with the false color clear sky imagery used in validation, and in these areas we used the corresponding GSW land-water indicator Λ^{GSW} directly to derive the land mask.

3.1.1. Global overall evaluation

We show the new combined land mask at a global scale in Fig. 4a, and highlight the differences from the MOD44Wv5 land mask. Here, the areas interpreted as land (or ice) in both data sets are displayed in dark gray. The areas where the MOD44Wv5 data set indicates water, but the combined land mask indicates land (or ice) are shown in yellow. There are some areas where MOD44Wv5 indicates land, but the combined land mask indicates water, and those are shown in magenta color. Most of the differences come from the polar regions and arid inland areas with seasonal waters. It is clear that the MOD44Wv5 land mask has fewer land pixels, especially around Antarctica (Fig. 4c). Here, the extent of land in the MOD44Wv5 data set is based on the bedrock elevation data, and thus does not include large areas permanently covered by ice. However, this difference does not reflect a deficiency, but rather a different purpose for the MOD44Wv5 data set.

3.1.2. Evaluation over global coastal regions

In order to evaluate the performance of the updated land mask in coastal regions, we show some regional examples comparing the combined global land mask with the MOD44Wv5 land mask, and the clear sky false color imagery from 2019 in Fig. 5. The first example (Fig. 5a) shows the MOD44Wv5 land mask for the Atlantic coast of Georgia and South Carolina, with coastal wetlands clearly identified as water. The updated land mask (Fig. 5b) treats most of the coastal wetlands as land, and that is confirmed by the corresponding false color imagery in Fig. 5c. Although parts of these wetlands are indeed covered by water, significant presence of vegetation means that ocean color retrievals are nearly always contaminated by the high reflectance signal from the vegetation.
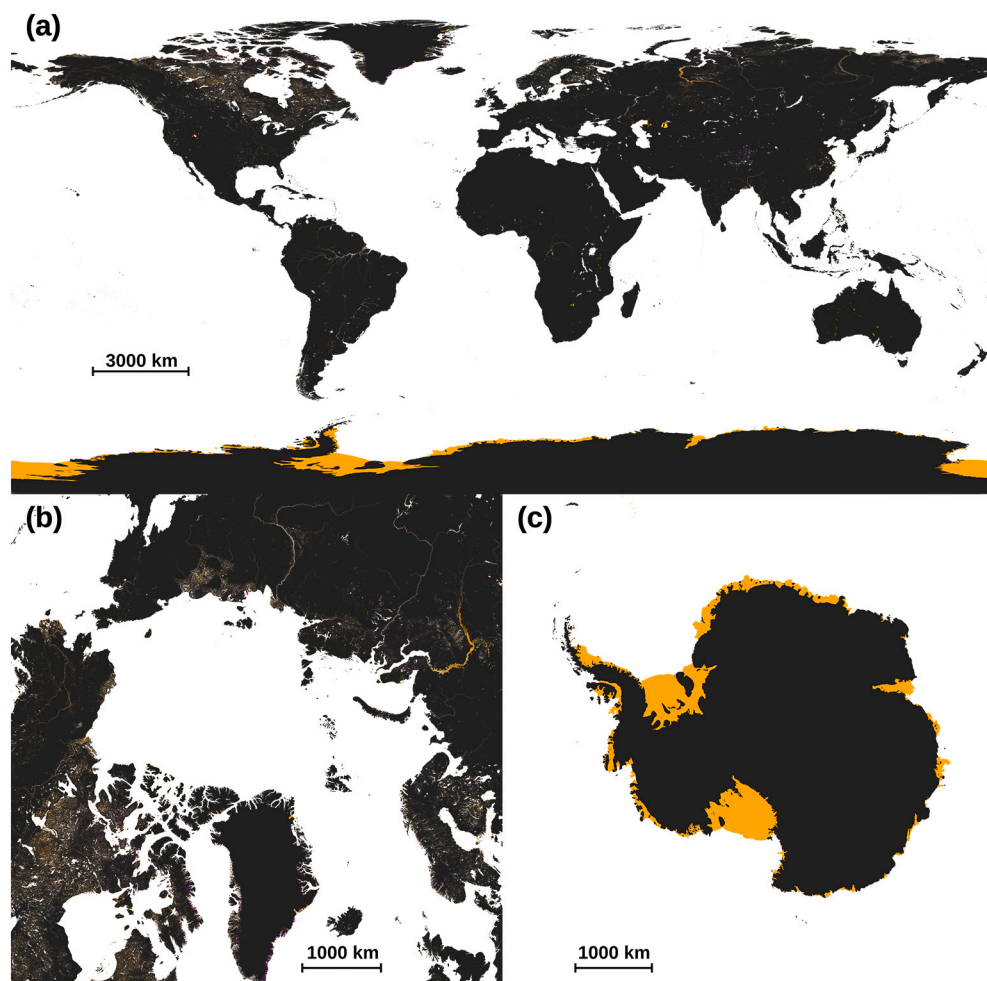
The coastline of the Bohai Sea has undergone many recent changes, most of which are not reflected in the MOD44Wv5 data (Fig. 5d). The new combined land mask (Fig. 5e) captures recent changes to the coastline and the reclaimed lands, as confirmed by the corresponding clear sky false color imagery in Fig. 5f.

Fig. 5g–i shows the southeastern part of Greenland. The updated land mask (Fig. 5h) shows longer fjords due to receding glaciers (as compared to the MOD44Wv5 land mask in Fig. 5g), again confirmed by the false color imagery (Fig. 5i).

In the example showing the Ganges River Delta region (Fig. 5j–l), the MOD44Wv5 land mask (Fig. 5j) overestimates the extent of the river as well as the many channels. The updated land mask (Fig. 5k) is more consistent with the false color imagery (Fig. 5l). This area too is subject to rapid changes of the coastline due to a large amount of sediment carried by the river and deposited in the delta region.

A more complete comparison of all the land mask data sources and the derived land mask for these regional examples is provided in the Supplementary Material (Figs. S1–S4). In addition to the cases shown in Fig. 5, we also note that MOD44Wv5 land mask tends to miss several islands in the Pacific (e.g., the Fatu Hiva Island in Marquesas), Arctic (e. g., the Henrietta Island), and Antarctic (the entire chain of the South Sandwich Islands).

3.1.3. Evaluation over global inland waters

We also perform a similar evaluation of the updated land mask for the inland waters, and show some examples in Fig. 6. In the first example (Fig. 6a), the MOD44Wv5 land mask data show a much larger extent for the Aral Sea that could have been accurate decades ago before the

Fig. 4. (a) The global image of the combined land mask including differences from the MOD44Wv5 land mask. Dark gray areas denote land (or ice) in both data sets; yellow color denotes areas identified as water in MOD44Wv5, but land (or ice) in the combined land mask, and magenta color denotes land in MOD44Wv5, but water in the combined land mask. While both land masks are mostly in agreement, the combined land mask identifies more land (or ice) in the polar regions. Panels (b) and (c) show the Arctic and Antarctic regions, respectively, including the differences from MOD44Wv5 data set with the same color coding. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
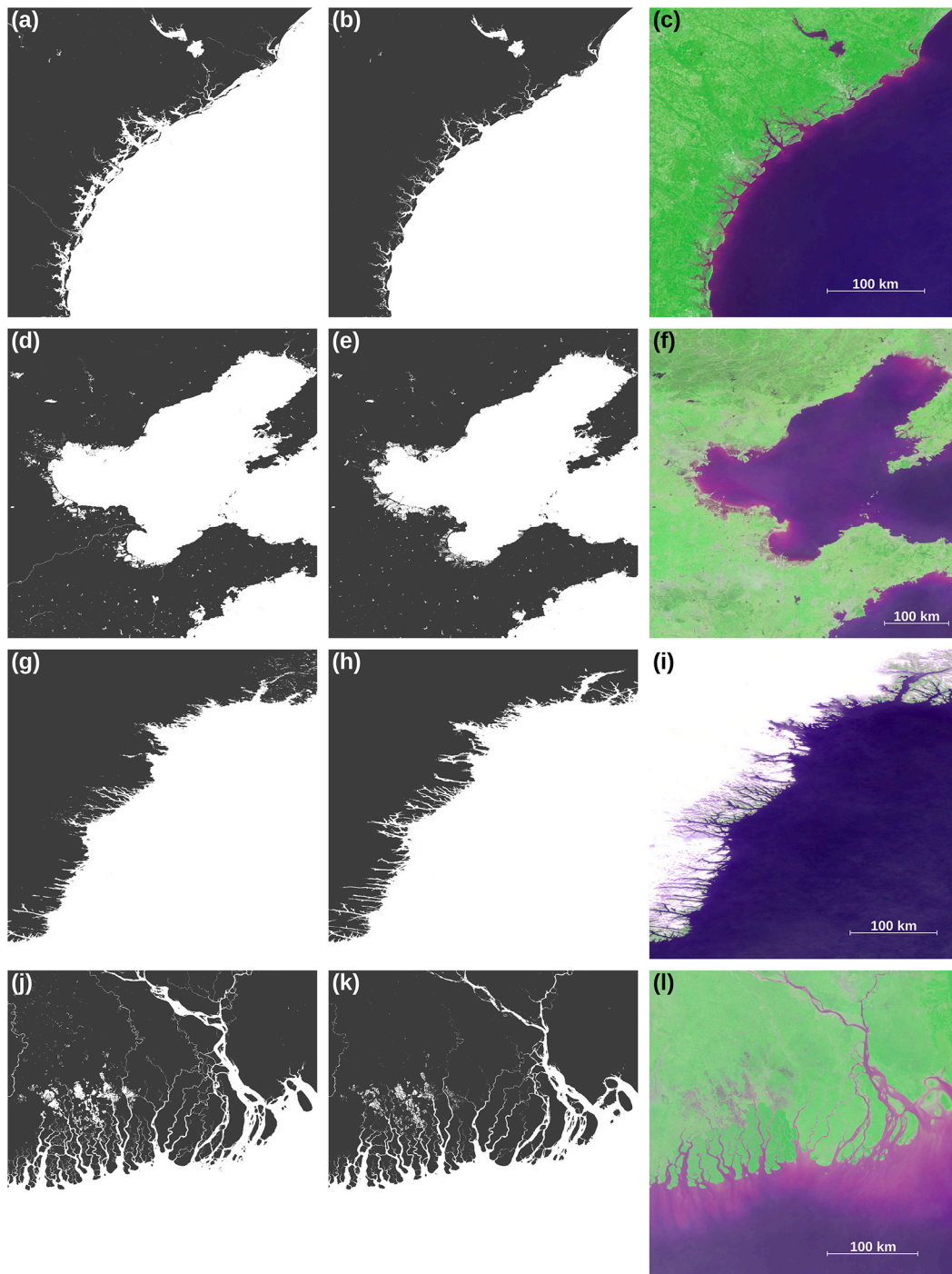
Fig. 5. Comparison of MOD44Wv5 land mask (left column), the new combined land mask (center column), and the clear sky false color imagery (right column) derived from 2019 data for (a)–(c) Atlantic coast of the southeast US (Georgia, South Carolina), (d)–(f) changes to the coastline of the Bohai Sea due to human development, (g)–(i) receding glaciers on the southeast coast of Greenland, and (j)–(l) changes to the coastline in Ganges Delta area.

dramatic shrinkage. The new combined land mask (Fig. 6b) is more accurate, and also matches quite well with the clear sky false color imagery from 2019 (Fig. 6c). It should be noted that the Aral Sea is still subject to frequent changes of water surface area, and cannot be accurately described by a single static data set.

Fig. 6d shows the MOD44Wv5 land mask for the floodplains of the Ob River in Russia, which appears to exaggerate the extent of the river water surface. In contrast, the new combined land mask (Fig. 6e) is more consistent with the recent clear sky false color imagery (Fig. 6f).

The surface area of many lakes on the Tibetan Plateau has increased

over the last few decades (Song et al., 2014), and that is reflected in the updated land mask (Fig. 6h), and the false color imagery (Fig. 6i) for this region, as compared to the older MOD44Wv5 land mask (Fig. 6g).

In the last example, the MOD44Wv5 land mask appears to exaggerate the extent of the inland waters in the western United States (Fig. 6j), while the new combined land mask (Fig. 6k) is more consistent with a recent clear sky false color imagery (Fig. 6l). The MOD44Wv5 land mask also misclassifies the ice and fields of solidified lava over the North and South Sister volcanoes as water (Fig. 6j, upper left). Similar misclassification happens for other volcanoes in Kamchatka, Alaska, New
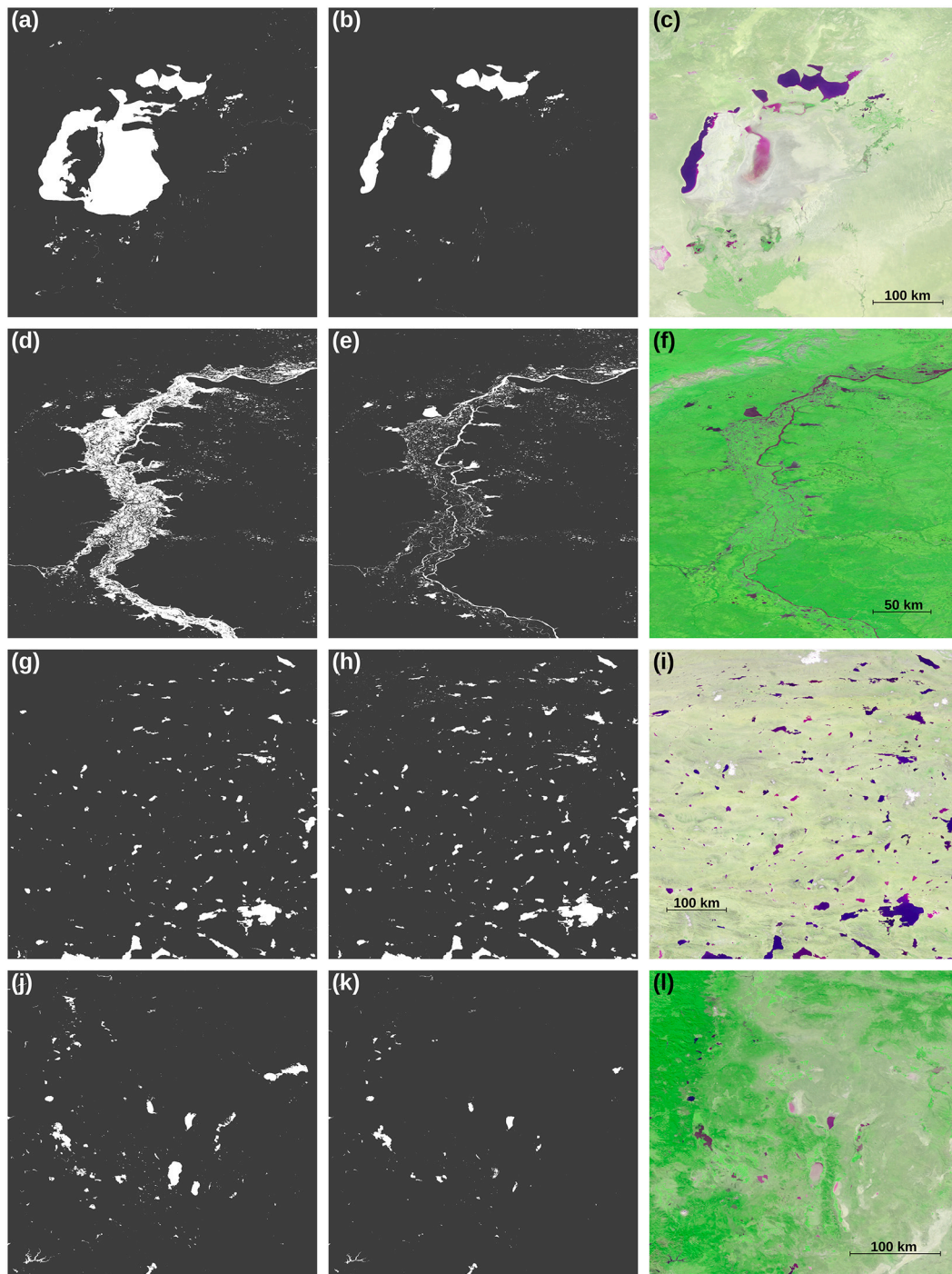
Fig. 6. Comparison of MOD44Wv5 land mask (left column), the new combined land mask (center column), and the clear sky false color imagery (right column) derived from 2019 data for (a)–(c) the Aral Sea which has shrunk considerably from the area shown in MOD44Wv5 data, (d)–(f) overestimation of the surface area in Ob River branching with MOD44Wv5 land mask, (g)–(i) increased surface area in the lakes on the Tibetan Plateau, and (j)–(l) decreased extent of lakes in the western US.

Zealand, and other places.

A more complete comparison of all the land mask data sources and the derived land mask for these regional examples is provided in the Supplementary Material (Figs. S5–S8). In addition to the types of changes shown in Fig. 6, older MOD44Wv5 land mask also does not include many recently created water reservoirs, such as the Indira Sagar Reservoir (India), the Alqueva Lake (Portugal), the Murum Dam Reservoir (Malaysia), and others.

To provide quantitative comparison of all the input data sources, and the derived land masks, we show the statistics of the percentage of land pixels for all of the regional examples shown in Figs. 3, 5, and 6 in Table 2. As mentioned above, the OSM land mask is only used for the polar areas, thus only OSM data for southeast Greenland are shown. The GFC land mask does not include the polar areas, thus no data for this area are available. As seen from Table 2, the MOD44Wv5 and MOD44Wv6 land masks underestimate the extent of land. Unsurprisingly, the percentage of land pixels in the new derived land mask matches the closest to the GSW data, which has the highest weight in

Table 2
Land pixel percentages for the regional examples corresponding to Figs. 3, 5, and 6.

Region	MOD44Wv5	MOD44Wv6	GSW	GFC	OSM	New
Mekong Delta	34.931	33.733	35.467	35.359	–	35.299
Southeast US	69.592	69.578	70.956	71.179	–	70.920
Bohai Sea	60.035	59.924	61.070	61.183	–	60.814
Southeast Greenland	45.628	44.620	44.689	–	44.266	44.662
Ganges Delta	51.100	51.096	52.864	52.565	–	52.370
Aral Sea	92.679	95.830	97.633	96.604	–	96.916
River Ob	92.006	92.319	96.471	97.080	–	96.410
Tibetan Plateau	96.586	95.960	95.535	97.065	–	96.131
Western US	98.304	98.888	99.366	99.477	–	99.354

Table 1.

3.2. Evaluation of effect on the ocean color retrievals

3.2.1. Global and regional VIIRS-derived Chl-a data

For the evaluation of the effects of various land mask data sets on VIIRS-derived ocean color products, we use the MSL12 ocean color data processing system with the NIR and shortwave infrared (NIR-SWIR) combined atmospheric correction algorithm (Gordon and Wang, 1994; Jiang and Wang, 2014; Wang, 2007; Wang and Shi, 2007). Specifically, we compare VIIRS-SNPP-derived global and regional Chl-a data (Wang and Son, 2016) using various land mask inputs in MSL12, and evaluate Chl-a data coverage and corresponding data quality. It is particularly noted that in the VIIRS ocean color data processing all are the same except for the inputs of land mask data sets. Therefore, differences in VIIRS-SNPP-derived Chl-a data are all due to the use of different land mask data in the MSL12 ocean color data processing. It should also be noted that, through all these years, VIIRS-SNPP ocean color products have been well evaluated and validated, showing high quality data over global open oceans and reasonable retrievals over coastal and inland waters (Barnes et al., 2019; Hlaing et al., 2013; Mikelsons et al., 2020; Wang et al., 2020).

3.2.2. Evaluation of global ocean color data

In order to evaluate the effect of using various land mask choices on the ocean color retrievals, we have processed VIIRS-SNPP data using the MSL12 ocean color data processing system (Wang et al., 2013) and derived the Level-2 Chl-a concentration data (Hu et al., 2012; O'Reilly et al., 1998; O'Reilly and Werdell, 2019; Wang and Son, 2016) with three choices of land mask: (1) MOD44Wv5 land mask used up till present in MSL12 operational ocean color retrievals, (2) MOD44Wv6 land mask, and (3) the new combined land mask developed in this study. We have processed data for the 24 days evenly scattered throughout 2019 for a better representative sample. The quantitative comparison of number of valid global ocean color retrievals is shown in Table 3. Since the changes of the land mask mostly affect only the coastal and inland waters, we also show the results for retrievals with water depth less than 1 km and 250 m.

We note that MOD44Wv6 produces more retrievals due to added water pixels in land mask (positive values in "Change of water pix." column), however, it loses an even larger number due to attempted retrievals over land pixels misclassified as water by this land mask (indicated by negative values in "Change of masked pix." column). The new

combined land mask yields more ocean color retrievals, both due to more water pixels and fewer misclassified land pixels. While the improvement seems relatively small relative to the total number of global retrievals, it becomes more significant for shallower coastal and inland waters. In fact, for regions with water depth < 250 m, the total number of pixels changed with the new land mask is about 3.0×10^5, which may be quite significant for water property data over regional coastal and inland waters.

3.2.3. Ocean color products from global coastal regions

A visual comparison of ocean color retrievals for four selected scenes is shown in Fig. 7: the left column corresponds to the MOD44Wv5 land mask, the center corresponds to the MOD44Wv6 land mask, and imagery in the right column is obtained with the new combined land mask.

Fig. 7a shows VIIRS-SNPP Chl-a retrievals obtained with the MOD44Wv5 land mask from the Georgia and the South Carolina (US) Atlantic coast recorded acquired on April 15, 2019, at 18:42 UTC. Despite large areas of wetland classified as water, few ocean color retrievals are recorded for those areas. In addition, some of the retrievals near shore are also masked out by the cloud shadow and stray light flag (Jiang and Wang, 2013). Both of these artifacts are due to misclassification of land or vegetation rich wetland areas as water by the MOD44Wv5 land mask. As a result, the high reflectances at the misclassified pixels are interpreted by MSL12 as due to clouds, and masked out. In addition, much larger surrounding areas are also masked out due to assumed likely contamination by straylight and cloud shadows (Hu et al., 2020; Jiang and Wang, 2013), thus reducing the number of valid ocean color retrievals. The same effect is also evident in Fig. 7b, showing VIIRS-derived Chl-a using the MOD44Wv6 land mask, which interprets even larger wetland areas as water pixels. However, Chl-a retrievals obtained with the new combined land mask (Fig. 7c) are more consistent, providing more retrievals over the wetlands and fewer data gaps overall.

Fig. 7d–f shows the VIIRS-SNPP-derived Chl-a data from February 15, 2019, at 12:09 UTC, obtained with the three land masks for the North Sea. Again, while using either the MOD44Wv5 (Fig. 7d) or MOD44Wv6 (Fig. 7e) land mask results in several areas of valid retrievals masked out due to perceived cloud contamination, fewer such cases are visible in retrievals with the updated land mask in Fig. 7f.

Fig. 7g shows VIIRS-SNPP Chl-a retrievals from the Bohai Sea on September 1, 2019, obtained with the MOD44Wv5 land mask. Since the MOD44Wv5 land mask is somewhat outdated, it does not capture the coastal land reclamation developments in the area over the last two

Table 3
Comparison of number of retrievals with MOD44Wv5, MOD44Wv6, and the new combined land masks for different maximum water depths in global ocean and inland waters.

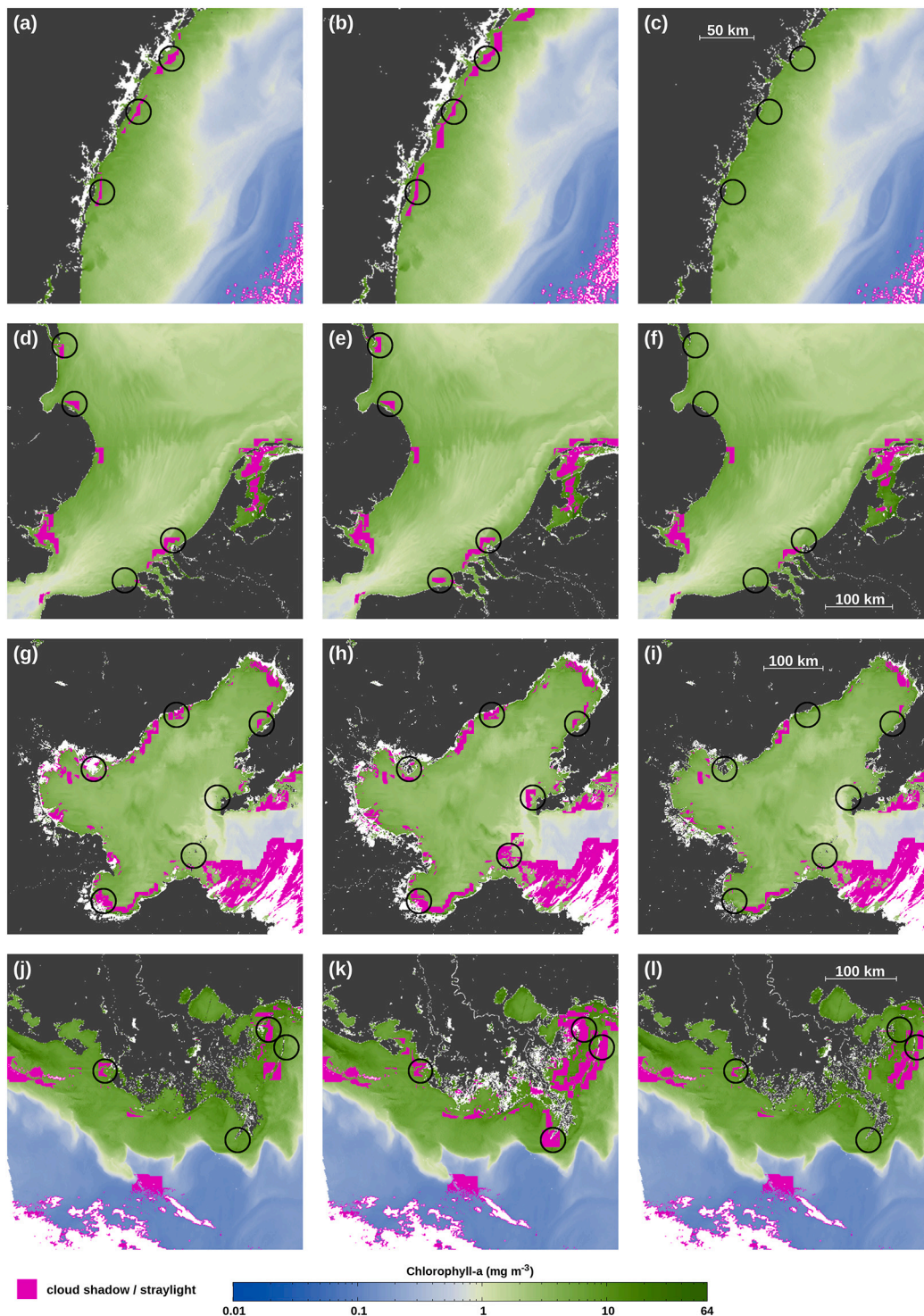| Water depth	MOD44Wv5	MOD44Wv6			Combined		
	Total retrievals	Relative change	Change of water pix.	Change of masked pix.	Relative change	Change of water pix.	Change of masked pix.
All	$1.59 \cdot 10^9$	−0.118%	+0.013%	−0.131%	+0.021%	+0.005%	+0.016%
< 1000 m	$2.14 \cdot 10^8$	−0.825%	+0.099%	−0.924%	+0.169%	+0.039%	+0.130%
< 250 m	$1.53 \cdot 10^8$	−1.032%	+0.138%	−1.170%	+0.194%	+0.055%	+0.139%

Fig. 7. Comparison of Chl-a retrievals in coastal areas from VIIRS-SNPP measurements using different land mask data sets. Land is displayed in dark gray, missed retrievals due to cloud shadow/straylight flag in magenta, and no retrievals over water due to other conditions (clouds, sun glint, etc.) in white. Some of the differences in retrievals are emphasized by the black circles. Left column: using the MOD44Wv5 land mask; center column: using the MOD44Wv6 land mask, and right column: using the new combined land mask data set. Scene identification: (a)–(c) Georgia and South Carolina coast, April 15, 2019, 18:42 UTC; (d)–(f) the North Sea, February 15, 2019, 12:09 UTC; (g)–(i) the Bohai Sea, September 1, 2019, 05:06 UTC; and (j)–(l) the Mississippi delta, May 15, 2019, 19:20 UTC. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

decades, and thus misclassifies many land pixels as water. This produces many failed retrievals with the MOD44Wv5 land mask, resulting in even larger areas masked out due to perceived cloud contamination (Wang and Shi, 2006) (Fig. 7g). Although newer, the MOD44Wv6 land mask is actually even more biased toward water, and thus produces even more misclassified pixels (Fig. 7h), resulting in even larger masked out areas. In contrast, Chl-a retrievals obtained with the more up-to-date combined land mask (Fig. 7i) show fewer data gaps in coastal areas.

The last scene, Fig. 7j–l shows Chl-a retrievals from May 15, 2019, in the Mississippi Delta, a complex coastal region with highly turbid, but also shallow waters. Here, the loss of ocean color retrievals due to pixel misclassification by the MOD44Wv6 land mask is especially evident and very significant in the area east from the delta (Fig. 7k). Comparing the retrievals with the MOD44Wv5 and the new combined land masks in Fig. 7j and l, respectively, we see both of these land masks result in some, though much smaller, loss of retrievals in different areas. This illustrates the overall difficulty of deriving a single land mask data set in such complex coastal areas. A more complete comparison with the corresponding true and false color imagery for the examples in Fig. 7 is provided in the Supplementary Material (Fig. S9).

3.2.4. Ocean color products from global inland waters

In addition to coastal waters, we also perform a similar comparison and evaluation of ocean color retrievals obtained with the three land mask choices for various inland bodies of water.

Fig. 8a–c shows the differences in the land masks and the ocean color retrievals for the Aral Sea observed by VIIRS-SNPP on May 1, 2019. The MOD44Wv5 land mask shows the extent of the Aral Sea before the large scale depletion due to human activities. On the other hand, the MOD44Wv6 land mask (Fig. 8b) shows a much smaller and partitioned Aral Sea, yet still overestimates its present extent. Both of these land masks result in failed and masked out Chl-a retrievals due to land/water misclassifications. In comparison, the new updated land mask (Fig. 8c) shows a more realistic picture of the present state of the Aral Sea and fewer failed or masked out Chl-a retrievals. We note again that the exact extent of the Aral Sea changes seasonally and interannually, and thus may not be exactly described by a static land mask.

A similar effect is seen in Fig. 8d–f, showing Chl-a retrievals derived with the three land masks from VIIRS-SNPP data over the northeast Caspian Sea recorded on September 15, 2019. Again, the MOD44Wv5 and MOD44Wv6 land masks significantly overestimate the extent of the
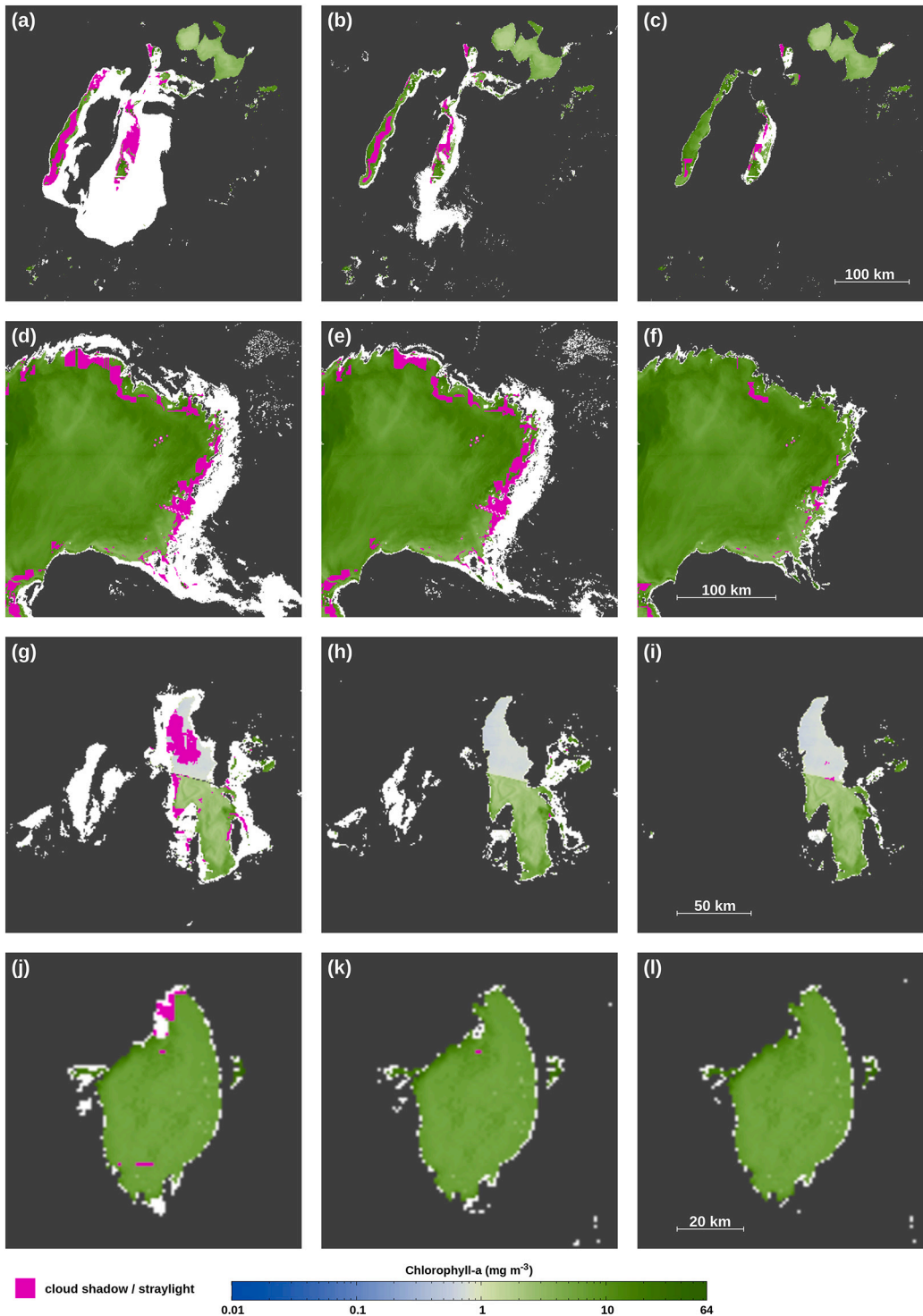

Fig. 8. Chl-a retrievals from VIIRS-SNPP measurements using the MOD44Wv5 land mask (left column) and the MOD44Wv6 land mask (center column), compared with the new combined land mask (right column). The land is displayed in dark gray, missed Chl-a retrievals due to cloud shadow and straylight flag in magenta, and no retrievals over water due to other conditions (clouds, sun glint, etc.) in white. Scene identification: (a)–(c) Aral Sea, May 1, 2019, 08:38 UTC; (d)–(f) northeast Caspian Sea, September 15, 2019, 09:10 UTC; (g)–(i) the Great Salt Lake, September 1, 2019, 20:21 UTC; and (j)–(l) Lake Okeechobee, April 15, 2019, 18:41 UTC. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

water surface and result in many failed or masked out Chl-a retrievals (Fig. 8d and e). The updated land mask (Fig. 8f) is more reflective of the present extent of the water surface and produces more consistent Chl-a retrievals with fewer data gaps.

Fig. 8g–i shows a similar comparison of Chl-a retrievals for the Great Salt Lake in the US, produced from VIIRS-SNPP observations recorded on September 1, 2019. The MOD44Wv5 land mask again overestimates the extent of the lake (Fig. 8g), and results in failed and masked out Chl-a retrievals. The MOD44Wv6 land mask actually represents the present extent of the Great Salt Lake quite well (Fig. 8h), but it also includes several seasonal bodies of water around it, which in recent years are rarely covered by water. The new combined land mask (Fig. 8i) likewise produces consistent retrievals over the Great Salt Lake, and does not include areas that are rarely covered by water.

In the last example, we show the comparison of Chl-a retrievals for Lake Okeechobee in Florida acquired on April 15, 2019. We note that slight changes in the land mask from MOD44Wv5 (Fig. 8j) to MOD44Wv6 (Fig. 8k) may result in a noticeable change in Chl-a retrievals in the north part of the lake. In this case, both MOD44Wv6 and the new updated land mask (Fig. 8l) produce the most consistent Chl-a retrievals. A more complete comparison with the corresponding true and false color imagery for the examples in Fig. 8 is provided in the Supplementary Material (Fig. S10).

In addition to the multiple scenes of imagery, we also show a quantitative comparison of the number of Chl-a retrievals derived using the MOD44Wv5 land mask and the new combined land mask for a few selected coastal and inland water regions in Table 4:

The data in Table 4 are based on the same 24 evenly distributed days in the year 2019. The geographical extent is defined by the northwest and southeast corners for a rectangular area in geographical projection. We see that the new combined land mask yields significantly more Chl-a retrievals in the coastal of the East China Sea, as well as for the area around the Aral Sea, and the lakes on the Tibetan Plateau. However, it provides slightly fewer Chl-a retrievals in the Mississippi Delta and the Amazon River basin. Both land masks are very close for the US East Coast region and the Great Lakes, and there is little change in the number of Chl-a retrievals for these areas.

4. Discussion

The method proposed and employed in this work relies on using multiple independent land mask data sources to eliminate the artifacts in the new combined land mask data set. The assumption is that in the independent land mask datasets, any artifacts are not correlated and likely to occur in different areas, and thus can be eliminated in the combined land mask data by combining the data sources according to some weighted majority rule. This assumption is mostly valid in all areas except for the polar regions where additional (OSM) dataset is required due to the lack of the data in most other data sources. Overall, the GSW data set is the most recent and most comprehensive. However, it too has some artifacts and areas of missing data. In general, combining a recent and high-quality land mask data set with older and outdated data sets may degrade the quality of the combined data set, as compared to the high quality one, especially in the areas where the latter is free from

artifacts. We have attempted to mitigate this by assigning higher weight to the GSW data, as compared to the other data sources.

Since some of the data sources have been derived from the same satellite observations, there may be a concern as to what extent are the data sources truly independent. As we have noted, and shown in many examples, the land/water masks derived from the same satellite data tend to differ significantly due to differences in the methodology used to derive them. However, using data derived from the same satellite measurements may still include some sampling bias, due to particular spatial-temporal coverage of satellite observations. Including additional data sources derived from other satellite measurements would allow to mitigate this concern.

Combining or assimilating multiple land mask data sources requires some adjustable parameters. We note again that the choice of such parameters depends on the application, and no single land mask dataset can completely describe the dynamics of the global surface waters. The new combined land mask data set in this work was derived mostly for the use with medium spatial resolution satellite ocean color measurements, though it is expected to be an improvement for other types of satellite-based measurements of ocean properties. We find that for the ocean color measurements, false (mis)identification of land as water is rather costly as it results in masking out larger areas by the ocean color retrieval algorithm due to perceived cloud contamination (Hu et al., 2020; Jiang and Wang, 2013). Similar considerations may be valid for other types of water properties measurements. On the other hand, the measurements of various land properties would likely require avoiding the contamination due to water surfaces.

We note that for ocean color remote sensing, most of the changes to the number of satellite retrievals arise due to excessive cloud shadow and stray light masking when a land pixel is misidentified as water (Hu et al., 2020; Jiang and Wang, 2013). Thus, the optimal land mask for maximal number of retrievals indirectly depends on the details of the cloud shadow and straylight masking algorithms used.

It should be mentioned that the design of a land mask (or water mask) data set involves a compromise between the need to maximize the water surface subject to ocean color measurements, and the requirement to avoid land contamination. Due to seasonal changes of the water surface area, it is not possible to completely satisfy both of these requirements in a single static data set. Thus, some seasonal inland waters may be lost in a static land mask, including the new land mask.

Overall, we find that the new combined land mask data set derived in this study is a significant improvement over the aging MOD44Wv5 land mask, especially in the areas with recent changes, such as coastal areas altered by development and land reclamation, and artificially created new inland water reservoirs. The new combined land mask data set scores even better compared to the comparatively recent MOD44Wv6 data, mostly due to overrepresentation of the seasonal inland waters in the latter.

Naturally, this work can be further extended by incorporating the seasonal variation into a consistent global land mask data set. However, at present, the GSW is the only data set known to us to include the water surface seasonality. Additionally, the present version is not truly global (as it excludes the Antarctic region), and also it is not free from artifacts. Thus, other data sources (such as false color imagery) are likely needed

Table 4

Comparison of number of retrievals with MOD44Wv5 and the combined land masks for selected coastal and inland regions.

Region	Geographical extent	MOD44Wv5 retrievals	Combined land mask retrievals	Change (%)
Amazon River	(1°N, 73°W) – (7°S, 49°W)	99,846	99,190	−0.65%
Aral Sea	(47°N, 56°E) – (41°N, 63°E)	130,727	154,002	+17.8%
East China Sea	(41°N, 117°E) – (31°N, 127°E)	3,489,260	3,542,822	+1.5%
Great Lakes	(49°N, 92.2°W) – (41.3°N, 76°W)	1,536,649	1,536,960	−0%
Mississippi Delta	(31°N, 95°W) – (28°N, 85°W)	912,297	908,186	−0.5%
Tibetan Plateau	(36.5°N, 79°E) – (29°N, 93°E)	137,602	154,430	+12.2%
US East Coast	(40°N, 77.5°W) – (34°N, 75°W)	592,640	592,561	−0.01%

for such a task. An updated version of the GSW data set with complete global coverage and minimized artifacts would be a great resource for ocean color remote sensing.

Likewise, it is straightforward to extend the methodology in this work to produce an updated higher spatial resolution global land mask using the existing data sources. However, with fewer high spatial resolution data sources and more spatial detail, it may be more difficult to ensure the consistency and quality of such results.

5. Conclusions

We have introduced a method to assimilate multiple independent land- or water-mask data sets into a combined land mask, whereby minimizing the artifacts in each data set, incorporating changes from more recent data sets, and increasing the global coverage. We have used this method to derive an updated, more complete, and more consistent land mask data set for the satellite ocean color remote sensing. The quality of the new combined land mask is evaluated and validated using the comparison with global clear sky true color and false color imageries, and by evaluating the effect of the updated land mask data set on the global ocean color retrievals. Though this work is focused on improving the global land mask data specifically for the satellite ocean color retrievals, the methodology can be applied to derive land- or water-mask data for different areas of study.

Author contribution statement

Karlis Mikelsons: Conceptualization, methodology, software, validation, investigation, writing - original draft, writing - review & editing, visualization. **Menghua Wang**: Conceptualization, writing - original draft. **Xiao-Long Wang**: Validation, investigation, writing - review & editing. **Lide Jiang**: Conceptualization, data curation, writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rse.2021.112356.

References

Barnes, B.B., Cannizzaro, J.P., English, D.C., Hu, C., 2019. Validation of VIIRS and MODIS reflectance data in coastal and oceanic waters: an assessment of methods. Remote Sens. Environ. 220, 110–123.

Carroll, M.L., Townshend, J.R., DiMiceli, C.M., Noojipady, P., Sohlberg, R.A., 2009. A new global raster water mask at 250 m resolution. Int. J. Digit. Earth 2, 291–308.

Carroll, M.L., DiMiceli, C.M., Townshend, J.R., Sohlberg, R.A., Elders, A.I., Devadiga, S., Sayer, A.M., Levy, R.C., 2017. Development of an operational land water mask for MODIS collection 6, and influence on downstream data products. Int. J. Digit. Earth 10, 207–218.

Donlon, C., Berruti, B., Buongiorno, A., Ferreira, M.-H., Femenias, P., Frerick, J., Goryl, P., Klein, U., Laur, H., Mavrocordatos, C., Nieke, J., Rebhan, H., Seitz, B., Stroede, J., Sciarra, R., 2012. The global monitoring for environment and security (GMES) Sentinel-3 mission. Remote Sens. Environ. 120, 37–57. https://doi.org/10.1016/j.rse.2011.07.024.

Esaias, W.E., Abbott, M.R., Barton, I., Brown, O.B., Campbell, J.W., Carder, K.L., Clark, D.K., Evans, R.L., Hodge, F.E., Gordon, H.R., Balch, W.P., Letelier, R., Minnet, P.J., 1998. An overview of MODIS capabilities for ocean science observations. IEEE Trans. Geosci. Remote Sens. 36, 1250–1265.

Goldberg, M. D., Kilcoyne, H., Cikanek, H., and Mehta, A. (2013), Joint polar satellite system: the United States next generation civilian polar-orbiting environmental satellite system, J. Geophys. Res. Atmos., 118, 13463–13475, doi:13410.11002/12013JD020389.

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

Hale, G.M., Querry, M.R., 1973. Optical constants of water in the 200nm to 200μm wavelength region. Appl. Opt. 12, 555–563.

Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O., Townsend, J.R.G., 2013. High-resolution global maps of 21st-century forest cover change. Science 342, 850–853.

Hlaing, S., Harmel, T., Gilerson, A., Foster, R., Weidemann, A., Arnone, R., Wang, M., Ahmed, S., 2013. Evaluation of the VIIRS Ocean color monitoring performance in coastal regions. Remote Sens. Environ. 139, 398–414.

Hu, C., Lee, Z., Franz, B.A., 2012. Chlorophyll a algorithms for oligotrophic oceans: A novel approach based on three-band reflectance difference. J. Geophys. Res. 117, C01011 https://doi.org/10.01029/02011JC007395.

Hu, C., Barnes, B.B., Feng, L., Wang, M., Jiang, L., 2020. On the interplay between ocean color data quality and data quantity: impacts of quality control flags. IEEE Geosci. Remote Sens. Lett. 17, 745–749. https://doi.org/10.1109/lgrs.2019.2936220.

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

Jiang, L., Wang, M., 2014. Improved near-infrared ocean reflectance correction algorithm for satellite ocean color data processing. Opt. Express 22, 21657–21678. https://doi.org/10.1364/OE.22.021657.

Li, S., Sun, D., Goldberg, M.D., Stefanidis, A., 2013. Derivation of 30-m-resolution water maps from TERRA/MODIS and SRTM. Remote Sens. Environ. 134, 417–430.

Mikelsons, K., Wang, M., 2018. Interactive online maps make satellite ocean data accessible. Eos Trans. AGU 99. https://doi.org/10.1029/2018EO096563.

Mikelsons, K., Wang, M., Jiang, L., 2020. Statistical evaluation of satellite ocean color data retrievals. Remote Sens. Environ. 237, 111601. https://doi.org/10.1016/j.rse.2019.111601.

O'Reilly, J.E., Werdell, P.J., 2019. Chlorophyll algorithms for ocean color sensors - OC4, OC5 & OC6. Remote Sens. Environ. 229, 32–47.

O'Reilly, J.E., Maritorena, S., Mitchell, B.G., Siegel, D.A., Carder, K.L., Garver, S.A., Kahru, M., McClain, C.R., 1998. Ocean color chlorophyll algorithms for SeaWiFS. J. Geophys. Res. 103, 24937–24953.

Pekel, J.F., Cottam, A., Gorelick, N., Belward, A.S., 2016. High-resolution mapping of global surface water and its long-term changes. Nature 540, 418–422.

Qi, L., Hu, C., Mikelsons, K., Wang, M., Lance, V., Sun, S., Barnes, B.B., Zhao, J., Zande, D.V., 2020. In search of floating algae and other organisms in global oceans and lakes. Remote Sens. Environ. 239, 111659. https://doi.org/10.1016/j.rse.2020.111659.

Salomonson, V.V., Barnes, W.L., Maymon, P.W., Montgomery, H.E., Ostrow, H., 1989. MODIS: advanced facility instrument for studies of the earth as a system. IEEE Trans. Geosci. Remote Sens. 27, 145–153.

Shi, W., Wang, M., Jiang, L., 2011. Spring-neap tidal effects on satellite ocean color observations in the Bohai Sea, Yellow Sea, and East China Sea. J. Geophys. Res. 116, C12032 https://doi.org/10.1029/2010JC007234.

Shi, W., Wang, M., Jiang, L., 2013. Tidal effects on ecosystem variability in the Chesapeake Bay from MODIS-aqua. Remote Sens. Environ. 138, 65–76. https://doi.org/10.1016/j.rse.2013.07.002.

Song, C., Huang, B., Ke, H., Richards, K.S., 2014. Remote sensing of alpine lake water environment changes on the Tibetan plateau and surroundings: a review. ISPRS J. Photogramm. Remote Sens. 92, 26–37.

Tanaka, K., Okamura, Y., Amano, T., Hiramatsu, M., Shiratama, K., 2009. Development status of the second-generation global imager (SGLI) on GCOM-C. In: Proc. SPIE, 7474, sensors, systems, and Next-Generation Satellites XIII, 74740N. https://doi.org/10.1117/12.830963.

Wang, M., 1999. A sensitivity study of SeaWiFS atmospheric correction algorithm: effects of spectral band variations. Remote Sens. Environ. 67, 348–359.

Wang, M., 2007. Remote sensing of the ocean contributions from ultraviolet to near-infrared using the shortwave infrared bands: simulations. Appl. Opt. 46, 1535–1547.

Wang, M., Franz, B.A. 2000. Comparing the ocean color measurements between MOS and SeaWiFS: a vicarious intercalibration approach for MOS. IEEE Trans. Geosci. Remote Sens. 38, 184–197.

Wang, M., Shi, W., 2006. Cloud masking for ocean color data processing in the coastal regions. IEEE Trans. Geosci. Remote Sens. 44, 3196–3205. https://doi.org/10.1109/tgrs.2006.876293.

Wang, M., Shi, W., 2007. The NIR-SWIR combined atmospheric correction approach for MODIS Ocean color data processing. Opt. Express 15, 15722–15733. https://doi.org/10.1364/oe.15.015722.

Wang, M., Son, S., 2016. VIIRS-derived chlorophyll-a using the ocean color index method. Remote Sens. Environ. 182, 141–149. https://doi.org/10.1016/j.rse.2016.05.001.

Wang, M., Isaacman, A., Franz, B.A., McClain, C.R., 2002. Ocean color optical property data derived from the Japanese Ocean color and temperature scanner and the French

polarization and directionality of the Earth's Reflectances: a comparison study. Appl. Opt. 41, 974–990.

Wang, M., Liu, X., Tan, L., Jiang, L., Son, S., Shi, W., Rausch, K., Voss, K., 2013. Impact of VIIRS SDR performance on ocean color products. J. Geophys. Res. Atmos. 118, 10347–10360. https://doi.org/10.1002/jgrd.50793.

Wang, M., Shi, W., Watanabe, S., 2020. Satellite-measured water properties in high altitude Lake Tahoe. Water Res. 178, 115839. https://doi.org/10.1016/j. watres.2020.115839.