Remote Sensing of Environment 225 (2019) 175–192

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Adaptation of the dark spectrum fitting atmospheric correction for aquatic applications of the Landsat and Sentinel-2 archives

Quinten Vanhellemont

Royal Belgian Institute of Natural Sciences, Operational Directorate Natural Environments, Vautierstraat 29, Brussels 1000, Belgium

ARTICLE INFO

Keywords:
Water colour
Water quality
Long time-series
Multi-sensor
Aerosol correction
Automation

ABSTRACT

The dark spectrum fitting (DSF) atmospheric correction method for aquatic application of metre-scale resolution optical satellite imagery is adapted to Landsat and Sentinel-2 (L/S2), including an automated tiled processing of full scene imagery and an optional image based glint correction. The DSF uses multiple dark targets in the subscene to construct a "dark spectrum" which is used to estimate the atmospheric path reflectance (ρ_{path}) according to the best fitting aerosol model. This method is fully automated and can be used for full mission archive processing, as demonstrated here for a study region in the North Sea. The new approach overcomes common issues with the SWIR based exponential extrapolation approach (EXP). An evaluation of both methods is presented using L_w measurements from 19 sites in the AERONET-OC network over a 15 year period and 5 satellite sensors. Overall, the DSF performs better than the EXP, with a notable improvement in the blue spectral region. The tiled processing allows for a smooth ρ_{path} estimation for full and merged L/S2 scenes, over clear and turbid coastal waters, inland waters, and land. The DSF selects the most appropriate band automatically, i.e. the one giving the lowest atmospheric path reflectance, and hence largely avoids amplification of glint and adjacency effects in the atmospheric correction. After application of the DSF, sun glint reflectance can be estimated from the SWIR bands, and the application of a sun glint correction significantly improves data availability for these nadir viewing sensors. A consistent processing across sensors allows for the exploitation of the > 30 year L/S2 archive, including Landsat 5 imagery dating back to 1984. A practical application of the DSF and the L/S2 archive is presented, where the remotely sensed water turbidity from 5 satellites is compared with in situ measurements from a long-term (2000–present) monitoring station in the southern North Sea.

1. Introduction

The launch of Landsat 8 (L8) in 2013 and Sentinel-2 (S2) A and B in 2015 and 2017 sparked the interest of both the water and land remote sensing communities with the unprecedented spatial, spectral, and temporal coverage combined with impressive radiometric quality and free data access. Several atmospheric correction (AC) algorithms were developed for land and water applications of these systems, as evidenced by the 14 algorithms participating in the L8 and S2 Atmospheric Correction Intercomparison Exercise (Doxani et al., 2018). The opening of the full Landsat archive (Woodcock et al., 2008) and subsequent consolidation efforts (Wulder et al., 2016), especially for global Landsat 5 (L5) data, added an interesting temporal dimension for coastal and inland water remote sensing, allowing for long-term (30+ years) studies. Coupled with easy data access through open portals such as USGS EarthExplorer, the Copernicus Science Hub, and several private initiatives such as Google Earth Engine (GEE), now for the first time the entire image archive is available for free to individual researchers. This includes L5 imagery dating back to 1984, and data from earlier Landsats (1 through 4) dating back to the early 1970s.

Atmospheric correction methods commonly used for water applications separate the water and atmospheric signals using different approaches. The aerosol contribution can be estimated after Rayleigh correction through the use of a "black pixel" assumption in the NIR (Antoine and Morel, 1999; Gordon and Wang, 1994) or SWIR bands (Gao et al., 2007). Non-zero NIR water reflectance can be modeled using an iterative approach based on the derived chlorophyll a concentration (Bailey et al., 2010), but this method fails in extremely turbid waters (Dogliotti et al., 2016). The use of a fixed water reflectance model and spatial homogeneity of aerosol type has been suggested for moderately turbid waters (Hu et al., 2000; Ruddick et al., 2000). The aerosol optical thickness can be imposed from external measurements or model results (Harmel et al., 2018), or can be derived by spectral unmixing of end-member spectra (De Keukelaere et al., 2018). Other methods model both the water and atmospheric components simultaneously through iterative fitting (Steinmetz et al., 2011)

E-mail address: quinten.vanhellemont@naturalsciences.be.

https://doi.org/10.1016/j.rse.2019.03.010
Received 4 October 2018; Received in revised form 21 November 2018; Accepted 6 March 2019
Available online 13 March 2019

or by the use of neural networks (Schiller and Doerffer, 1999; Schroeder et al., 2007).

There is still a lack of a single robust atmospheric correction algorithm for water applications and processing software which can process these bulk data from Landsat and Sentinel-2 in an automated and consistent manner. Through ACOLITE (Vanhellemont and Ruddick, 2014, 2015a, 2016), we made a first effort to make a unified processor publicly available for water applications of this multi-sensor and long-term archive. The software and updates are provided on the ACOLITE forum: http://odnature.naturalsciences.be/remsem/software-and-data/acolite/. The simple AC was based on zero water reflectance in the shortwave-infrared (SWIR, wavelengths > 1 μm) part of the spectrum, and was originally developed for remote sensing of water turbidity in turbid and extremely turbid waters (Vanhellemont and Ruddick, 2015a) using Landsat and Sentinel-2 type sensors (L/S2). The signal in the SWIR was assumed to be entirely caused by aerosol and Rayleigh scattering, and after Rayleigh correction, the remaining signal was extrapolated from the SWIR to the visible (VIS) and near-infrared (NIR) bands using an exponential function (EXP). After ACOLITE was made publicly available it was quickly picked up by the community for processing of L/S2 imagery for aquatic applications. While the EXP AC was successfully used in other turbid environments (Braga et al., 2017; Brando et al., 2015; Caballero et al., 2018; Lee et al., 2016; Liu et al., 2017; Luo et al., 2018; Manzo et al., 2018; Novoa et al., 2017; Ody et al., 2016), the performance in other regions was not always satisfactory, often resulting in negative reflectances (e.g. Bernardo et al., 2017; Martins et al., 2017, and many unpublished results). Performance in the blue bands is generally quite poor due to the exponential extrapolation from the SWIR. Indeed for retrieval of chlorophyll-a in oceanic waters using blue and green bands, Gordon and Wang (1994) preferred the use of model-generated LUTs over the simpler exponential extrapolation. Similarly, performance for the older L5/7 imagery was found to be rather poor, likely due to the SWIR band quality and noise introduced by the per pixel atmospheric correction. The main reason the exponential extrapolation from the SWIR fails is the assumption that the SWIR reflectance is solely caused by aerosol and Rayleigh scattering. In the case of sun glint (very common for nadir-viewing sensors) and adjacency effects (common for inland and nearshore waters) the additional signal will be considered as aerosols, and will be included in the extrapolation to the VIS/NIR bands. In this case, not only is the aerosol reflectance overestimated, the assumption of an exponential shape will no longer be valid. Poor signal-to-noise (SNR) ratio in the SWIR bands (especially for L5/7) can cause problems if the AC relies on these bands for the aerosol estimation (see also Pahlevan et al., 2017a), especially when used in a per-pixel method. Even with a smoothing of the SWIR bands (as done by e.g. Franz et al., 2015; Pahlevan et al., 2017b; Vanhellemont and Ruddick, 2016) unwanted effects could be introduced near bright SWIR targets (clouds, land, ships). Fixing the atmospheric reflectance as spatially uniform over too large an area, for example using a single estimate on a 185 × 185 km Landsat WRS2 scene, is likely inaccurate, and having a spatially varying atmosphere is preferred.

Overall, making ACOLITE open access (and open source as of July 2018) has been a success, and the interaction with the user community leads to a better identification of user priorities, bugs and issues. There can be some negative side effects, such as bugs affecting the performance in other studies, e.g. Dörnhöfer et al. (2016) used an early release with S2 support where a bug in the geometry affected especially the blue band retrievals. Furthermore, multiple settings in a processor (e.g. choice of bands used in the AC) can lead to confusion among the users and could lead to sub-optimal settings for their particular case. Fixing the red/NIR water reflectance ratio to a known model (Neukermans et al., 2009; Ruddick et al., 2006; Vanhellemont and Ruddick, 2014) has limited general applicability, especially when the red/NIR ratio is affected by chlorophyll-a absorption or reflectance saturation (Luo et al., 2018; Vanhellemont and Ruddick, 2015a). Hence, a generic, robust, and automated band selection would be preferred to leaving the choice to the user.

In this paper, an AC algorithm for L/S2 sensors (specifically for Landsat 5/7/8 and Sentinel-2A/B) is presented which will address these common problems with a robust automated band selection process, and an aerosol correction which will allow for spatial variability of aerosols (both type and concentration) without impacting the noise level in the output product. To this aim, the dark spectrum fitting (DSF) atmospheric correction by Vanhellemont and Ruddick (2018) was adapted and evaluated for L/S2. The DSF was originally developed for water applications of “very high resolution” metre-scale optical satellites but showed some potential for the application to L/S2, due to their better spectral coverage (notably including bands in the SWIR region). Because of the coarser pixel size on L/S2, the ground-level object shadows typically selected by DSF for metre-scale sensors will no longer be available for the atmospheric correction. However, thanks to the presence of SWIR bands, where the water leaving radiance reflectance, ρ_w, is zero, the DSF should perform well in cases where the SWIR reflectance is not significantly affected by adjacency effects. Additionally, due to the dynamic band selection, the DSF will select other bands (typically blue or red) if the NIR/SWIR adjacency effects are too severe. Since the DSF is not a pixel based approach (although it could arguably work well that way for the L/S2 band sets when ignoring noise aspects), a tile-based version of the algorithm was implemented to process whole or even several merged L/S2 scenes with spatially variable aerosols. The scene size from these sensors is significantly larger than those from very high resolution imagers, about 185 × 185 km for a Landsat tile, and about 110 × 110 km for a Sentinel-2 granule and hence a single path reflectance estimate for the scene is likely insufficient. An optional image based sun glint correction of the surface reflectance is included for all L/S2 sensors.

The performance of both EXP and DSF algorithms is evaluated for the time period 2003–present using the global AERONET-OC network (Zibordi et al., 2009) for all considered sensors (L5/7/8 and S2A/B). Finally, the temporal aspect of this archive is demonstrated using a time-series covering a period from 2000-present of turbidity derived from the 5 considered satellite sensors, and turbidity measured at the Warp SmartBuoy (Mills et al., 2003) located in the Thames Estuary in the southern North Sea.

2. Data and methods

2.1. Satellite imagery

Imagery from the full archive of Landsat and Sentinel-2 series of satellites (L/S2) was processed with ACOLITE. Landsat data from the Thematic Mapper on Landsat 5 (L5/TM), ranging from 1984 to 2011, the Enhanced Thematic Mapper Plus on Landsat 7 (L7/ETM+), 1999–present, and the Operational Land Imager on Landsat 8 (L8/OLI), 2013–present were used. Landsat has a spatial resolution of 30 m, with 3 visible bands (blue, green, and red), 1 near-infrared (NIR) and 2 shortwave infrared (SWIR) bands on L5 and L7. The OLI on L8 adds another visible band (blue at 443 nm), and a cirrus detection band at 1.3 μm (Roy et al., 2014). L7 and L8 have a panchromatic channel at 15 m spatial resolution, which offers some sharpening potential for aquatic applications (Vanhellemont and Ruddick, 2015b). Recently, the extraction of additional spectral information from panchromatic channels has been suggested (Castagna et al., 2018). Data from the Multi-Spectral Imager on both Sentinel-2A (S2A/MSI), 2015–present, and Sentinel-2B (S2B/MSI), 2017–present were used. The MSI is a 13 band imager with spatial resolution of 10 (4 bands), 20 (6 bands) or 60 (3 bands) m (Drusch et al., 2012). It has 4 visible bands, 5 red-edge and NIR bands, 2 SWIR bands, a water vapour band at 945 nm, and a cirrus detection band at 1.3 μm. The band centre wavelengths of all sensors are given in Table 1 and the Relative Spectral Responses are plotted in Fig. 1. Sentinel-2 imagery was resampled to 10 m during processing by

Table 1
Band averaged wavelengths (nm) for the bands from each satellite sensor used in the present study. Thermal channels on Landsat sensors have been excluded, as well as the panchromatic and cirrus bands. For the MSI on S2A/B the cirrus and water vapour bands have been excluded.

Sensor	Blue1	Blue2	Green	Red	RE1	RE2	RE3	NIR	NIR2	SWIR1	SWIR2
L5/TM		486	571	660				839		1678	2217
L7/ETM+		479	561	661				835		1650	2208
L8/OLI	443	483	561	655				865		1609	2201
S2A/MSI	443	492	560	665	704	740	783	865	833	1614	2202
S2B/MSI	442	492	559	665	704	739	780	864	833	1610	2186

replicating pixels (no interpolation), although processing at 20 and 60 m is also supported in ACOLITE (using neighbourhood averaging for resampling higher resolution bands). Imagery from all sensors was downloaded from Google Cloud Services (GCS). All available data from Sentinel-2 was processed, using the most recent available processing baseline for each scene. For Landsat, " Collection 1" data from both Tiers 1 and 2 was used, excluding the Real Time data. Tier 1 includes the highest quality scenes that include precision terrain correction and have accurate georegistration (less than 12 m root mean square error). Tier 2 are the other scenes which do not meet Tier 1 criteria due to, among other reasons, cloud cover and lack of ground control points. Since coastal and offshore scenes often fall in the second Tier, both categories are used here. Cloudy scenes are removed by using a threshold on the SWIR reflectances. The " Collection 1" data from USGS hosted on GCS may not include all Landsat 5 data ever recorded, as the archive is still being consolidated from global receiving stations (Wulder et al., 2016), notably for European coverage. However, thanks to large consolidation efforts in 2017–2018, the global L5 data availability from the Collection 1 corresponds reasonably well with the era of operational AERONET-OC stations used as an in situ reference in the present paper. The following naming has been adopted: "scenes" for the satellite data as provided by the space agencies, "subscenes" for spatial

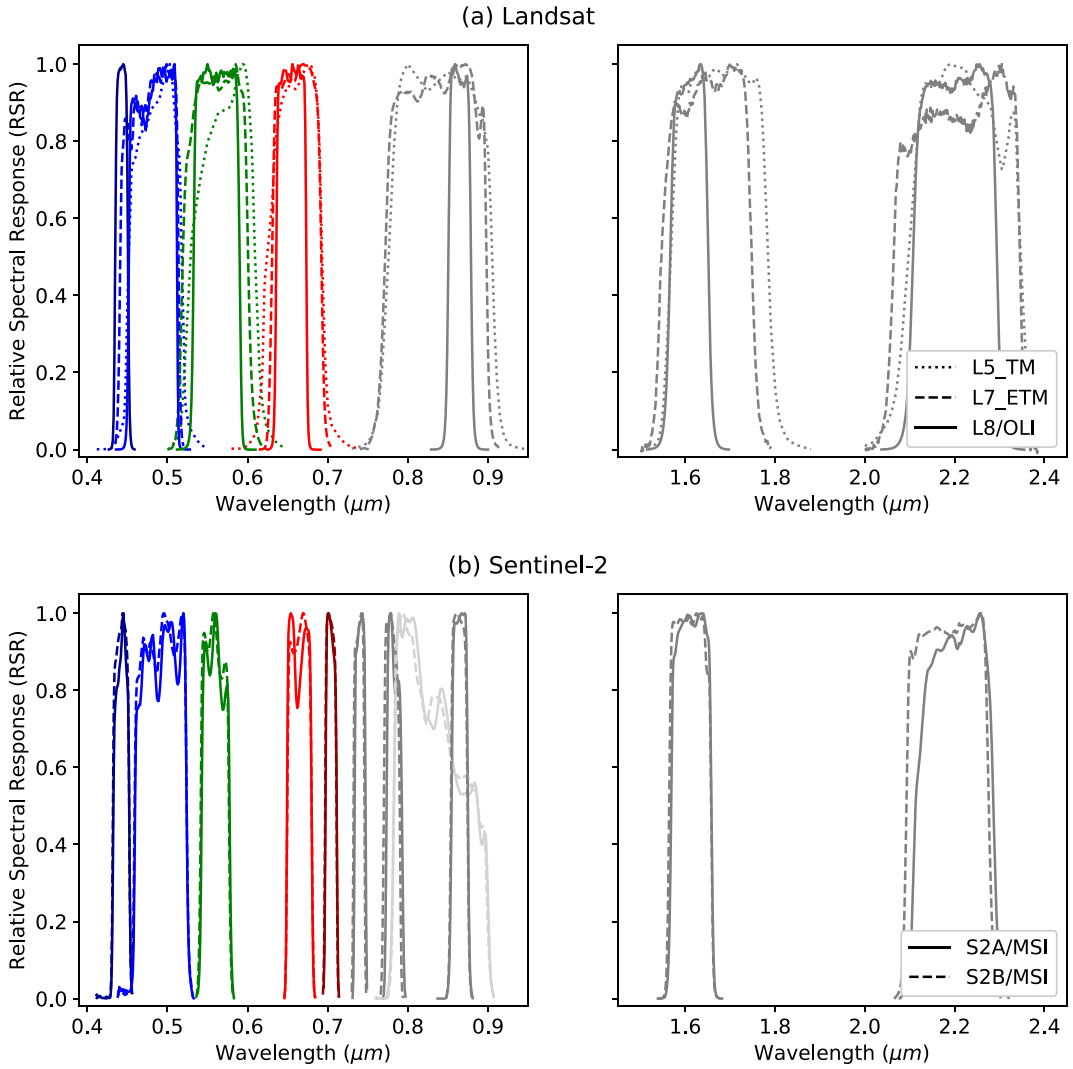

Fig. 1. Relative Spectral Responses for the bands in Table 1. The two blue bands are plotted in dark and light blue, the green and red bands in green and red. The MSI RE1 band is plotted in dark red, and the high resolution NIR band in light grey. The other RE, NIR and SWIR bands are plotted in darker grey. For S2A/MSI the January 2018 update of the Relative Spectral Response was used. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

subsets of these scenes, whether as user-defined region of interest (ROI) or as a tile in the tiled processing. A ROI covered by multiple scenes from the same overpass that are merged during processing is also considered a "subscene".

2.2. In situ data

In situ radiometry was obtained from all sites in the AERONET-OC network (Zibordi et al., 2009). These sites are equipped with a CIMEL radiometer which has been modified to also record water-leaving radiance (L_w) in addition to the optical parameters of aerosols. Data were downloaded in all three quality levels (lev10, lev15, and lev20) from the AERONET website (http://aeronet.gsfc.nasa.gov). These quality levels indicate (Zibordi et al., 2009): raw data with complete measurement sequences (lev10), cloud-screened data passing empirical quality control thresholds on the L_w (lev15) and fully quality assured data including post-deployment calibration and spectral consistency tests, with visual inspection and availability of aerosol optical thickness measurements (lev20). Matchups were identified for all three data levels, and were defined as satellite overpasses bounded by in situ measurements in a 4 hour window centred on the overpass time (i.e. overpass time ± 2 h). AERONET data from the closest bounding observations were linearly interpolated to the satellite overpass time. In situ normalised water-leaving radiance (L_{wn}) measurements are available with and without an f/Q correction for bidirectional effects of the L_w, where f is a factor relating apparent optical properties (i.e. the reflectance) to the inherent optical properties (i.e. absorption and scattering) of the water, and where Q describes the anisotropy of the in-water radiance. The standard AERONET-OC f/Q correction uses a chlorophyll-a based approach (Zibordi et al., 2009). The f/Q correction has the largest impact at high reflectances, which likely originate from non-phytoplankton dominated systems, and would likely overcorrect most spectra from turbid sediment-dominated waters of interest in the present study. Hence, non-f/Q corrected L_{wn} were used as (1) there is no normalisation applied to the satellite observations, and (2) the f/Q correction uses a chlorophyll-a concentration estimate, which may not be valid in the more turbid water stations (Zibordi et al., 2009). The in situ L_{wn} are converted to water-leaving radiance reflectances, ρ_w:

$$\rho_w = \pi \cdot L_{wn} / F0 \tag{1}$$

where F0 is the extraterrestrial solar irradiance (Thuillier et al., 2003) for a narrow 10 nm square band centred on the CIMEL wavelengths. ρ_w values are linearly interpolated to form a "hyperspectral" dataset which is then resampled to the relative spectral response (RSR) of the satellite sensor. The resampled interpolated dataset is then finally band shifted using coefficients retrieved from the simulated dataset by Nechad et al. (2015). This band shifting approach is presented in Supplementary Data 1.

In situ turbidity (FNU) was measured at the Warp SmartBuoy (Mills et al., 2003) located approximately at 1.05133° E, 51.5347° N using a SeaPoint turbidimeter mounted at 1 m depth and recording data every 15 or 30 m(depending on deployment) as the mean average of a 10 minute burst measurement. The instrument measures turbidity as side-scattering at 880 nm, with a 15–150 ° acceptance angle and peak sensitivity at 90°. Data were obtained from the CEFAS website (https://www.cefas.co.uk/cefas-data-hub/smartbuoys/) as quality controlled post-recovery data spanning the deployments between 2000 and 2017. Matchups were defined as measurements bounding the satellite overpass, with time differences between the samples and the overpass time less than the deployment sampling window of 15 or 30 min. The bounding measurements were then linearly interpolated to the overpass time.

2.3. Atmospheric correction methods

The atmospheric correction (AC) aims to separate the top-of-atmosphere (TOA) observation by the satellite sensor into the signal from the atmosphere and the signal from the surface in order to retrieve surface reflectances, ρ_s, or water-leaving radiance reflectances, ρ_w. The distinction between ρ_s and ρ_w is here made as some methods only retrieve the reflectance over water pixels, and others over both water and land. For the present study the retrieved ρ_s and ρ_w should be equivalent. ρ_s for water pixels is assumed to have the water surface fully treated, although sun glint may still be present and require separate correction (Section 2.3.4). Two atmospheric correction methods available in ACOLITE will be compared, the SWIR based exponential extrapolation method by Vanhellemont and Ruddick (2014, 2015a), hereafter EXP, and a multi-band "dark spectrum fitting" technique by Vanhellemont and Ruddick (2018), hereafter DSF. The DSF was developed for metre scale resolution sensors, and is in the present paper adapted and evaluated for the decametre resolution sensors on Landsat and Sentinel-2. The satellite image data were provided by the space agencies as TOA reflectance, ρ_t. ρ_t is corrected for gas transmittance (t_{gas}), here taken as the product of the band averaged ozone and water vapour transmittances, and water pixels are corrected for sky reflectance reflected at the air-water interface (see Vanhellemont and Ruddick, 2018, for details). In the following two sections, ρ_t is assumed to be have the gas and sky reflectance corrections applied. The water vapour (945 nm) and cirrus bands (1.3 µm) are excluded from both methods due to the low atmospheric transmittance in these bands.

2.3.1. The SWIR based exponential extrapolation (EXP)

The SWIR based exponential extrapolation (EXP) method assumes the water reflectance is zero in both SWIR bands. The multiple scattering aerosol reflectance, ρ_{am}, can hence be estimated over water pixels. For SWIR1 and SWIR2,

$$\rho_w = 0 \quad \text{and} \quad \rho_{am} = \rho_{rc}, \tag{2}$$

where ρ_{rc} is the Rayleigh corrected reflectance:

$$\rho_{rc}(\lambda) = \rho_t(\lambda) - \rho_r(\lambda), \tag{3}$$

with ρ_r the Rayleigh reflectance estimated from the LUT presented in Section 2.4. The ratio of the ρ_{am} in the SWIR, ϵ, is then used to estimate the spectral dependency of the aerosol reflectance:

$$\epsilon = \frac{\rho_{am}(SWIR1)}{\rho_{am}(SWIR2)} \tag{4}$$

The ρ_{am} can then be calculated at all other wavelengths by use of an exponential extrapolation to calculate the ratio to ρ_{am} (SWIR2), ϵ (λi,*SWIR2*):

$$\epsilon(\lambda i, SWIR2) = \epsilon^{\delta i}, \tag{5}$$

where δi is given by:

$$\delta i = \frac{\lambda(SWIR2) - \lambda i}{\lambda(SWIR2) - \lambda(SWIR1)}. \tag{6}$$

The water-leaving radiance reflectance, ρ_w, can then be computed in all bands using:

$$\rho_w(\lambda i) = \frac{1}{t_{Ray}(\lambda i)}[\rho_{rc}(\lambda i) - \epsilon(\lambda i, SWIR2) \cdot \rho_{am}(SWIR2)], \tag{7}$$

where t_{Ray} is the two-way Rayleigh transmittance. Due to the reliance on water pixels, the method is not so well suited to derive land reflectances, especially when the distance to the nearest water pixels is larger than the spatial variability of the aerosols. The EXP supports pixel-by-pixel processing for water targets, but this may result in noisy outputs, due to the low signal and low signal-to-noise ratio. Fixing of the ρ_{am} and ϵ over subscenes is recommended.

2.3.2. The dark spectrum fitting (DSF) algorithm and updates for L/S2

The dark spectrum fitting (DSF) algorithm uses two assumptions to estimate the atmospheric path reflectance, ρ_{path}, from a satellite scene

or subscene:

1. The atmosphere is homogeneous over a certain extent, i.e. ρ_{path} is constant within the considered (sub)scene.
2. The (sub)scene contains pixels with $\rho_s \approx 0$ in *at least one* of the sensor bands, where ρ_{path} can be estimated.

A representative dark spectrum, ρ_{dark} is constructed from the lowest observed ρ_t in each band. The ρ_{dark} is used to select the most appropriate band/model combination to estimate ρ_{path} in a three step process:

1. For the scene specific sun and viewing geometry, the spectral ρ_{path} is computed for a range of aerosol models (Section 2.4) and the aerosol optical thickness at 550 nm, τ_a. For each band, denoted by the band weighted wavelength λ, and each of the aerosol types in the LUT, the observed $\rho_{dark}(\lambda)$ will be bounded by two $\rho_{path}(\lambda)$ values, corresponding to τ_a steps in the LUT. These bounding τ_a values are then linearly interpolated to the $\rho_{dark}(\lambda)$ to give an estimated τ_a for this band and aerosol model.
2. For a given aerosol model, the τ_a estimate from the band giving the lowest non-zero τ_a will be used, as the higher values obtained using the other bands will give negative ρ_s for the dark pixels in the band giving the lowest τ_a.
3. An aerosol model is then selected using the Root Mean Squared Difference (RMSD) between the ρ_{dark} and the estimated ρ_{path}. The RMSD is computed for each band pair containing the fitted band. The aerosol model and band combination giving the lowest RMSD in any of those pairwise comparisons is considered to be best fitting and is finally selected for the atmospheric correction.

With the best fitting band and aerosol model combination selected, the parameters required for the atmospheric correction are then retrieved from the LUT for all bands using the computed τ_a at 550 nm:

- ρ_{path}: the atmospheric path reflectance,
- t_{tot}: the two-way total atmospheric transmittance,
- s_a: the spherical albedo of the atmosphere,

which allows for the computation of directional surface reflectance, ρ_s:

$$\rho_s = \frac{\rho_{pc}}{t_{tot} + s_a \cdot \rho_{pc}}, \tag{8}$$

with ρ_{pc} the "path-corrected" reflectance:

$$\rho_{pc} = \frac{\rho_t}{t_{gas}} - \rho_{path} - \rho_{sky}, \tag{9}$$

where t_{gas} is the gas transmittance, and ρ_{sky} is an estimate of the air-water interface sky reflectance, which is set to 0 for land pixels and estimated analytically for water pixels (as in Vanhellemont and Ruddick, 2018).

2.3.3. Tiled DSF processing

One of the assumptions of the DSF is the spatial constraint on the aerosol variability. Although a single ρ_{path} can be used to process a full scene or merged product, an automated processing using an internal tiling was implemented in order to incorporate spatial variability of the atmosphere over large spatial extents. Tile size of approximately 6 × 6 km are used, i.e. the tile size was set to 200 by 200 pixels for Landsat, and to 600 by 600 (10 m) pixels for Sentinel-2. For each tile separately, the DSF is used to estimate the path reflectance according to Section 2.3.2. After the tiled processing, an additional filtering step is performed to remove tiles with very low τ_a estimates (τ_a 550 nm < 0.005), indicative of cloud shadow presence, where not only the surface but also the atmosphere is shaded. Tiles with less than 10% pixel coverage (i.e. those at the swath edges) are also removed. For the final full-scene processing, the retrieved parameters are linearly interpolated between the different tile centres of adjacent tiles to form smooth full-scene datasets. This approach allows for spatial variability of the aerosol estimates, with a smooth transition over land and water targets. The interpolation also makes the continuous mixing of different aerosol models selected in adjacent tiles possible.

2.3.4. Glint correction for the DSF

Sun glint on the water surface is a recurring problem for nadir viewing sensors. Much of the Landsat and Sentinel-2 archive is affected by sun glint, especially for lower latitude areas. As the DSF selects the band giving the lowest estimate of atmospheric path reflectance, pixels and bands with severe sun glint are avoided in the ρ_{path} estimation in the atmospheric correction. This means that the glint signal will be still present in the resulting surface reflectance, and by assuming zero water-leaving reflectance in a reference band, λ_{ref}, the glint reflectance, ρ_g, can be estimated:

$$\rho_g(\lambda_{ref}) = \rho_s(\lambda_{ref}) \tag{10}$$

ρ_g can then be computed in other bands by taking into account the ratios of direct atmospheric transmittance and Fresnel reflectance at the water surface:

$$\rho_g(\lambda) = \rho_s(\lambda_{ref}) \cdot \frac{T_{dir}(\lambda)}{T_{dir}(\lambda_{ref})} \cdot \frac{r(\theta, \lambda)}{r(\theta, \lambda_{ref})}, \tag{11}$$

where T_{dir} is the two way direct transmittance of the atmosphere, computed from the total optical thickness of the atmosphere, τ_t, retrieved from the LUT (see Section 2.4):

$$T_{dir} = e^{-\tau_t(\lambda)/\cos\theta_s} \cdot e^{-\tau_t(\lambda)/\cos\theta_v}, \tag{12}$$

and $r(\theta,\lambda)$ is the Fresnel reflectance coefficient for air-incident rays at wavelength λ:

$$r(\theta) = 0.5 \cdot \left\{ \frac{\sin^2(\theta - \theta_t)}{\sin^2(\theta + \theta_t)} + \frac{\tan^2(\theta - \theta_t)}{\tan^2(\theta + \theta_t)} \right\}, \tag{13}$$

for the incidence angle θ for the sun (θ_s) and viewing (θ_v) zenith angles and relative azimuth angle $\Delta\phi$:

$$\cos 2\theta = \cos\theta_s \cos\theta_v + \sin\theta_s \sin\theta_v \cos\Delta\phi \tag{14}$$

with the angle of transmittance, θ_t, given by:

$$\theta_t = \sin^{-1}\left(\frac{1}{n_w(\lambda)\sin\theta}\right), \tag{15}$$

where $n_w(\lambda)$ is the band averaged refractive index of water with respect to air, taken from the Water Optical Properties Processor (WOPP) by Röttgers et al. (2016). Harmel et al. (2018) have shown the importance of using a spectral n_w, and demonstrate a robust sun glint correction using the SWIR bands on Sentinel-2A/MSI. In the present paper, ρ_g is estimated for all sensors in both SWIR bands (ρ_s at 1.6 and 2.2 µm). The SWIR band giving the lowest ρ_g is then used in the correction, to avoid negative reflectances in the other SWIR band. For clear water sites, the NIR bands could be used as reference bands in the glint correction, or an alternative method such as presented by Hedley et al. (2005) could be used. These are however not generally applicable over turbid waters due to non-zero NIR reflectance.

2.4. Updates to the LUT

The look-up table (LUT) used in the DSF was constructed using the 6SV radiative transfer model (Kotchenova et al., 2006; Vermote et al., 2006) containing atmospheric path reflectance, ρ_{path}, two-way diffuse atmospheric transmittance, t_{du}, the spherical albedo of the atmosphere, s_a, and the total optical thickness of the atmosphere, τ_t, for the default Continental (model 1) and Maritime (model 2) aerosol models. The LUT

supports the addition of more aerosol models if required. The LUT generated by Vanhellemont and Ruddick (2018) was updated to include extra wavelengths in the SWIR, for better representation of the relatively wide Landsat and Sentinel-2 bands at 1.6 µm and 2.2 µm(see Fig. 1). The LUT now contains 18 wavelengths: 0.39, 0.41, 0.44, 0.47, 0.51, 0.55, 0.61, 0.67, 0.75, 0.865, 1.04, 1.24, 1.55, 1.61, 1.66, 2.10, 2.25, and 2.40 µm. For each sensor, the LUT is interpolated from the 18 wavelengths to a hyperspectral (1 nm step) dataset between 0.39 and 2.40 µm. This hyperspectral dataset is then resampled to the relative spectral response (RSR) of the sensor bands. The LUT is generated without the effects of trace atmospheric gases and hence the resulting datasets are spectrally smooth, and can be accurately interpolated.

2.5. Processing of satellite data for validation

For the comparison with AERONET-OC data, imagery with bounding in situ measurements within 2 h of the overpass time was downloaded from GCS, with potential matchups identified for 19 sites across the 5 sensors. Imagery was cropped to an approximately 10 by 10 km region of interest (ROI) centred on the site. For some sites, overlapping scenes from the same orbit were merged to a single dataset. The merged and cropped images were processed using the DSF with a single ρ_{path} estimate over the ROI, and using the EXP with the aerosol concentration and type fixed (using the median values) over the ROI. The DSF with glint correction (DSF + GC) was used as a third processing configuration. The ρ_{dark} for the DSF is estimated from the intercept of an Ordinary Least Squares regression of the 1000 darkest pixels in each band, as done by Vanhellemont and Ruddick (2018). For turbidity time-series processing, scenes covering the Warp SmartBuoy location (Landsat tiles 200/024 and 201/024 and Sentinel-2 granule T31UCT) were downloaded from GCS, and processed for a 10 × 10 km area centred on the station with a fixed subscene ρ_{path} estimated using the DSF + GC. ρ_t and ρ_s or ρ_w were extracted for a 3 × 3 pixel box centred on the site location.

Extracted satellite data were filtered according to the 3 × 3 box mean average values. Data were removed where the ρ_s or $\rho_w < 0$ in any band, and where the ρ_t 1.6 µm > 0.01. The negatives are removed to exclude data with an obvious problem in the atmospheric correction (overestimation of atmospheric reflectance) or the satellite imagery itself. Negative reflectances are relatively uncommon in the DSF results due to the construction of the algorithm, and are often the result of cloud shadow contamination. The water leaving radiance in 1.6 µm band is expected to be zero (Wang, 2007), and hence the signal in the SWIR bands can be attributed to (1) atmospheric scattering, (2) glint on the water surface, and (3) haze, clouds or the presence of above water objects. By excluding the pixels with high ρ_t 1.6 µm, quite an accurate non-water filtering can be achieved (Vanhellemont and Ruddick, 2015a; Wang and Shi, 2006). The used threshold results in a quite restrictive data filtering, and will also exclude scenes with a clear atmosphere and significant sun glint contribution. Depending on a specific study and study area, and in the interest of data availability, the quality control may be relaxed.

Due to the potential optical contamination caused by the AERONET-OC site structure, inaccuracies in the provided coordinates for some sites, and pixel shifts caused by viewing parallax or imperfect geolocation of the satellite imagery for offshore locations, the coordinates for matchup extraction were manually shifted based on cloud-free L/S2 imagery. The shifted coordinates are provided in Supplementary Data 2, as well as some RGB examples showing some of the larger sites and the used reference location. These shifts are done on a spatial scale (< 300 m) over which the water masses around the stations are typically homogeneous. Pixels directly around the station could be used, but automated detection of structure and shadow pixels mixed with varying coverage of water proved to be unreliable. Imperfect image geolocation and viewing parallax may skew the location of the station structures from scene to scene, and using a fixed set of pixels directly around the site could be unreliable. By applying this shift, the influence of currents around the structure are also avoided.

For the Warp SmartBuoy site, the mean average turbidity in Formazine Nephelometric Units (FNU) was computed using the algorithm of Nechad et al. (2009):

$$T = \frac{A \cdot \rho_w}{1 - \frac{\rho_w}{C}}, \tag{16}$$

where ρ_w is the water-leaving radiance reflectance, here taken as the ρ_s retrieved after the DSF atmospheric correction. A and C are band specific calibration coefficients, using band-weighted coefficients from the hyperspectral calibration provided by Nechad et al. (2009), with the coefficients provided in Supplementary Data 3. Both red and NIR bands were used to compute turbidity, for MSI the 10 m NIR band at 833 nm was used.

3. Results

3.1. Application of the DSF

The DSF aerosol selection process is illustrated for a S2A scene over Zeebrugge in Fig. 2, showing the ρ_{path} according to several τ_a steps in the LUT (τ_a at 550 nm between 0.05 and 1.3) for the Continental and Maritime model, coloured from yellow to red with increasing τ_a. Both plots show the ρ_{dark} (black dots) and the Rayleigh reflectance, $\rho_{Rayleigh}$ (dotted blue). The τ_a estimated from the linear interpolation of the LUT steps to the ρ_{dark} is plotted for selected bands. For the Maritime model (right plot), band 11 (1.6 µm) would be selected as the fitted band in step (2), as it gives the lowest τ_a (0.10). For the Continental model (left plot), band 1 (443 nm) gives the lowest τ_a (0.24). For both models, a ρ_{path} is calculated from the retrieved τ_a. The ρ_{path} is then compared with the ρ_{dark} for the two values in all band pairs containing the selected band (as in step 3 in Section 2.3.2). The model and τ_a which gives the lowest RMSD for any of these band pairs will be selected for the atmospheric correction. In the example here, the best fitting bands are 11 and 12 (1.6 and 2.2 µm) for the Maritime model, and 1 and 8 (443 and 833 nm) for the Continental model. The RMSD are respectively $6.3 \cdot 10^{-4}$ and $1.6 \cdot 10^{-4}$. Although the τ_a at 550 nm is lower for the Marine model, the ρ_{path} from the Continental model is found to be better fitting, and hence the Continental model with a τ_a of 0.24 will be used in the processing.

3.2. Tiled processing

The automated tiled processing of larger scenes consists of running the DSF on subscenes, retrieving a ρ_{path} for each subscene, and interpolating the results to form a full scene dataset. An example of the tiled processing is given in Fig. 3 for a Landsat scene over the turbid coastal waters in the southern North Sea. The bands used in this example to estimate the ρ_{path} vary from red/NIR/SWIR bands (4, 5, 6, 7) over the turbid coastal waters to the visible bands (typically band 1) over land (middle panel). This result shows a fairly typical switch between the use of longer wavelength bands over water, and the shorter wavelength bands over land. The retrieved ρ_{path} (right panel) is highest over the offshore waters around 52.2° N, 2.9 ° E and the English, and Belgian/French land masses. This indicates the method retrieves the spatial variability of aerosols within the scene, which is not related to the average brightness of the underlying surface. Typically Landsat scenes are split in about 40 × 40 tiles, and Sentinel-2 scenes in 19 × 19 tiles; resulting in running the DSF respectively 1600 and 361 times. A full scene ρ_{path} is generated by interpolation of the individual tile retrievals. This allows for smoothly tracing the variability of the aerosol reflectance throughout the scene.

A transect from the Thames river mouth to the Belgian coastal zone is shown in Fig. 4, and shows the estimated variability of ρ_{path}, and the

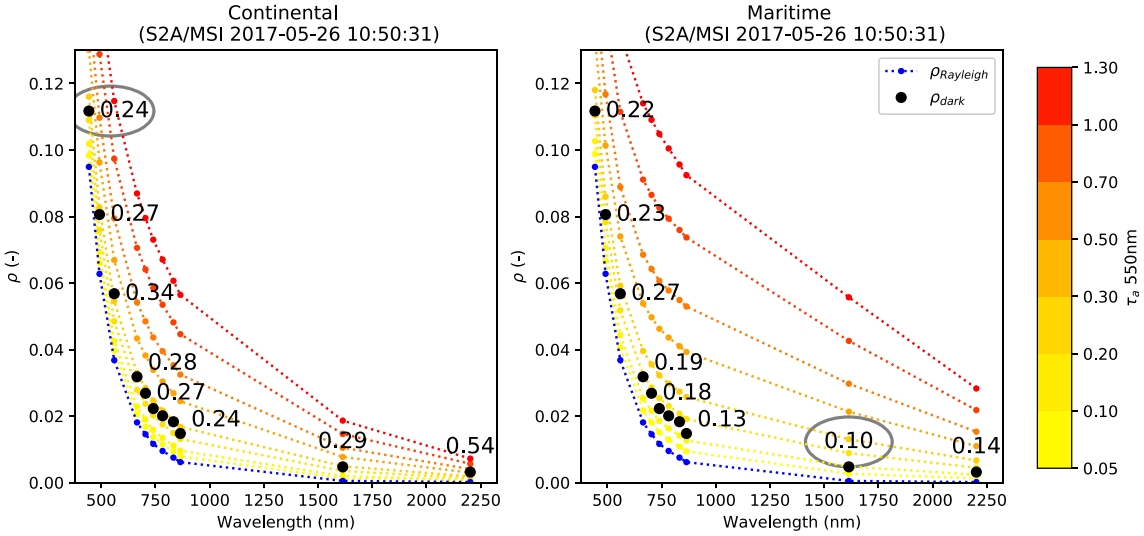

Fig. 2. The dark spectrum fitting (DSF) algorithm applied to a Sentinel-2A scene over Zeebrugge harbour (2017-05-26), using the Continental (left) and Maritime aerosol models (right). The dashed blue line is the Rayleigh reflectance ($\rho_{Rayleigh}$), the yellow to red lines are the path reflectance (ρ_{path}) for increasing τ_a. The grey circles show the selected band (with minimum τ_a) for each model. τ_a values are plotted for selected bands, some NIR values were suppressed for legibility. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

retrieved ρ_s. The ρ_{path} is not affected by the high ρ_s in the English (1–2 ° longitude) and Belgian turbid coastal waters (> 3° longitude). The variability across the ρ_s transect are changes in resuspended sediment concentration largely related to changes in bathymetry (i.e. higher turbidity is found over the shallow sand banks). By using the scene-wide fixed τ_a, in certain regions the ρ_{path} is underestimated and hence the ρ_s overestimated. This is especially important in the clearer waters in the example presented here, where the overestimation of ρ_s can be > 50–100 % of the signal at 865 nm, e.g. around 2.5° longitude in the transect in Fig. 4.

3.3. Glint correction

The optional glint correction applied after the DSF is illustrated using a L8/OLI image in Figs. 5 and 6. This example shows that by using the SWIR bands to estimate glint reflectance at the surface, ρ_g, the correction is not affected by below-surface turbidity features. More realistic water spectra are retrieved after the glint correction, even for very high ρ_g (0.06–0.10 at 1.6 µm). The red band ρ_g for these points ranges between 0.04 and 0.12, which will render the red band derived turbidity product unusable if not corrected for. Computed turbidity ranges between 26 and 180 FNU without and between 4 and 27 FNU with glint correction. For P1 and P2 located in Galveston Bay, retrieved turbidities were respectively 62 and 74 FNU before, and 26 and 27 FNU after glint correction. The average May turbidity for the autonomous turbidity station closest to Galveston Bay was 31 FNU (USGS Site 294643095035200, Lynchburg Road, Baytown, TX, measurements between 2016 and 2018). The glint correction is further evaluated in the radiometry validation section below. It should be noted that the spectra presented in Fig. 6 would all be removed using the quality threshold established in Section 2.5.

3.4. Radiometry matchups

The retrieved water reflectances were evaluated using in situ radiance measurements from global AERONET-OC sites. In total, 1869 merged scenes (i.e. total number of overpasses, the number of individual scene files was higher) with bounding in situ AERONET-OC data at lev10 were processed: 207, 681 and 510 for L5/7/8 and 415 and 56 for S2A/B. This dataset was reduced to 785 and 498 scenes with

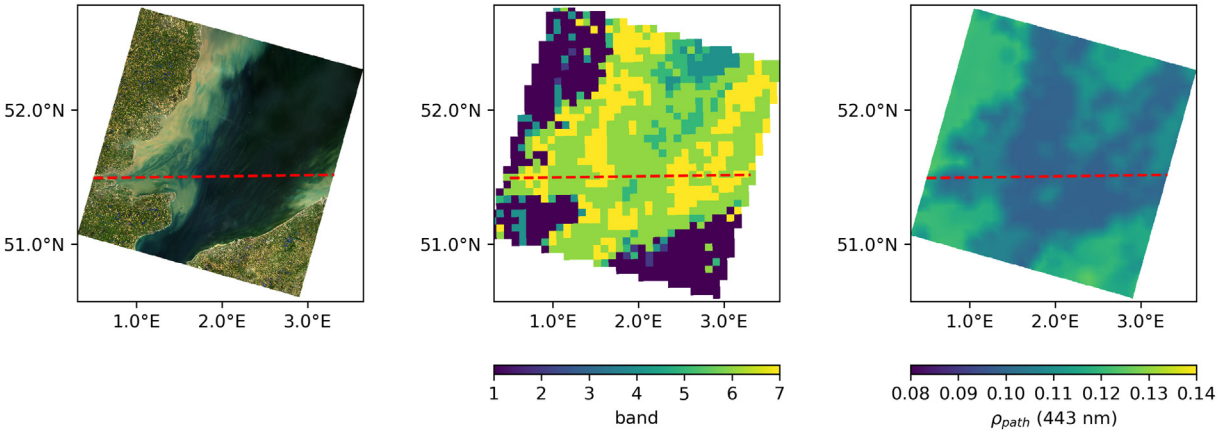

Fig. 3. A Landsat 8 OLI scene (2018-04-19 10:45 UTC, path/row 200/024) over the southern North Sea and Thames Estuary, (left) ρ_s RGB composite (655, 561, 483 nm) showing suspended sediments (greenish hues) in the coastal waters, (middle) the 6 × 6 km tiling used in the processing, showing for each tile the band giving the lowest τ_a at 550 nm, and (right) the ρ_{path} in the blue band (443 nm) interpolated from the tiles. The dashed horizontal line is the transect shown in Fig. 4. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

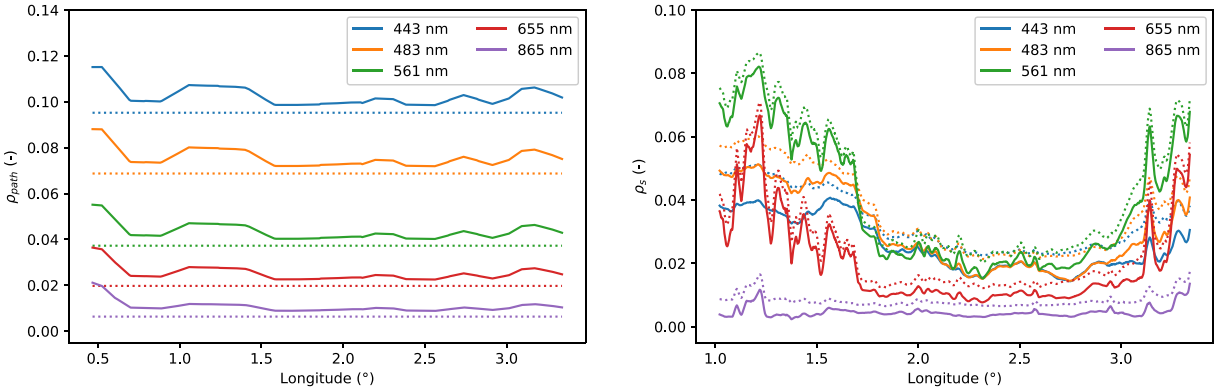

Fig. 4. A transect from the Landsat scene over the southern North Sea and Thames Estuary shown in Fig. 3, showing the estimated ρ_{path} (left panel) and the retrieved ρ_s (right panel) for the VNIR bands. The dotted lines show the retrieval for a scene-wide fixed τ_a, and the solid lines show the retrieval from the tiled processing. The ρ_s data were smoothed using a 50 pixel (1.5 km) moving average, in order to reduce the vertical variability in the 8000+ pixel plot.

bounding in situ lev15 and lev20 data. After quality filtering of the satellite data using the threshold filtering on retrieved ρ_t at 1.6 µm, i.e. removing cloudy and severely glinted scenes, 328 matchups with lev20 AERONET-OC data were retained; respectively 37, 138, and 98 for L5/7/8 and 55 for S2A. No lev20 matchups were available yet with S2B/MSI, due to the recency of the satellite. Matchups with lev20 AERONET-OC data are given in Table 2 and Fig. 7 for L5/TM, in Table 3 and Fig. 8 for L7/ETM+, in Table 4 and Fig. 9 for L8/OLI, and in Table 5 and Fig. 10 for S2A/MSI. The tables include the Reduced Major Axis (RMA) regression slopes (m) and offsets (b), the root mean squared difference (RMSD), the mean absolute difference (MAD), and the mean absolute relative difference (MARD). Results are given for the exponential extrapolation (EXP), the dark spectrum fitting without (DSF) and with glint correction (DSF + GC) methods. For L7, the number of matchups was lower for EXP than for DSF due to the larger number of negative reflectances retrieved by EXP, likely as a result of the striped imagery caused by the scan line corrector failure in 2003. Matchups with the not fully quality controlled lev15 data are provided for completeness in Supplementary Data 4. These contain the only matchups that are presently available for S2B/MSI, and in general have a larger in situ data range for the other sensors, while retaining a similar performance in terms of statistics. A visual representation of the RMSD and MARD per sensor band and atmospheric correction method is also provided in Supplementary Data 4.

Reasonable results are retrieved overall, with an RMSD between satellite and in situ generally < 0.01 for the DSF, with the EXP having similar errors for the red and green bands. For the blue bands, the errors for the EXP become much larger, sometimes exceeding 0.02. Typically the minimum RMSD is achieved for the red bands, with a minimum of 0.005 found for the DSF and DSF + GC. The RMSD ranges in the visible are 0.62–2.57 for EXP, 0.58–1.28 for DSF, and 0.47 –1.21 for DSF + GC ($\times 10^{-2}$). The RMSD for the NIR band are of similar magnitude, generally 1.1–1.6 for EXP, 0.8–1.1 for DSF and 0.6 –1.1 for DSF + GC ($\times 10^{-2}$). This is mainly due to the low signal encountered in that band, and increases for sensors with lower signal to noise ratio in that band. The MARD shows a similar pattern across sensors and methods, with about 30–100 % difference in the visible bands. The NIR band MARD is generally > 100%, and occasionally > 170%, again as a result of generally low NIR water signals. Typically the relative differences increase with decreasing signal size, and lowest MARD are found for the green bands that have the highest reflectances in the current dataset. The performance of EXP is the worst, with MARD ranging between 37 and 101 % for the visible and generally > 150% for the NIR bands. For DSF and DSF + GC the MARD reduce to respectively 28–69% and 24–73 % for the visible bands. DSF performance is more or less comparable to that of EXP for the NIR bands; with perhaps some improved performance for L5/TM and L7/ETM+. The RMSD and MARD are in most cases higher for EXP than for DSF, and the lowest values are found for the DSF + GC across all sensors and bands, with the exception of shortest blue band on S2A/MSI where a slightly lower MARD is found for DSF. The same trends are also noticeable in the RMA slope difference between DSF and EXP, showing a more reliable estimate of ρ_s by the DSF (slopes closer to 1). RMA offsets decrease from EXP to DSF and are smallest for DSF + GC. The most consistent performance between the DSF and EXP is found for L8 and S2A, which have more narrow bands and better radiometric performance compared to L5 and L7. For L5 and L7, the improvement from EXP to DSF is drastic, with visible band slopes improving from 1.2–1.7 to 0.9 –1.2, and giving about half

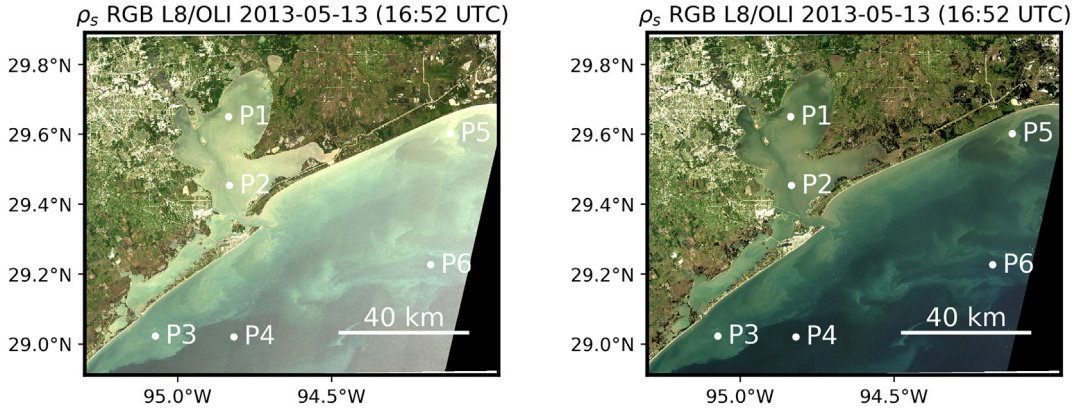

Fig. 5. ρ_s RGB composites from a L8/OLI image covering Houston Harbour (Texas, USA), with significant sun glint contamination. (Left) standard DSF, (right) DSF + glint correction. Reflectance spectra for the plotted points with and without glint correction are given in Fig. 6.

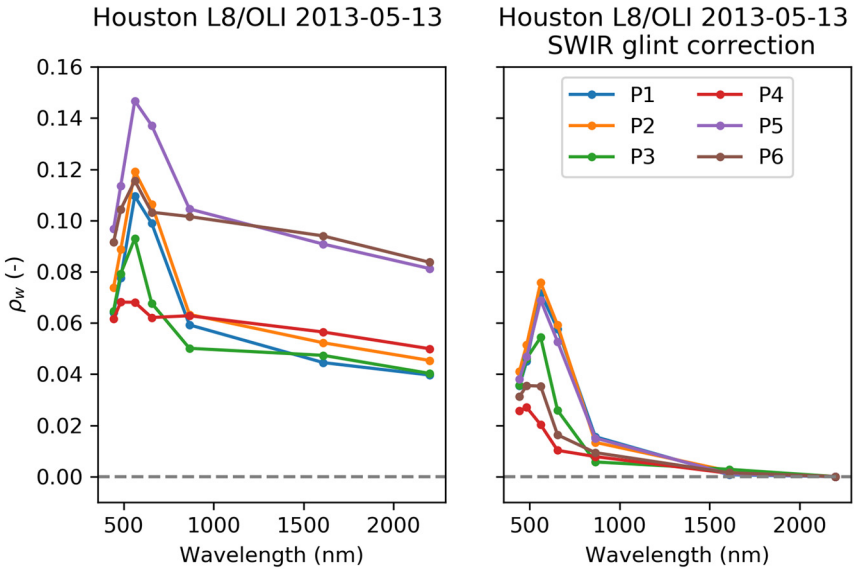

Fig. 6. Spectra extracted from Fig. 5, DSF derived ρ_s without (left) and with (right) glint correction applied.

Table 2
Landsat 5/TM matchups with Level 2 AERONET-OC data. n: number of matchups, m, b: RMA regression slope and offset, RMSD: Root Mean Squared Difference, MAD: Mean Average Difference, MARD: Mean Average Relative Difference. Data were rejected when the extracted box average ρ_t 1.6 µm > 0.01.

Band	ac	Level	n	m	b	R^2	RMSD	MAD	MARD
486	EXP	L2	37	1.59	$1.38 \cdot 10^{-2}$	0.53	$2.59 \cdot 10^{-2}$	$2.36 \cdot 10^{-2}$	0.89
	DSF	L2	37	1.20	$3.87 \cdot 10^{-4}$	0.61	$8.17 \cdot 10^{-3}$	$3.71 \cdot 10^{-3}$	0.36
	DSF + GC	L2	37	1.19	$-5.43 \cdot 10^{-4}$	0.62	$7.49 \cdot 10^{-3}$	$2.56 \cdot 10^{-3}$	0.32
571	EXP	L2	37	1.22	$1.39 \cdot 10^{-2}$	0.61	$2.05 \cdot 10^{-2}$	$1.85 \cdot 10^{-2}$	0.67
	DSF	L2	37	0.98	$6.14 \cdot 10^{-3}$	0.74	$8.35 \cdot 10^{-3}$	$5.76 \cdot 10^{-3}$	0.34
	DSF + GC	L2	37	0.98	$4.74 \cdot 10^{-3}$	0.76	$7.25 \cdot 10^{-3}$	$4.28 \cdot 10^{-3}$	0.29
660	EXP	L2	37	1.60	$8.55 \cdot 10^{-3}$	0.24	$1.52 \cdot 10^{-2}$	$1.33 \cdot 10^{-2}$	0.94
	DSF	L2	37	1.08	$2.02 \cdot 10^{-3}$	0.38	$5.45 \cdot 10^{-3}$	$2.63 \cdot 10^{-3}$	0.45
	DSF + GC	L2	37	1.04	$6.55 \cdot 10^{-4}$	0.43	$4.50 \cdot 10^{-3}$	$9.39 \cdot 10^{-4}$	0.39
839	EXP	L2	37	−5.34	$2.23 \cdot 10^{-2}$	0.00	$1.58 \cdot 10^{-2}$	$1.46 \cdot 10^{-2}$	1.70
	DSF	L2	37	3.74	$3.83 \cdot 10^{-3}$	0.00	$8.25 \cdot 10^{-3}$	$7.15 \cdot 10^{-3}$	1.49
	DSF + GC	L2	37	3.22	$2.52 \cdot 10^{-3}$	0.01	$6.31 \cdot 10^{-3}$	$5.22 \cdot 10^{-3}$	1.36

the relative errors. The DSF and DSF + GC slopes are very consistent, indicating that the glint correction was not sensitive to the below-water signal in the visible bands. The decrease in RMA offset for DSF + GC compared to DSF further confirms the improvement offered by the glint correction.

3.5. Turbidity time-series and matchups

In total, 1857 scenes from Landsat and Sentinel-2 covering the Warp location were obtained, of which ~1184 were acquired during the SmartBuoy deployment history (2000–present). Of these, 390 passed the automated quality control and were used to compute water turbidity. Time-series of in situ and satellite-derived turbidity are presented in Fig. 11. An annual turbidity cycle is measured in situ, with high turbidities in winter (most of the time > 25 FNU), and low turbidities in summer (typically < 25 FNU). The satellite observations follow this cycle quite closely, although there is a strong sampling bias for the clearer summer conditions (see also Fettweis and Nechad, 2011), with less cloud cover, calmer conditions, and a more favourable solar illumination. The in situ measurements indicate a strong temporal variability (on the scale from minutes to hours), due to the site location in the tidally influenced and shallow southern North Sea. A full time-series of L5 and L7 turbidity (1984–2018, respectively 328 and 151 products) is provided in Supplementary Data 5, showing a quite consistent turbidity retrieval over the years with a similar seasonal cycle.

In total, 278 scenes with bounding in situ measurements were available, 82/99/48 for L5/7/8 and 39/10 for S2A/B. Overview scatterplots for the red and NIR bands across the sensors are shown in Fig. 12. Matchup plots per sensor are provided in Supplementary Data 6. Encouraging results are found for the red band derived turbidity, with a global RMSD across sensors of 11.2 FNU (MARD of 45%). The 10, 50, and 90th percentiles of the matchup data were 3.7, 14.2, and 36.2 FNU for the in situ measurements, and 5.3, 17.4, and 41.7 FNU for the satellite data. For the NIR band, derived turbidity errors are larger, giving an RMSD or 18.1 FNU and MARD of 63% across sensors. This is mainly caused by the poor performance of the older sensors, where L5 and L7 have a NIR band RMSD of 25.6 and 16.0 and a MARD of 78% and 63%. For the newer generation of sensors (L8/S2A/S2B), the red and NIR turbidity retrievals are much more consistent, with an RMSD of 6.4 –11.6 FNU for the red and 5.6 –12.4 FNU for the NIR band, giving a MARD of respectively 35.5–47% and 40–56 %. The RMSD from the reflectance matchups can also be translated into uncertainties for turbidity retrieval using the algorithm of Nechad et al. (2009). A red band RMSD of 0.007 gives uncertainties < 3 FNU at low turbidities (0–12 FNU) and 3–10 FNU at moderate turbidities (12–50 FNU). At high turbidity, the uncertainty increases rapidly due to the asymptote of the turbidity model. These ranges translate into a general applicability of these sensors for turbidity retrieval with uncertainties < 25% in the range of 12–110 FNU. A red band RMSD of 0.005 would extend this range to 7–195 FNU. Reflectance uncertainties of similar magnitude were found for the NIR bands, which would result in larger errors in a turbidity retrieval algorithm. For most AERONET-OC sites, the turbidity

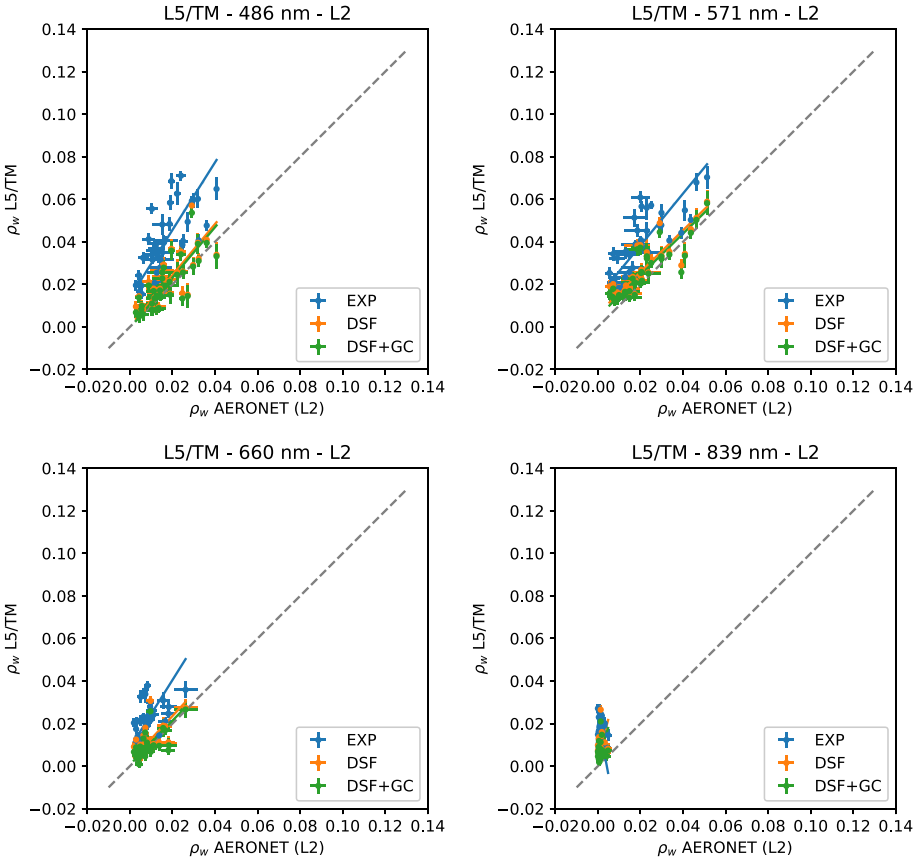

Fig. 7. Landsat 5/TM matchups with Level 2 AERONET-OC data using the ρ_t 1.6 µm > 0.01 rejection criterion for the satellite data. The solid lines are the Reduced Major Axis regression lines and the grey dashed line the 1:1 line. Error bars are the standard deviation (vertical) in a 3 × 3 pixel box around the station, and (horizontal) the available in situ data in the daylight window containing the overpass time.

Table 3
Same as Table 2, but for Landsat 7/ETM+.

Band	ac	Level	n	m	b	R^2	RMSD	MAD	MARD
479	EXP	L2	121	1.69	$9.13 \cdot 10^{-3}$	0.61	$2.26 \cdot 10^{-2}$	$1.95 \cdot 10^{-2}$	0.83
	DSF	L2	138	1.03	$2.91 \cdot 10^{-3}$	0.69	$6.99 \cdot 10^{-3}$	$3.36 \cdot 10^{-3}$	0.34
	DSF + GC	L2	138	1.04	$1.53 \cdot 10^{-3}$	0.71	$6.31 \cdot 10^{-3}$	$2.08 \cdot 10^{-3}$	0.31
561	EXP	L2	121	1.17	$9.72 \cdot 10^{-3}$	0.69	$1.58 \cdot 10^{-2}$	$1.30 \cdot 10^{-2}$	0.57
	DSF	L2	138	0.88	$5.62 \cdot 10^{-3}$	0.80	$6.93 \cdot 10^{-3}$	$3.29 \cdot 10^{-3}$	0.28
	DSF + GC	L2	138	0.89	$3.83 \cdot 10^{-3}$	0.82	$6.13 \cdot 10^{-3}$	$1.67 \cdot 10^{-3}$	0.24
661	EXP	L2	121	1.57	$7.97 \cdot 10^{-3}$	0.38	$1.41 \cdot 10^{-2}$	$1.20 \cdot 10^{-2}$	0.98
	DSF	L2	138	1.04	$4.65 \cdot 10^{-3}$	0.46	$6.90 \cdot 10^{-3}$	$4.96 \cdot 10^{-3}$	0.61
	DSF + GC	L2	138	1.04	$2.80 \cdot 10^{-3}$	0.50	$5.52 \cdot 10^{-3}$	$3.10 \cdot 10^{-3}$	0.48
835	EXP	L2	121	4.52	$6.00 \cdot 10^{-3}$	0.03	$1.23 \cdot 10^{-2}$	$1.08 \cdot 10^{-2}$	1.59
	DSF	L2	138	2.59	$3.66 \cdot 10^{-3}$	0.00	$6.96 \cdot 10^{-3}$	$5.90 \cdot 10^{-3}$	1.37
	DSF + GC	L2	138	2.37	$1.92 \cdot 10^{-3}$	0.00	$5.11 \cdot 10^{-3}$	$3.85 \cdot 10^{-3}$	1.17

is too low to warrant the switch to a longer wavelength in the turbidity retrieval.

4. Discussion

4.1. Adaptation of the DSF

The Dark Spectrum Fitting (DSF) algorithm by Vanhellemont and Ruddick (2018) was successfully adapted for the atmospheric correction of decametre scale Landsat and Sentinel-2 (L/S2) imagery. Although the spatial resolution of these satellite sensors is not sufficient to resolve ground-level object shadows, the DSF was able to estimate the atmospheric path reflectance thanks to the more extensive band set, especially thanks to the bands in the SWIR. A new aerosol model selection criterion was adopted, improving on the approach of Vanhellemont and Ruddick (2018) who used the model giving the lowest τ_a as a result of the limited band set on metre-scale sensors. For L/S2 processing, the lowest RMSD between the ρ_{path} and the ρ_{dark} for any band pair including the selected band is used as model selection criterion. A tiled processing was adopted allowing the path reflectance to vary over the satellite scene, with no significant addition of noise if the SWIR bands are used in the atmospheric correction. The tiled processing allows for smooth ρ_{path} retrievals over varied land and water targets.

The general assumptions used in the DSF are similar to current marine atmospheric correction methods, for example the existence of "black" water pixels in the NIR (Gordon and Wang, 1994) and SWIR (Gao et al., 2007; Wang, 2007), and the spatial homogeneity of aerosols (Hu et al., 2000; Ruddick et al., 2000), but are here combined in a novel and generic way. The DSF uses no a priori selection of the "black" band, but rather the optimal band is selected during the processing. Furthermore, the method could use land based targets for the AC if this would give a better estimate of the ρ_{path}. For extremely turbid waters, waters with significant sun glint, or for inland waters with high

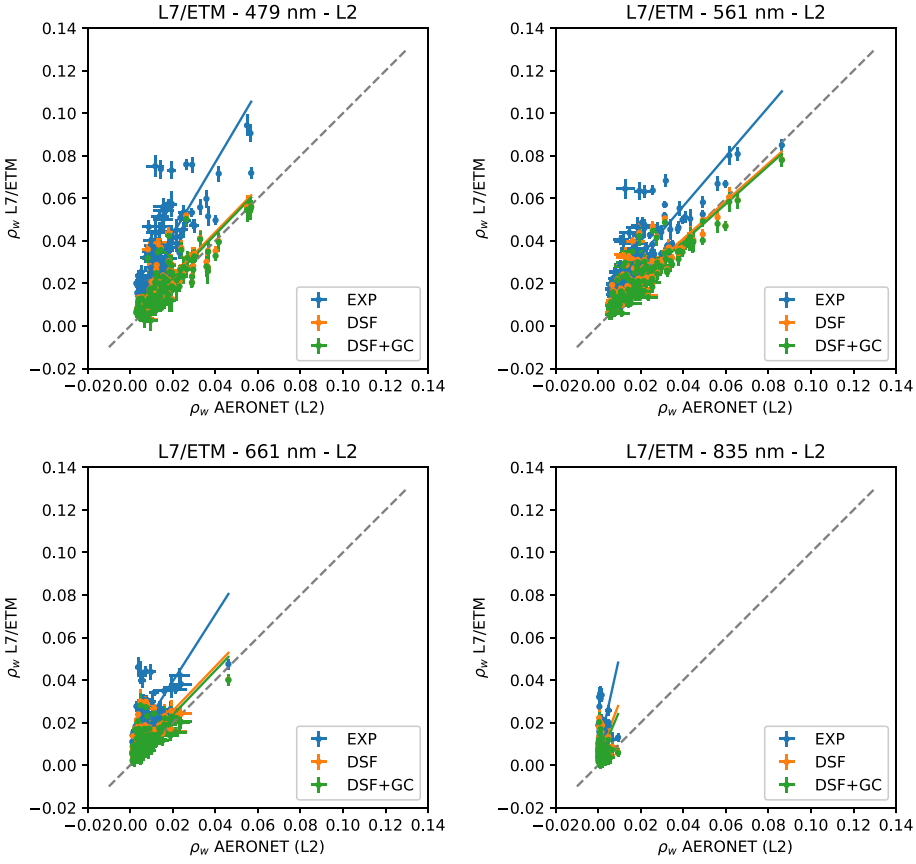

Fig. 8. Same as Fig. 7, but for Landsat 7/ETM+.

Table 4
Same as Table 2, but for Landsat 8/OLI.

Band	ac	Level	n	m	b	R^2	RMSD	MAD	MARD
443	EXP	L2	98	1.23	$1.45 \cdot 10^{-2}$	0.52	$1.88 \cdot 10^{-2}$	$1.72 \cdot 10^{-2}$	0.94
	DSF	L2	98	1.26	$3.96 \cdot 10^{-3}$	0.44	$1.10 \cdot 10^{-2}$	$6.99 \cdot 10^{-3}$	0.60
	DSF + GC	L2	98	1.24	$3.07 \cdot 10^{-3}$	0.46	$1.01 \cdot 10^{-2}$	$5.86 \cdot 10^{-3}$	0.57
483	EXP	L2	98	1.07	$1.29 \cdot 10^{-2}$	0.73	$1.56 \cdot 10^{-2}$	$1.40 \cdot 10^{-2}$	0.71
	DSF	L2	98	1.09	$6.37 \cdot 10^{-3}$	0.72	$1.05 \cdot 10^{-2}$	$7.74 \cdot 10^{-3}$	0.48
	DSF + GC	L2	98	1.08	$5.16 \cdot 10^{-3}$	0.74	$9.29 \cdot 10^{-3}$	$6.41 \cdot 10^{-3}$	0.43
561	EXP	L2	98	0.86	$8.84 \cdot 10^{-3}$	0.90	$8.41 \cdot 10^{-3}$	$5.76 \cdot 10^{-3}$	0.37
	DSF	L2	98	0.92	$5.42 \cdot 10^{-3}$	0.90	$7.04 \cdot 10^{-3}$	$3.70 \cdot 10^{-3}$	0.28
	DSF + GC	L2	98	0.92	$3.80 \cdot 10^{-3}$	0.91	$5.97 \cdot 10^{-3}$	$2.07 \cdot 10^{-3}$	0.24
655	EXP	L2	98	0.80	$5.77 \cdot 10^{-3}$	0.87	$5.90 \cdot 10^{-3}$	$3.89 \cdot 10^{-3}$	0.55
	DSF	L2	98	0.88	$4.29 \cdot 10^{-3}$	0.81	$5.86 \cdot 10^{-3}$	$3.17 \cdot 10^{-3}$	0.49
	DSF + GC	L2	98	0.87	$2.52 \cdot 10^{-3}$	0.85	$4.60 \cdot 10^{-3}$	$1.34 \cdot 10^{-3}$	0.45
865	EXP	L2	98	1.38	$2.82 \cdot 10^{-3}$	0.08	$4.03 \cdot 10^{-3}$	$3.26 \cdot 10^{-3}$	1.27
	DSF	L2	98	1.87	$2.85 \cdot 10^{-3}$	0.03	$5.03 \cdot 10^{-3}$	$3.87 \cdot 10^{-3}$	1.28
	DSF + GC	L2	98	1.38	$1.39 \cdot 10^{-3}$	0.10	$2.94 \cdot 10^{-3}$	$1.83 \cdot 10^{-3}$	1.26

adjacency effects, the use of a dark land target and a blue band rather than a water target and a NIR or SWIR band can result in a lower error on the ρ_{path} estimate. For example, for a dark target and a known τ_a, a same magnitude, spectrally flat reflectance error, caused e.g. by non-zero ρ_s or the presence of haze/glint, will give different impacts when a blue or a NIR band is used in the estimation of the aerosol optical thickness. For a typical aerosol model, a same absolute error would result in a lower difference between the true and the retrieved τ_a for the blue than for the NIR band, due to the larger atmospheric signal in the blue. This is in essence the reason why the DSF may shift between bands to retrieve the lowest ρ_{path} for a given dark spectrum.

Sun glint is very common in nadir viewing satellite sensors, and contaminates many coastal scenes in the Landsat and Sentinel-2 archives. Sun glint effects are typically amplified in the EXP method, by assuming the entire SWIR signal to be atmospheric. This can cause underestimation of ρ_s or even result in retrieval negative reflectances after atmospheric correction. The DSF avoids severe glint in the estimation of path reflectance, by dynamically choosing the bands used to derive ρ_{path}. The glint effects are still present in the DSF derived ρ_s, but the glint signal can be estimated in the SWIR bands that should have no signal from below the water surface (Harmel et al., 2018; Wang, 2007). With a successful glint correction, the DSF has the possibility to provide more useful data from the Landsat and Sentinel-2 archives. The glint correction is not enabled by default, as in some cases, the residual SWIR reflectance may be caused by adjacency effects. Perhaps using two SWIR bands it may be possible to estimate both adjacency and glint effects from the DSF retrieved ρ_s.

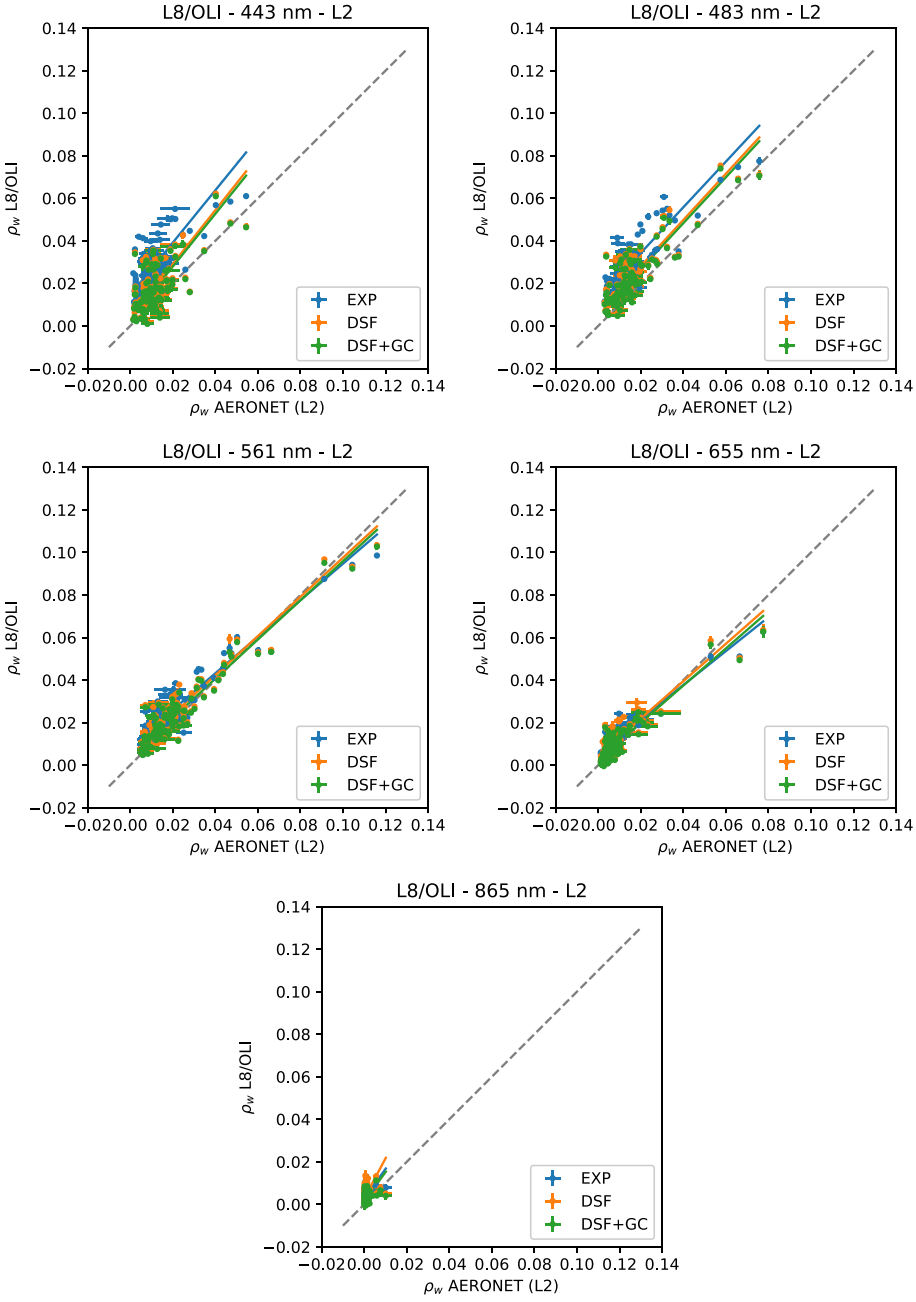

Fig. 9. Same as Fig. 7, but for Landsat 8/OLI.

4.2. Evaluation of the atmospheric correction

The performance of the DSF atmospheric correction adapted to L/S2 was evaluated here using two sets of in situ data: (1) AERONET-OC radiometry from sites worldwide and (2) turbidity data from the Warp SmartBuoy. In the comparison with radiometry, the DSF with and without glint correction was compared, as well as the previous EXP algorithm. Satellite data were filtered based on the extracted 3 × 3 box mean average ρ_t at 1.6 µm.

4.2.1. Radiometry

Matchup results were presented for lev20 AERONET-OC data measured by sites worldwide, for EXP, DSF, and DSF + GC. Overall, the DSF gives better correspondence than the EXP with the in situ reflectance in the VIS bands (blue-red) across all sensors. Lower absolute and relative errors are found, with RMA regression slopes closer to 1, with lower offsets and higher linear correlation coefficients. The improvement is especially noticeable in the blue bands around 440 and 480 nm, where the EXP generally underestimates the ρ_{path} and hence overestimates the ρ_w. Exceptions to this general trend are found where the data ranges are very small, and linear correlation statistics are not very meaningful (e.g. blue band on L5, and NIR bands across the board). Matchup performance is best for the green and red channels, which have generally the highest reflectance range, and which are of general of interest to retrieve turbidity at low to moderate turbidities (Novoa et al., 2017; Shen et al., 2010). Lower performance in the blue channels is to be expected, as the largest atmospheric signal has to be removed. The NIR performance is generally the worst, which is a combination of the low expected water signal, larger relative noise level, and large spectral width of the bands on these sensors (Fig. 1). The glint correction further improves on the DSF, and DSF + GC generally gives the best performance out of all three algorithms. Glint effects can be quite obvious in the

Table 5
Same as Table 2, but for Sentinel-2A/MSI.

Band	ac	Level	n	m	b	R^2	RMSD	MAD	MARD
444	EXP	L2	55	1.51	$1.38 \cdot 10^{-2}$	0.39	$2.10 \cdot 10^{-2}$	$1.90 \cdot 10^{-2}$	1.01
	DSF	L2	55	1.48	$4.67 \cdot 10^{-3}$	0.41	$1.29 \cdot 10^{-2}$	$9.59 \cdot 10^{-3}$	0.69
	DSF + GC	L2	55	1.48	$3.49 \cdot 10^{-3}$	0.40	$1.21 \cdot 10^{-2}$	$8.43 \cdot 10^{-3}$	0.76
497	EXP	L2	55	1.10	$1.16 \cdot 10^{-2}$	0.68	$1.47 \cdot 10^{-2}$	$1.32 \cdot 10^{-2}$	0.67
	DSF	L2	55	1.10	$6.18 \cdot 10^{-3}$	0.74	$9.81 \cdot 10^{-3}$	$7.77 \cdot 10^{-3}$	0.46
	DSF + GC	L2	55	1.10	$4.80 \cdot 10^{-3}$	0.72	$8.82 \cdot 10^{-3}$	$6.35 \cdot 10^{-3}$	0.42
560	EXP	L2	55	0.88	$9.43 \cdot 10^{-3}$	0.83	$8.92 \cdot 10^{-3}$	$6.99 \cdot 10^{-3}$	0.37
	DSF	L2	55	0.93	$6.14 \cdot 10^{-3}$	0.87	$6.88 \cdot 10^{-3}$	$4.75 \cdot 10^{-3}$	0.28
	DSF + GC	L2	55	0.92	$4.69 \cdot 10^{-3}$	0.86	$6.01 \cdot 10^{-3}$	$3.07 \cdot 10^{-3}$	0.24
664	EXP	L2	55	1.03	$6.36 \cdot 10^{-3}$	0.64	$7.51 \cdot 10^{-3}$	$6.54 \cdot 10^{-3}$	0.75
	DSF	L2	55	1.13	$4.90 \cdot 10^{-3}$	0.71	$6.75 \cdot 10^{-3}$	$5.77 \cdot 10^{-3}$	0.68
	DSF + GC	L2	55	1.11	$3.13 \cdot 10^{-3}$	0.69	$5.27 \cdot 10^{-3}$	$3.85 \cdot 10^{-3}$	0.53
865	EXP	L2	55	4.67	$6.70 \cdot 10^{-3}$	0.45	$1.07 \cdot 10^{-2}$	$9.90 \cdot 10^{-3}$	1.73
	DSF	L2	55	5.31	$5.79 \cdot 10^{-3}$	0.68	$1.05 \cdot 10^{-2}$	$9.54 \cdot 10^{-3}$	1.72
	DSF + GC	L2	55	5.18	$3.94 \cdot 10^{-3}$	0.61	$8.77 \cdot 10^{-3}$	$7.58 \cdot 10^{-3}$	1.63

matchup scatterplots, where a vertical trend can be observed in the EXP and DSF results. This vertical trend largely disappears in the DSF + GC results as this glint is accounted for. The most severe glint effects are excluded from the matchup dataset due to the filtering based on the ρ_t at 1.6 µm.

4.2.2. Glint correction

By applying the SWIR based sun glint correction after the DSF, the correspondence to the in situ measurements improves: similar RMA slopes and correlation coefficients are retrieved, while the RMA offset and the measures of differences (RMSD, MAD, MARD) both are reduced. The extension of the DSF with a SWIR based sun glint correction overall results in more useful data from L/S2 archive, especially for lower latitude sites, where most nadir viewing imagery is highly contaminated by glint. For clearer water scenes, where ρ_w can be assumed zero in the NIR, a glint correction based on the NIR band may be applied as well, typically leading to less noise introduced by the glint correction. It should be noted that by disabling the constraint on ρ_t 1.6 µm in the data filtering, the number of matchups increases by about 25–50 %, depending on the sensor, but performance degrades significantly (not shown), especially for EXP and DSF. The results for the DSF + GC with and without data filtering applied are most comparable. However, since the ρ_t 1.6 µm threshold does not only remove sun glinted scenes, but also those with cloud and haze, there are significantly more outliers when relaxing the data filtering. A more robust method for pixel identification and satellite data quality control is needed.

4.2.3. Bidirectional effects

No correction for the bidirectionality of the water has been performed for either the in situ or the satellite data. The sun zenith angle for both satellite and in situ are comparable due to the restriction on matchup timing, but there are some differences between the viewing zenith and azimuth angle relative to the sun that should be kept in mind. The bidirectional reflectance model of Park and Ruddick (2005) was used to estimate the reflectances for the typical AERONET-OC in-air viewing zenith angle of 40° and the range of viewing zenith angles, 0–7.5° for Landsat, and 0–12 ° for Sentinel-2. Park and Ruddick (2005) present the bidirectional water reflectance for a given view and illumination geometry as a function of the backscattering albedo, ω_b, and a phase function parameter, γ_b, equivalent to the fraction of particulate to total backscatter, $\frac{b_{bp}}{b_b}$. For turbid waters ($\omega_b = 0.4$, $\gamma_b = 0.99$), the relative differences were found to be < 5% for sun zenith angles of 0–30 ° and < 8% for sun zenith angles up to 60° when viewing perpendicular to the solar principal plane. Relative differences were < 5% viewing along the principal plane with the sun at the back of the sensor for sun angles up to 60°. For clear waters ($\omega_b = 0.04$, $\gamma_b = 0.6$) the differences where generally < 3% for sun angles between 0 and 30 ° in the perpendicular plane, but could reach 16% when viewing along the principal plane. In the present study, a 5–8 % difference between the in situ and satellite could be expected due to the lack of correction of bidirectionality in both datasets.

4.2.4. Resampling of AERONET-OC data

Another source of uncertainties in the matchup analysis is caused by the comparison of broad bands on the satellite sensors with the narrow spectral bands on the in situ stations. The in situ data are interpolated to form a hyperspectral dataset, which is then resampled to the sensor bands and finally shifted by dividing by a correction factor derived from simulated data (presented in Supplementary Data 1). In general, the interpolated dataset underestimates the true broad-band value, and the shifts to broad-band equivalent reflectance are < 1. For the blue bands around 440 and 490 nm, the broad-band reflectance is well-represented by the interpolated AERONET-OC values, with a mean average relative difference (MARD) of 0.9–1.3 % across sensors. For these bands a slight overestimation is found and a low correction factor is applied (1.00 –1.01). This good correspondence is largely caused by the good coverage by the CIMEL instrument in the blue spectral range (bands at 413, 441, 488, and 530 nm). Additionally, in turbid waters the spectral variability in the blue part of the spectrum is generally quite low due to saturation effects (Doxaran et al., 2002; Luo et al., 2018), and as a result of this spectral stability a more reliable comparison between broad and narrow bands can be made. The green and red bands at 560 and 660 nm show a MARD ranging from 6.7 to 9.1 % and 9.1–22.9 % respectively, with the largest differences in the red band for the Landsat sensors (14.9–22.9 %). Correction factors are accordingly larger for the red and green bands, generally between 0.87 and 0.94. The errors for the interpolated data are low for the Landsat 8 and Sentinel-2 NIR bands (3.5–7.5 % MARD) as they are quite narrow and closely located to an AERONET-OC band at around 865 nm, although the correction factors are quite large (0.75 –0.82). The NIR bands on L5 and L7 show much larger expected errors, respectively 65 and 67%, with however lower correction factors applied (0.95 and 0.97). This is mainly due to the lack of spectral coverage by AERONET-OC inside these broad bands. The convolution of the linearly interpolated in situ dataset and band shifting to broad-band equivalent reflectance used in the present paper gives a reasonable performance, but ideally hyperspectral in situ data should be used for the validation of broad-band sensors. If the absorption and scattering properties of the water are known, other band shifting methods could be attempted, e.g. through knowledge of inherent optical properties (Mélin and Sclep, 2015) or chlorophyll-a concentration (Zibordi et al., 2015).

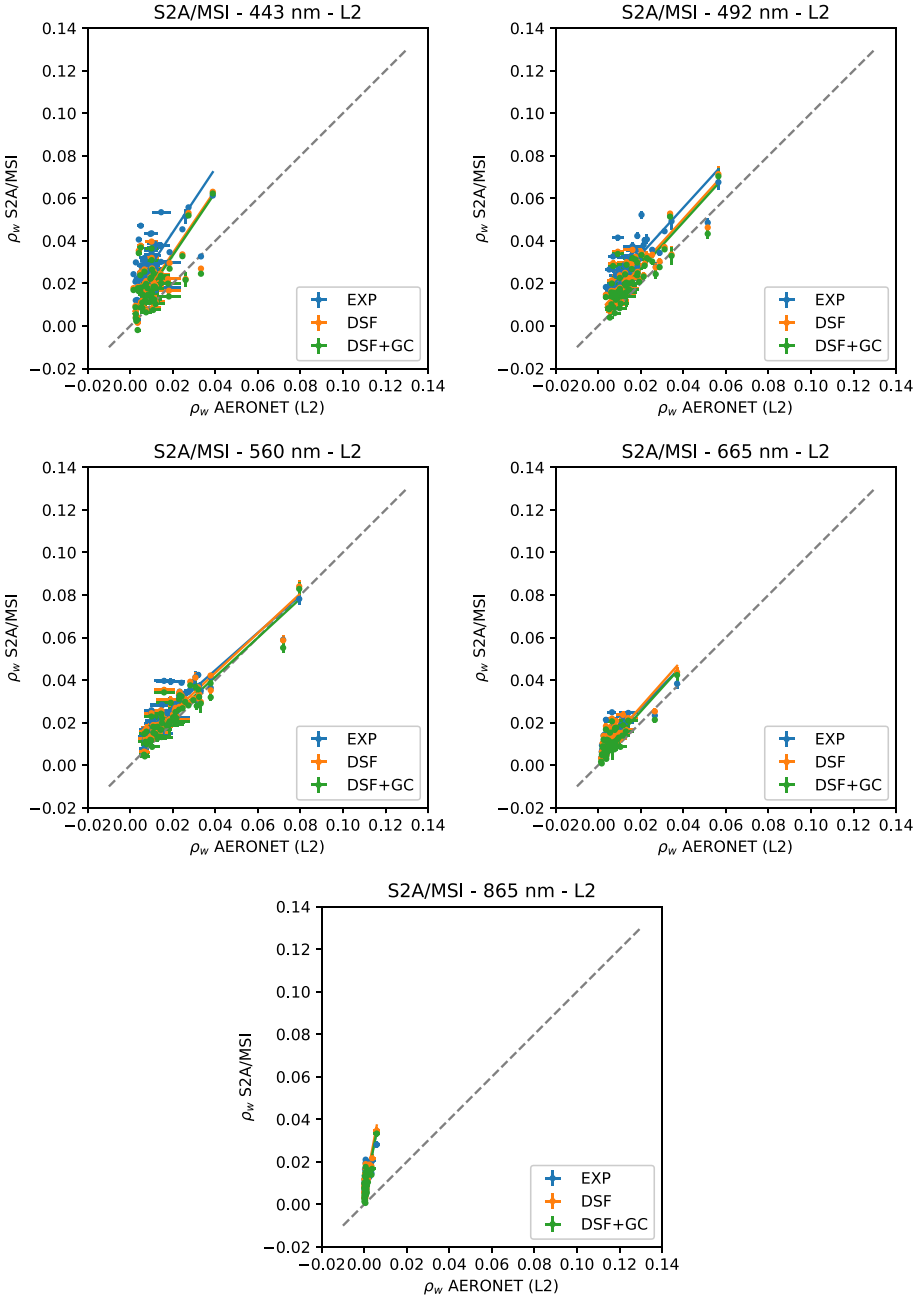

Fig. 10. Same as Fig. 7, but for Sentinel-2A/MSI.

4.2.5. Adjacency effects

In the matchups presented here, the satellite derived NIR reflectance is generally much higher compared to the in situ reflectance. This unlikely to be caused by the lack of spectral coverage by the in situ radiometer and the performed band shifting alone. Although the resampling of the NIR bands gives noisier results (especially for L5 and L7), the linear interpolation should both overestimate (L8, S2A, S2B) and underestimate (L5, L7) the true band averaged reflectance. The larger range observed in the NIR data could be caused by the calibration performance at low radiances, by residual glint signal, or by environmental or adjacency effects, i.e. the atmospheric scattering of photons from bright NIR targets (e.g. land or clouds) into the field of view of the sensor. These adjacency effects are more pronounced in the NIR and SWIR, due to the larger contrast between water (nearly black) and land or clouds in these bands. The DSF largely avoids amplification of adjacency effects in the atmospheric correction, by shifting from NIR or SWIR bands to lesser affected bands (e.g. blue and red for vegetation) to determine the aerosol optical thickness. The adjacency effects will be still present, but could potentially be estimated from the residual signal in the NIR and SWIR, much like the glint correction presented here. In the tiled processing, a background signal could be computed per tile, to estimate the spectral shape of the adjacency effect. Further investigation is needed to correct these adjacency effects, e.g. by using assumed spectral relationships of water (Sterckx et al., 2011).

4.2.6. Data ranges

The reflectance range and the used sites are not consistent between all sensors and this also introduces variability in the performance estimates per sensor. For example the reflectance ranges in the matchup data are much larger for L7, L8 and S2A (0–0.10) than for L5 and S2B (0–0.05), leading to better linear correlation coefficients (which are quite sensitive to data range), but larger RMSD values (due to the

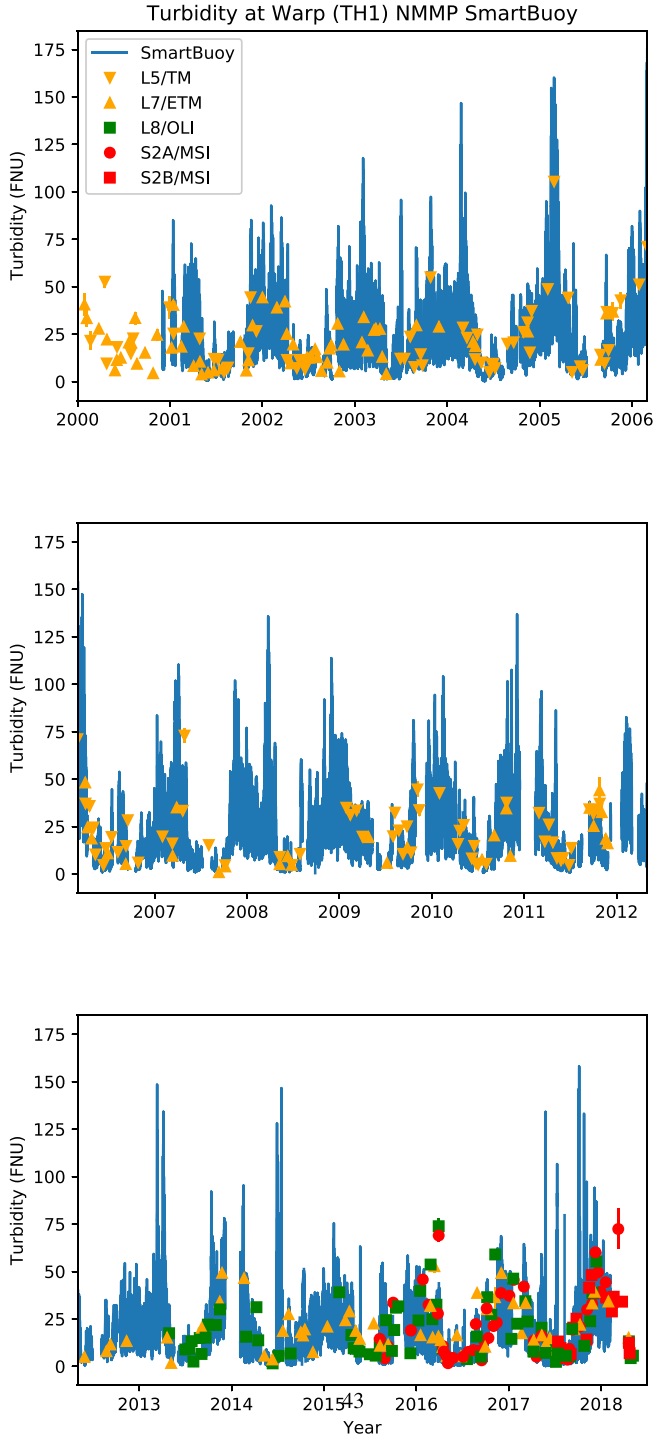

Fig. 11. Time-series of turbidity measured at the Warp location by in situ SmartBuoy deployments (solid blue line) and turbidity estimated from the red bands on Landsat and Sentinel-2 (various symbols). The vertical error bars are the standard deviation in a 3 × 3 pixel box around the station. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

inclusion of higher reflectances). The matchups with lev15 data (shown in Supplementary Data 4), have a higher data range, but have a similar performance to the lev20 matchups in terms of RMA slope, and may even have lower RMSD, MAD, and MARD. The matchup coordinates were manually shifted from the reported station position, to increase valid pixel availability from the satellite imagery, and to reduce the effects of the platforms themselves (especially for the larger platforms such as Gloria, LISCO, Thornton-C_Power). The used shifts are < 300 m and it is expected that coastal waters are largely homogeneous over this length scale, although some very high spatial variability can be found in shallow environments (Dorji and Fearns, 2017; Vanhellemont and Ruddick, 2018). Furthermore, the in situ measurements are not timed exactly at the satellite overpass and for tidally influenced sites this could introduce additional errors by the interpolation of the bounding measurements. In tidally influenced systems, the temporal variability may significantly exceed the spatial variability around the site (e.g. Vanhellemont et al., 2013, 2014).

4.2.7. Alternative methods

Similar performances are reported for matchups with AERONET-OC data and the iCOR processing of L8/OLI and S2A/MSI by De Keukelaere et al. (2018), who find a slightly higher RMSD compared to DSF and DSF + GC of 8.5×10^{-3} for the green and red bands, and $1.2–1.4 \times 10^{-2}$ for the blue bands. Their reported relative errors are lower likely due to the more equal distribution of high and low reflectance points in their 14 image matchup dataset. For S2A processing, the GRS algorithm by Harmel et al. (2018) used an external τ_a (measured or modeled) and a SWIR-based image derived sun glint correction. Because of the different units used, a direct comparison between the GRS and the DSF + GC is difficult, although the Normalised Root Mean Squared Difference (NRMSD, in %) reported by Harmel et al. (2018) compares well with the Mean Average Relative Differences (MARD, in %) reported here for the DSF + GC. Comparisons per band for respectively GRS and DSF + GC show similar relative differences with in situ data: 78% and 72% for B1 (442 nm), 38% and 42% for B2 (492 nm), 32% and 32% for B3 (560 nm), 74% and 57% for B4 (664 nm) and 569 and 172 % for B9. It should be noted that the GRS algorithm depends on inputs of measured or simulated τ_a, while the DSF is fully image based. The multi-sensor processing system used by the NASA Ocean Biology Processing Group, SeaDAS/l2gen was also adapted for processing of L8 (Franz et al., 2015; Pahlevan et al., 2017c) and S2A (Pahlevan et al., 2017b). The red band RMSD from the DSF + GC matchups presented here are larger than those presented by Pahlevan et al. (2017c) and (Pahlevan et al., 2017b). They found RMSD of respectively 1.5 and 2.2×10^{-3} for L8 and S2A, using their preferred atmospheric correction method selecting the aerosol model using the 865 nm and 1.6 μ m channels. Higher RMSD values are retrieved with the DSF + GC processing presented here, respectively 4.7 and 5.7×10^{-3} for L8/OLI and S2A/MSI. Results from Pahlevan et al. (2017b,c) focus on clearer waters, and include red band reflectances generally < 0.03, and < 0.01 for the bulk of their points, while the upper limits in the datasets presented here are 0.06 –0.10. (Quoted values were converted to ρ_w from the reported Rrs: $\rho_w = Rrs \cdot \pi$.) For the red band on L8, Pahlevan et al. (2017c) present median percentage differences of 59% before and 161% after the application of vicarious gains, whereas in the present paper the MARD (with no application of gains) for this band is 48% (DSF + GC). For the S2A red band, Pahlevan et al. (2017b) present a relative difference of 24%, while in the present paper a 57% MARD is retrieved. The slopes from the results presented here are significantly better (especially for L8), which is likely also a result of the larger reflectance range. Pahlevan et al. (2018) present the SeaDAS/l2gen processing of L5 and L7, with a mean relative difference (after application of vicarious gains) to matchup AERONET-OC measurements in the visible bands of 10–14 % for L5 and 1–10 % for L7, with a reported RMSD of 2.4 –5.6 $\times 10^{-3}$ for both sensors (converted here to ρ_w). The DSF + GC RMSD for these sensors and bands is 4.5 –7.5 $\times 10^{-3}$. It should be noted that the ranges of in situ data used by Pahlevan et al. (2018) are also much lower than the one presented here, with red reflectances for their matchups generally < 0.015. All SeaDAS/l2gen processing improved dramatically with the application of system calibration gains, but still large variability of the performance is found between sensors (especially in terms of mean relative difference). Overall, results from the present study compare well with other

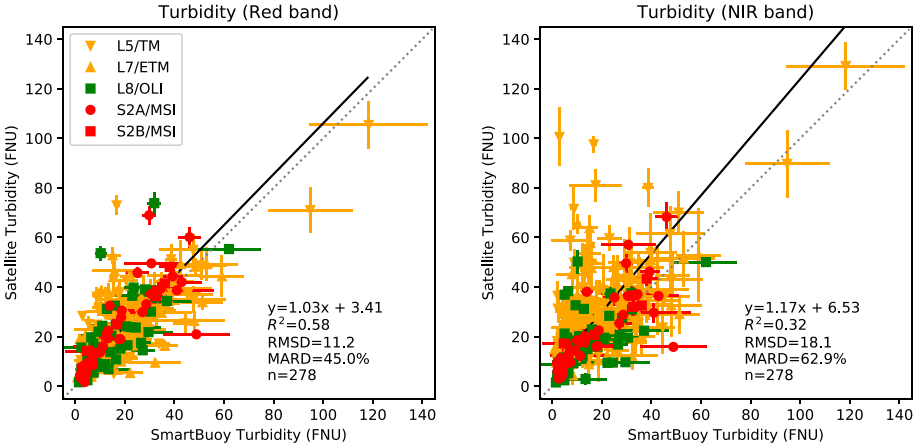

Fig. 12. A scatterplot of turbidity measured at the Warp location by in situ SmartBuoy deployments and turbidity estimated from the red (left) and NIR (right) bands on Landsat and Sentinel-2. The vertical error bars are the standard deviation in a 3 × 3 pixel box around the station. The horizontal error bars are the in situ standard deviation in a 6 hour window centred on the overpass time. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

approaches specifically for water processing. Likely, the different subsets of matchup data due to both in situ and satellite data filtering, the application of system vicarious gains, and differences in the band-shifting of in situ data, plays a significant role in the observed differences between the DSF and other methods. The strength of the DSF method is the unified processing of these 5 satellite sensors without the requirement of external τ_a inputs, while dynamically switching to the most appropriate bands over turbid waters. The optional glint performance further improves the comparison with in situ data. In the interest of water turbidity retrieval in sediment-dominated waters, a non-zero NIR water reflectance is to be expected, and hence a fixed NIR-SWIR atmospheric correction cannot be generally recommended.

4.2.8. Turbidity

A nearly 35 year time series of remotely sensed turbidity was presented (Supplementary Data 5). Such a long time-series opens perspectives on the study of impacts of port and offshore constructions, and the deepening of navigation channels performed since the mid-1980s. The last 18 years of the time-series were validated with in situ measurements from the Warp SmartBuoy (Figs. 11 and 12 and Supplementary Data 6). There is some uncertainty related to the actual buoy location between and within deployments, high variability at small spatial scales around the buoy, and variability in the first metre below the water surface (Vanhellemont et al., 2013). Nonetheless, acceptable errors are found, with an MARD of 45% for the red band retrieval. The RMSD from the red band turbidity matchups (11.2 FNU) corresponds quite closely to the one expected from the AERONET-OC measurements (3–10 FNU for turbidities of 10–50). The AERONET-OC matchups typically have a large subset of lower reflectance measurements compared to the turbidity matchups. The AERONET-OC red band reflectances are typically < 0.03, indicating turbidities are generally < 10 FNU, while 61% of the turbidity matchups have an in situ turbidity > 10 FNU. Although using a longer wavelength turbidity product is only needed at higher turbidities (Dogliotti et al., 2015; Novoa et al., 2017), the NIR derived turbidity was consistent with the red band derived turbidity for the newer satellite sensors (L8 and S2A/B). The NIR derived turbidity was much noisier with larger errors for L5/7. The NIR bands on TM and ETM+ are very wide and thus incorporate significant amounts of spectral variability. In combination with the rather poor signal-to-noise specification for water applications this leads to the considerable uncertainty in the NIR based turbidity retrieval for L5/7. For both the red and NIR bands, the glint correction is essential for lower latitude areas and the mid-latitudes in summer.

5. Conclusions

- The dark spectrum fitting (DSF) algorithm was successfully adapted to Landsat and Sentinel-2 satellites, allowing for the consistent processing of > 30 years of data from Landsat 5/7/8 and Sentinel-2A/B. Tiled processing over large scenes allows for a smooth ρ_{path} retrieval over land, coastal and inland water sites. The DSF algorithm is now available as the default and recommended atmospheric correction algorithm in ACOLITE.
- Landsat and Sentinel-2 provide largely nadir-viewing imagery, and are often contaminated by sun glint. An image based glint correction of the derived ρ_s may improve the availability of useable data from the archives, and can be applied after the DSF. It is not applied by default as for inland and coastal sites adjacency effects may also influence the retrieved SWIR reflectance (e.g. Sterckx et al., 2015).
- In comparison with autonomous AERONET-OC measurements, the DSF performed better than the EXP method for all satellite sensors, especially in the blue part of the spectrum. The DSF with glint correction performed best from all three algorithms, with an average difference to the in situ data of 24–29 % in the green band and 31–73 % in the other visible bands. A further comparison with other atmospheric correction algorithms is planned in a follow up to the ACIX exercise (Doxani et al., 2018).
- Satellite turbidity compared well with that measured by an autonomous in situ station over the past 18 years. The red band derived turbidity had an RMSD of 11.2 FNU (MARD 45%) over all sensors, demonstrating the usefulness of the entire nearly 35 year L/S2 archive for practical studies of water turbidity. The NIR band retrievals had a global RMSD of 18.1 FNU (MARD 63%), mainly due to the poorer performance of L5/7, which have very wide spectral bands and a relatively poor signal-to-noise specification. For more recent sensors (L8 and S2A/B), the red and NIR derived turbidity values were more consistent.

Acknowledgments

This work was funded by the Federal Belgian Science Policy Office (BELSPO) under the STEREO III programme PONDER project (SR/00/325). NASA/USGS and ESA/Copernicus are thanked for the acquisition and free distribution of Landsat and Sentinel-2 images used in this work. Google is thanked for the fast and easy distribution of satellite data via Google Cloud Services. The operators of the AERONET-OC stations worldwide are kindly thanked for their efforts in keeping the stations running and for the use of their data: Giuseppe Zibordi and the Joint Research Centre of the European Commission(Abu_Al_Bukhoosh, Galata_Platform, Gloria, Gustav_Dalen_Tower, Helsinki_Lighthouse, Venise), Rodney Forster (Blyth_NOAH), Brent Holben (COVE_SEAPRISM, GOT_Seaprism, USC_SEAPRISM_2), Tim Moore, Steve Ruberg, and Menghua Wang (Lake_Erie), Sam Ahmed and Alex Gilerson (LISCO), Thomas Schroeder (Lucinda), Hui Feng and Heidi Sosik (MVCO), Susanne Kratzer (Palgrunden), Dimitry Van der Zande (Thornton_C-power, Zeebrugge-MOW1), Burton Jones, Curtiss Davis,

and Nick Tufillaro (USC_SEAPRISM), and Alan Weidemann, Bill Gibson and Robert Arnone (WaveCIS_Site_CSI_6). The Australian Integrated Marine Observing System (IMOS) and CSIRO are acknowledged for funding the LJCO AERONET-OC site. CEFAS is kindly thanked for operating the SmartBuoy network and distributing the data freely. Kevin Ruddick is thanked for feedback on a first draft of this manuscript. The Associate Editor and four anonymous reviewers are thanked for their feedback which considerably improved the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rse.2019.03.010.

References

Antoine, D., Morel, A., 1999. A multiple scattering algorithm for atmospheric correction of remotely sensed ocean colour (MERIS instrument): principle and implementation for atmospheres carrying various aerosols including absorbing ones. Int. J. Remote Sens. 20 (9), 1875–1916.

Bailey, S.W., Franz, B.A., Werdell, P.J., 2010. Estimation of near-infrared water-leaving reflectance for satellite ocean color data processing. Optics express 18 (7), 7521–7527.

Bernardo, N., Watanabe, F., Rodrigues, T., Alcântara, E., 2017. Atmospheric correction issues for retrieving total suspended matter concentrations in inland waters using OLI/landsat-8 image. Adv. Space Res. 59 (9), 2335–2348.

Braga, F., Zaggia, L., Bellafiore, D., Bresciani, M., Giardino, C., Lorenzetti, G., Maicu, F., Manzo, C., Riminucci, F., Ravaioli, M., et al., 2017. Mapping turbidity patterns in the Po River prodelta using multi-temporal Landsat 8 imagery. Estuar. Coast. Shelf Sci. 198, 555–567.

Brando, V.E., Braga, F., Zaggia, L., Giardino, C., Bresciani, M., Matta, E., Bellafiore, D., Ferrarin, C., Maicu, F., Benetazzo, A., et al., 2015. High-resolution satellite turbidity and sea surface temperature observations of river plume interactions during a significant flood event. Ocean Sci. 11 (6), 909.

Caballero, I., Navarro, G., Ruiz, J., 2018. Multi-platform assessment of turbidity plumes during dredging operations in a major estuarine system. International Journal of Applied Earth Observation and Geoinformation 68, 31–41.

Castagna, A., Simis, S., Dierssen, H., Vanhellemont, Q., Sabbe, K., Vyverman, W., 2018. Extending the operational land imager/landsat 8 for freshwater research: retrieval of an orange band from PAN and MS bands. In: .

De Liesbeth, K., Sterckx, S., Adriaensen, S., Knaeps, E., Reusen, I., Giardino, C., Bresciani, M., Hunter, P., Neil, C., Van der Zande, D., Vaiciute, D., 2018. Atmospheric correction of landsat-8/OLI and sentinel-2/MSI data using iCOR algorithm: validation for coastal and inland waters. European Journal of Remote Sensing 51 (1), 525–542.

Dogliotti, A.I., Ruddick, K., Guerrero, R., 2016. Seasonal and inter-annual turbidity variability in the Río de la Plata from 15 years of MODIS: El Niño dilution effect. Estuar. Coast. Shelf Sci. 182, 27–39.

Dogliotti, A.I., Ruddick, K.G., Nechad, B., Doxaran, D., Knaeps, E., 2015. A single algorithm to retrieve turbidity from remotely-sensed data in all coastal and estuarine waters. Remote Sens. Environ. 156, 157–168.

Dorji, P., Fearns, P., 2017. Impact of the spatial resolution of satellite remote sensing sensors in the quantification of total suspended sediment concentration: a case study in turbid waters of Northern Western Australia. PloS one 12 (4), e0175042.

Dörnhöfer, K., Göritz, A., Gege, P., Pflug, B., Oppelt, N., 2016. Water constituents and water depth retrieval from Sentinel-2A — a first evaluation in an oligotrophic lake. Remote Sens. 8 (11), 941.

Doxani, G., Vermote, E., Roger, J.C., Gascon, F., Adriaensen, S., Frantz, D., Hagolle, O., Hollstein, A., Kirches, G., Li, F., et al., 2018. Atmospheric correction inter-comparison exercise. Remote Sens. 10 (2), 352.

Doxaran, D., Froidefond, J.M., Lavender, S., Castaing, P., 2002. Spectral signature of highly turbid waters: application with SPOT data to quantify suspended particulate matter concentrations. Remote sensing of Environment 81 (1), 149–161.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., et al., 2012. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. Remote Sensing of Environment 120, 25–36.

Fettweis, M.P., Nechad, B., 2011. Evaluation of in situ and remote sensing sampling methods for SPM concentrations, Belgian continental shelf (southern North Sea). Ocean Dyn. 61 (2-3), 157–171.

Franz, B.A., Bailey, S.W., Kuring, N, Werdell, P.J., 2015. Ocean color measurements with the Operational Land Imager on Landsat-8: implementation and evaluation in seaDAS. J. Appl. Remote. Sens. 9 (1), 096070.

Gao, B.C., Montes, M.J., Li, R.R, Dierssen, H.M., Davis, C.O., 2007. An atmospheric correction algorithm for remote sensing of bright coastal waters using MODIS land and ocean channels in the solar spectral region. IEEE Trans. Geosci. Remote Sens. 45 (6), 1835–1843.

Gordon, H.R., Wang, M., 1994. Retrieval of water-leaving radiance and aerosol optical thickness over the oceans with seawiFS: a preliminary algorithm. Applied Optics 33 (3), 443–452.

Harmel, T., Chami, M., Tormos, T., Reynaud, N., Danis, P.A., 2018. Sunglint correction of the multi-spectral instrument (MSI)-SENTINEL-2 imagery over inland and sea waters from SWIR bands. Remote Sens. Environ. 204, 308–321.

Hedley, J.D., Harborne, A.R., Mumby, P.J., 2005. Simple and robust removal of sun glint for mapping shallow-water benthos. Int. J. Remote Sens. 26 (10), 2107–2112.

Hu, C., Carder, K.L., Muller-Karger, F.E., 2000. Atmospheric correction of seawiFS imagery over turbid coastal waters: a practical method. Remote Sensing of Environment 74 (2), 195–206.

Kotchenova, S.Y., Vermote, E.F., Matarrese, R., Klemm Jr, F.J., 2006. Validation of a vector version of the 6s radiative transfer code for atmospheric correction of satellite data. Part I: path radiance. Applied Optics 45 (26), 6762–6774.

Lee, Z., Shang, S., Qi, L., Yan, J., Lin, G., 2016. A semi-analytical scheme to estimate Secchi-disk depth from landsat-8 measurements. Remote Sensing of Environment 177, 101–106.

Liu, H., Li, Q., Shi, T., Hu, S., Wu, G., Zhou, Q., 2017. Application of sentinel 2 MSI images to retrieve suspended particulate matter concentrations in Poyang Lake. Remote Sens. 9 (7), 761.

Luo, Y., Doxaran, D., Ruddick, K., Shen, F., Gentili, B., Yan, L., Huang, H., 2018. Saturation of water reflectance in extremely turbid media based on field measurements, satellite data and bio-optical modelling. Optics Express 26 (8), 10435–10451.

Manzo, C., Federica, B., Luca, Z., Ernesto, B.V., Claudia, G., Mariano, B., Cristiana, B., 2018. Spatio-temporal analysis of prodelta dynamics by means of new satellite generation: the case of Po River by landsat-8 data. International Journal of Applied Earth Observation and Geoinformation 66, 210–225.

Martins, V.S., Barbosa, C.C.F., Sander de Carvalho, L.A., Jorge, D.S.F., Lobo, F.L., Novo, E., 2017. Assessment of atmospheric correction methods for sentinel-2 MSI images applied to amazon floodplain lakes. Remote Sens. 9 (4), 322.

Mélin, F., Sclep, G., 2015. Band shifting for ocean color multi-spectral reflectance data. Optics Express 23 (3), 2262–2279.

Mills, D.K., Laane, R.W.P.M., Rees, J.M., Rutgers van der Loeff, M., Suylen, J.M., Pearce, D.J., Sivyer, D.B., Heins, C., Platt, K., Rawlinson, M., 2003. Smartbuoy: a marine environmental monitoring buoy with a difference. In: Elsevier Oceanography Series. 69. Elsevier, pp. 311–316.

Nechad, B., Ruddick, K.G., Neukermans, G., 2009. Calibration and validation of a generic multisensor algorithm for mapping of turbidity in coastal waters. In: SPIE Europe Remote Sensing, pp. 74730H.

Nechad, B., Ruddick, K.G., Schroeder, T., Oubelkheir, K., Blondeau-Patissier, D., Cherukuru, R.C.N., Brando, V.E., Dekker, A.G., Clementson, L.A., Banks, A.C., et al., 2015. Coastcolour round robin datasets: a database to evaluate the performance of algorithms for the retrieval of water quality parameters in coastal waters. Earth System Science Data 7, 319–348.

Neukermans, G., Ruddick, K., Bernard, E., Ramon, D., Nechad, B., Deschamps, P.Y., 2009. Mapping total suspended matter from geostationary satellites: a feasibility study with SEVIRI in the southern North Sea. Opt. Express 17 (16), 14029–14052.

Novoa, S., Doxaran, D., Ody, A., Vanhellemont, Q., Lafon, V., Lubac, B., Gernez, P., 2017. Atmospheric corrections and multi-conditional algorithm for multi-sensor remote sensing of suspended particulate matter in low-to-high turbidity levels coastal waters. Remote Sens. 9 (1), 61.

Ody, A., Doxaran, D., Vanhellemont, Q., Nechad, B., Novoa, S., Many, G., Bourrin, F., Verney, R., Pairaud, I., Gentili, B., 2016. Potential of high spatial and temporal ocean color satellite data to study the dynamics of suspended particles in a micro-tidal river plume. Remote Sens. 8 (3), 245.

Pahlevan, N., Balasubramanian, S., Sarkar, S., Franz, B., 2018. Toward long-term aquatic science products from heritage landsat missions. Remote Sens. 10 (9), 1337.

Pahlevan, N., Roger, J.C., Ahmad, Z., 2017a. Revisiting short-wave-infrared (SWIR) bands for atmospheric correction in coastal waters. Opt. Express 25 (6), 6015–6035.

Pahlevan, N., Sarkar, S., Franz, B.A., Balasubramanian, S.V., He, J., 2017b. Sentinel-2 multispectral instrument (MSI) data processing for aquatic science applications: demonstrations and validations. Remote Sens. Environ. 201, 47–56.

Pahlevan, N., Schott, J.R., Franz, B.A., Zibordi, G., Markham, B., Bailey, S., Schaaf, C.B., Ondrusek, M., Greb, S., Strait, C.M., 2017c. Landsat 8 remote sensing reflectance (rrs) products: evaluations, intercomparisons, and enhancements. Remote Sens. Environ. 190, 289–301.

Park, Y.J., Ruddick, K., 2005. Model of remote-sensing reflectance including bidirectional effects for case 1 and case 2 waters. Appl. Opt. 44 (7), 1236–1249.

Röttgers, R., Doerffer, R., McKee, D., Schönfeld, W., 2016. The water optical properties processor (WOPP): pure water spectral absorption, scattering, and real part of refractive index model.

Roy, D.P., Wulder, M.A., Loveland, T.R., Woodcock, C.E., Allen, R.G., Anderson, M.C., Helder, D., Irons, J.R., Johnson, D.M., Kennedy, R., et al., 2014. Landsat-8: science and product vision for terrestrial global change research. Remote Sensing of Environment 145, 154–172.

Ruddick, K.G., De Vera, C., Park, Y.J., Moore, G., et al., 2006. Seaborne measurements of near infrared water-leaving reflectance: the similarity spectrum for turbid waters. Limnol. Oceanogr 51 (2), 1167–1179.

Ruddick, K.G., Ovidio, F., Rijkeboer, M., 2000. Atmospheric correction of seawiFS imagery for turbid coastal and inland waters. Applied Optics 39 (6), 897–912.

Schiller, H., Doerffer, R., 1999. Neural network for emulation of an inverse model operational derivation of case II water properties from MERIS data. International Journal of Remote Sensing 20 (9), 1735–1746.

Schroeder, T., Behnert, I., Schaale, M., Fischer, J., Doerffer, R., 2007. Atmospheric correction algorithm for MERIS above case-2 waters. Int. J. Remote Sens. 28 (7), 1469–1486.

Shen, F., Verhoef, W., Zhou, Y., Salama, M.S., Liu, X., 2010. Satellite estimates of wide-range suspended sediment concentrations in Changjiang (Yangtze) estuary using MERIS data. Estuaries Coasts 33 (6), 1420–1429.

Steinmetz, F., Deschamps, P.Y., Ramon, D., 2011. Atmospheric correction in presence of sun glint: application to MERIS. Optics Express 19 (10), 9783–9800.

Sterckx, S., Knaeps, E., Ruddick, K., 2011. Detection and correction of adjacency effects in hyperspectral airborne data of coastal and inland waters: the use of the near infrared similarity spectrum. International Journal of Remote Sensing 32 (21), 6479–6505.
Sterckx, S., Knaeps, S., Kratzer, S., Ruddick, K., 2015. SIMilarity Environment Correction (SIMEC) applied to MERIS data over inland and coastal waters. Remote Sens. Environ. 157, 96–110.
Thuillier, G., Hersé, M., Labs, D., Foujols, T., Peetermans, W., Gillotay, D., Simon, P.C., Mandel, H., 2003. The solar spectral irradiance from 200 to 2400 nm as measured by the SOLSPEC spectrometer from the ATLAS and EURECA missions. Sol. Phys. 214 (1), 1–22.
Vanhellemont, Q., Greenwood, N., Ruddick, K., 2013. Validation of MERIS-derived turbidity and PAR attenuation using autonomous buoy data. ESA Special Publication 722, 12.
Vanhellemont, Q., Neukermans, G., Ruddick, K., 2014. Synergy between polar-orbiting and geostationary sensors: remote sensing of the ocean at high spatial and high temporal resolution. Remote Sens. Environ. 146, 49–62.
Vanhellemont, Q., Ruddick, K., 2014. Turbid wakes associated with offshore wind turbines observed with Landsat 8. Remote Sens. Environ. 145, 105–115.
Vanhellemont, Q., Ruddick, K., 2015a. Advantages of high quality SWIR bands for ocean colour processing: examples from Landsat-8. Remote Sens. Environ. 161, 89–106.
Vanhellemont, Q., Ruddick, K., 2015b. Pan-sharpening to improve spatial resolution of optical remote sensing with examples from Landsat-8 (30 m/15 m). In: Poster Presented at the International Ocean Color Science Meeting, Held 15–18 June 2015, San Francisco. Online Reference: Https://Www.Eposters.Net/Poster/Pan-Sharpening-To-Improve-Spatial-Resolution -Of-Optical-Remote-Sensing-With-Examples-From-Landsat-8.
Vanhellemont, Q., Ruddick, K., 2016. ACOLITE for Sentinel-2: Aquatic Applications of MSI Imagery. In: ESA Special Publication SP-740. Presented at the 1 Living Planet Symposium Held in Prague, Czech Republic.
Vanhellemont, Q., Ruddick, K., 2018. Atmospheric correction of metre-scale optical satellite data for inland and coastal water applications. Remote Sens. Environ. 216, 586–597. https://doi.org/10.1016/j.rse.2018.07.015.
Vermote, E., Tanré, D., Deuzé, J.L., Herman, M., Morcrette, J.J., Kotchenova, S.Y., 2006. Second simulation of a satellite signal in the solar spectrum-vector (6SV). 6S User Guide Version 3, 1–55.
Wang, M., 2007. Remote sensing of the ocean contributions from ultraviolet to near-infrared using the shortwave infrared bands: simulations. Appl. Opt. 46 (9), 1535–1547.
Wang, M., Shi, W., 2006. Cloud masking for ocean color data processing in the coastal regions. IEEE Trans. Geosci. Remote Sens. 44 (11) 3196-3105.
Woodcock, C.E., Allen, R., Anderson, M., Belward, A., Bindschadler, R., Cohen, W., Gao, F., Goward, S.N., Helder, D., Helmer, E., et al., 2008. Free access to Landsat imagery. Science 320 (5879), 1011.
Wulder, M.A., White, J.C., Loveland, T.R., Woodcock, C.E., Belward, A.S., Cohen, W.B., Fosnight, E.A., Shaw, J., Masek, J.G., Roy, D.P., 2016. The global landsat archive: status, consolidation, and direction. Remote Sens. Environ. 185, 271–283.
Zibordi, G., Mélin, F., Berthon, J.F., Holben, B., Slutsker, I., Giles, D., D'Alimonte, D., Vandemark, D., Feng, H., Schuster, G., et al., 2009. AERONET-OC: a network for the validation of ocean color primary products. J. Atmos. Oceanic Tech. 26 (8), 1634–1651.
Zibordi, G., Mélin, F., Berthon, J.F., Talone, M., 2015. In situ autonomous optical radiometry measurements for satellite ocean color validation in the western Black Sea. Ocean Sci. 11 (2), 275–286.