Radiometric validation of atmospheric correction for MERIS in the Baltic Sea based on continuous observations from ships and AERONET-OC

Ping Qin[a,b], Stefan G.H. Simis[a], Gavin H. Tilstone[a,*]

[a] Plymouth Marine Laboratory, Prospect Place, The Hoe, Plymouth PL1 3DH, UK

[b] College of Information Science and Engineering, Ocean University of China, Qingdao 266100, China

*Corresponding author: ghti@pml.ac.uk

Running head: Atmospheric correction models for MERIS in the Baltic Sea

Keywords: Atmospheric Correction, Baltic Sea, Remote Sensing Reflectance, MERIS, AERONET-OC, Shipborne radiometry.

ABSTRACT

The Baltic Sea is a semi-enclosed sea that is optically dominated by coloured dissolved organic material (CDOM) and has relatively low sun elevation which makes accurate ocean colour remote sensing challenging in these waters. The high absorption, low scattering properties of the Baltic Sea are representative of other optically similar water bodies including the Arctic Ocean, Yellow Sea, Black Sea, coastal regions adjacent to the CDOM-rich estuaries such as the Amazon, and highly absorbing lakes where radiometric validation is essential in order to develop accurate remote sensing algorithms. Previous studies in this region mainly focused on the validation and improvement of standard Chlorophyll-a (Chl *a*) and attenuation coefficient (k_d) ocean colour products. The primary input to derive these is the water-leaving radiance (L_w) or remote sensing reflectance (R_{rs}) and it is therefore fundamental to obtain the most accurate L_w or R_{rs} before deriving higher level products. To this end, the retrieval accuracy of R_{rs} from Medium Resolution Imaging Spectrometer (MERIS) imagery using six atmospheric correction processors was assessed through above-water measurements at two sites of the Aerosol Robotic Network for Ocean Colour (AERONET-OC; 363 measurements) and a shipborne autonomous platform from which the highest number of measurements were obtained (4986 measurements). The six processors tested were the CoastColour processor (CC), the Case 2 Regional processor for lakes (C2R-Lakes), the Case 2 Regional CoastColour processor (C2R-CC), the FUB/WeW water processor (FUB), the MERIS ground segment processor (MEGS) and POLYMER. All processors except for CC had small average absolute percentage differences (ψ) in the wavelength range from 490 nm to 709 nm (ψ <40%), while other bands had larger differences with ψ > 60%. Compared to *in situ* values, the $R_{rs}(709)/R_{rs}(665)$ band ratio had ψ < 30% for all processors. The most accurate R_{rs} in the 490 to 709 nm domain was obtained from POLYMER with ψ < 30% and coefficients of determination (R^2) > 0.6. Using a score system based on all statistical tests, POLYMER scored highest, while C2R-CC, C2R-Lakes and FUB had lower scores. This study represents the largest data base of in situ R_{rs}, the most comprehensive analysis of AC models for highly absorbing waters and for MERIS, conducted to date. The results have implications for the new generation of Copernicus Sentinel ocean colour satellites.

1 Introduction

Remote sensing has become an important tool to monitor the dynamics of optically active substances in the marine environment due to high coverage at both spatial and temporal scales (IOCCG, 2000). Some bio-optical and geophysical variables, such as the concentration of chlorophyll *a* (Chl *a*) as an indicator of phytoplankton biomass, suspended particulate matter, coloured dissolved organic matter (CDOM), as well as the bulk inherent optical properties of the visible surface layer, have been successfully retrieved from water-leaving radiance (L_w) or remote sensing reflectance (R_{rs}). L_w or R_{rs} at the sea surface is derived from the top-of-atmosphere (TOA) radiance after atmospheric correction (AC). In principle, the more accurate L_w or R_{rs}, the more accurate will be the derived biogeochemical products. The performance of atmospheric correction is therefore key to quality assured ocean colour data for monitoring issues of water quality, carbon cycling and climate change.

Various AC methods have been developed for remote sensing of the open-ocean, coastal seas, and inland waters. In the open ocean, AC mainly rely on the black pixel hypothesis (Gordon and Wang 1994) which assumes that the marine reflectance in the near infrared (NIR, 700–1000 nm) is negligible due to the relatively high absorption of water itself. The TOA radiance in the NIR wavebands is further influenced by absorption and scattering from atmospheric aerosols, and the reflectance in the short visible domain (400–700 nm) is extrapolated through spectral aerosol models. The black pixel assumption is too simplistic for most inland and coastal waters since the contributions to R_{rs} of suspended particulates can cause R_{rs} in the near-infrared wavelength range to depart from zero (Ruddick *et al.*, 2000; Hu *et al.*, 2000; Knaeps *et al.*, 2012). Alternative AC methods have been proposed to cope with a variety of the Case 2 waters, including the black pixel method by means of the short wave infrared or ultraviolet wavebands (Wang & Shi, 2007; Siegel *et al.*, 2000; He *et al.*, 2012), spectral optimization that utilizes a bio-optical model in conjunction with radiative transfer models (Steinmetz *et al.*, 2011; Callieco & Dell'Acqua, 2011), and artificial neural networks (Schiller & Doerffer, 1999; Doerffer, 2007; Schroeder *et al.*, 2007; Brockmann *et al.*, 2016).

MERIS on the European Space Agency ENVISAT mission, during its operation in 2002–2012, offered a wide dynamic range of products for both marine and terrestrial observations. It provided global coverage in 3 days, with observations at 15 bands at visible and NIR wavelengths designed to observe both open-ocean and coastal environments. It also provided data at full (~ 300 m) and reduced (~ 1200 m) resolution. The MERIS era marked the start of long-term remote sensing observations of water quality in optically complex environments. A range of atmospheric correction processors were developed for MERIS, designed for a wide range of applications from coastal to inland waters. These include the CoastColour (CC) processor (Doerffer and Schiller, 2007), the Case 2 Regional (C2R) processor (Doerffer and Schiller, 2008), the FUB/WeW water (FUB) processor (Schroeder *et al.*, 2007), and the Case 2 Regional CoastColour (C2R-CC) processor

(Brockmann *et al.*, 2016). In addition, the default MERIS ground segment (MEGS) processor has been continually updated (Aiken & Moore, 2000) to reflect the performance of the MERIS instrument over its lifespan. An alternative polynomial based algorithm (POLYMER) (Steinmetz *et al.*, 2011) has been increasingly used with MERIS and other sensors, though it was not the primary choice for optically complex waters.

The high-CDOM waters of the Baltic Sea are characteristic of water bodies with high riverine input, long water retention times, but low mineral particle loading, such as the Arctic Ocean, Yellow Sea, Black Sea, coastal regions adjacent to the CDOM-rich estuaries such as the Amazon, and highly absorbing lakes. In these environments, reflectance at short visible wavelengths is particularly low and may contribute as little as 0.4% of the TOA radiance, compared to 9.8% over open ocean waters (IOCCG, 2010). The performance of AC processors dedicated to high absorbing coastal waters, have thus far not been as successful as those applied to turbid waters which have a stronger reflectance signal. Regional re-tuning of some AC processors has improved their performance in some highly absorbing waters (Attila *et al*., 2013).

Previous research in the Baltic Sea evaluated the performance of standard and Case 2-specific Chl a (Harvey et al. 2015; D'Alimonte et al. 2012; Kratzer et al. 2008; Melin et al. 2007; Reinart and Kutser 2006) and k_d (Stramska and Swirgon 2014; Doron et al. 2011; Pierson et al. 2008) ocean colour products. Regionally calibrated blue-green ratio versions of OC4v6 (Pitarch *et al.*, 2016; Darecki and Stramsk, 2004) have allegedly improved the accuracy of Chl *a* retrieval in the Baltic Sea, but do not work for waters where CDOM dominates the absorption in the blue. Using longer wavelengths such as red-to-green (Wozniak 2014) and red-to-near-infra red (Koponen et al. 2007; Krawczyk et al. 1997, Matthews 2011) is therefore advisable in these optically complex, CDOM-rich waters. Ligi *et al.* (2016) assessed 30 empirical remote sensing algorithms for retrieving Chl *a* in the Baltic Sea through modelled and *in situ* reflectance data, and found that NIR-red band ratio algorithms performed best. Few papers have considered the performance of and improving the accuracy of the primary input, L_w or R_{rs}, of SeaWiFS, MODIS-Aqua and MERIS, used to derive Chl *a* and k_d products (D'Alimonte et al. 2014; Zibordi et al. 2009; Kratzer et al. 2008; Melin et al. 2007; Darecki & Stramski 2004; Ohde et al. 2002). Some studies have improved the performance of regional specific Chl *a* algorithms for the Baltic Sea using FUB and C2R processors coupled to AC neural networks has been improved (Beltrán-Abaunza *et al.*, 2014; Attila *et al.* 2013; Kratzer et al. 2008). Melin et al. (2013) and Bulgarelli et al. (2003) also showed that improvements in the aerosol libraries used in the AC processors for MERIS and SeaWiFS also improves retrieval of R_{rs},. Some studies have shown that the accuracy of both the shape and amplitude of L_w or R_{rs} are required otherwise improvements in green to near infrared bands but failure in the blue bands may result in reasonable Chl *a* concentration retrieval but a failure in the retrieval of other products, such as absorption by CDOM.

Another common challenge to achieve this is obtaining sufficient in situ data to carry out a comprehensive analysis of satellite R_{rs}. Both MOBY (Voss et al. 2007), BOUSOLLE (Antoine et al. 2008) and AERONET-OC (Zibordi, *et al*, 2009b) have undoubtedly aided the global assessment of ocean colour products. These platforms are fixed structures, close to the coast, and though temporal coverage from them is good, spatial coverage is limited. A growing network of autonomous radiometers deployed on research ships and ships of opportunity such as ferries could potentially fill these spatial gaps in data coverage, provided that the same high quality measurements on shipborne platforms are achieved as on the fixed platforms. To this end in this paper, by combining shipborne and AERONET-OC measurements, and using a rigorous quality control procedure for the ship data (Simis and Olsson 2013), we use the use the largest data base to date of in situ R_{rs} to evaluate the performance, accuracy and suitability of six AC processors for MERIS for the Baltic Sea. The retrieval accuracy at each band and spectral shape of CC, C2R, C2R-CC, FUB, MEGS and POLYMER processors were evaluated against *in situ* R_{rs} from two AERONET-OC measurement platforms and a prototype platform for continuous shipborne reflectance measurements operated from a research vessel, which has since been installed on two merchant vessels on the Alg@line network managed by the Finnish Environment Institute (SYKE).. The suitability of each processor at different locations as well as the seasonal bias in retrieval of R_{rs}, was also compared.

2 Data and Methods

2.1 Study area

The Baltic Sea is a semi-enclosed brackish marine water body located in Northern Europe between the maritime temperate and continental sub-Arctic zones (Fig. 1), and has partial, seasonal sea-ice cover. It covers an area of ~400 000 km^2 which includes the Gulf of Bothnia, Gulf of Finland, Gulf of Riga, Gulf of Gdansk and Kattegat Bay. The mean water depth over the region is approximately 54 m and tides are negligible due to limited connectivity with the Atlantic Ocean. One of the main characteristics of the Baltic Sea is the salinity gradient that increases from the north with salinity < 1 PSU to the south-west with salinity up to > 20 PSU. Riverine input is large and seasonal, with annual mean river runoff of ~14000 m^3/s (Leppäranta & Myrberg, 2009; Omstedt *et al.*, 2004). Eutrophication and pollution are significant in the region due to the terrestrial input of nitrogen and phosphorus and the limited water exchange with the North Sea.

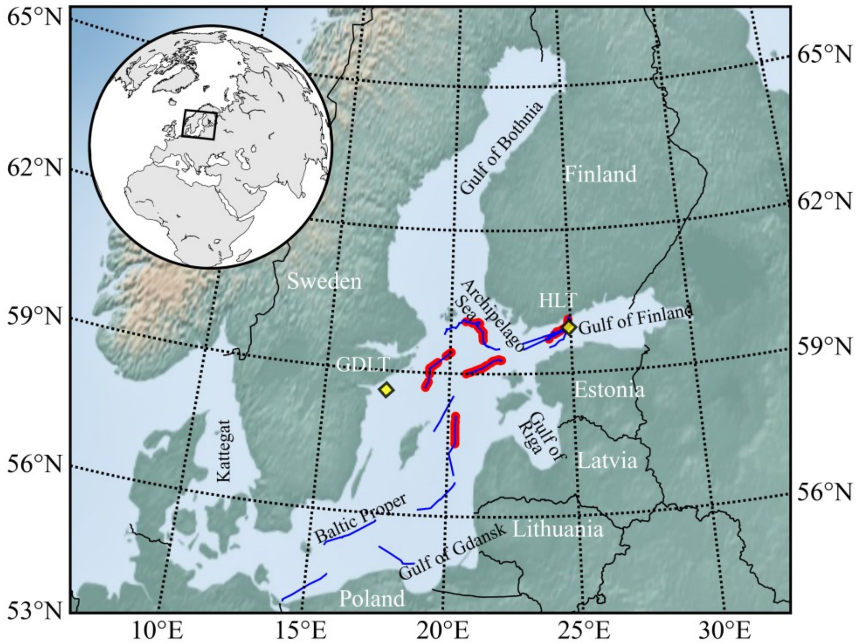

Fig. 1. Locations of *in situ* data from the research vessel (RV, blue lines) and two AERONET-OC sites (yellow markers): Gustaf Dalen Lighthouse Tower (GDLT) and Helsinki Lighthouse Tower (HLT). Red markers represent match-ups with the shipborne observations.

CDOM absorption coefficients at 440 nm are generally > 1.0 m^{-1}, with higher values in estuaries and bays, such as the Neva Bay where $a_{CDOM}(442)$ is 3.77 m^{-1} (Wozniak *et al*., 2014; Ylöstalo *et al*., 2016). There are generally two annual phytoplankton blooms in the Baltic Sea. The spring bloom is dominated by diatoms and dinoflagellates, and exhibits high peak biomass but this is generally short-lived from March to April. The summer bloom is dominated by cyanobacteria from July to September, when there is thermal stratification and cyanobacteria accumulate during prolonged calm weather (Kahru *et al.* 2015; Groetsch *et al*., 2014).

2.2 Shipborne observations

In situ radiometric observations from the shipborne platform were acquired during three cruises on R/V Aranda in the Baltic Sea during spring (April) 2011 and summer (July) 2010 and 2011 in the Gulf of Finland, the Baltic Proper and the Archipelago Sea (Fig. 1).

Three RAMSES spectro-radiometers (TriOS Optical Sensors, Rastede, Germany) were mounted on the bow of the research vessel. A RAMSES-ACC with cosine collector optics was directed upwards to record the downwelling irradiance above the water surface (E_d), and two RAMSES-ARC radiance sensors were used to measure the sky radiance (L_s) and the total upwelling radiance pointed at the surface of the water (L_t), at 140° and 40° zenith angles respectively. The azimuth angle in relation to the solar azimuth was kept as close to 135° as possible using a stepper motor platform to compensate for the solar azimuth (calculated from GPS time and location) and vessel heading, without pointing back at the ship, and was always > 90° (Simis & Olsson, 2013). Three sensors recorded the wavelength range of 320–950 nm with 3.3 nm spectral resolution and a field-of-view of 7° at 15-s intervals. Inter-calibration of the sensors was verified before each cruise by pointing the radiance sensors at a large white spectralon panel and simultaneously recording E_d on the roof of the laboratory on a day with clear skies. R_{rs} (sr^{-1}) was then calculated as follows:

$$R_{rs}(\lambda) = L_{w+}(\lambda)/E_d(\lambda) \tag{1}$$

$$L_{w+}(\lambda) = L_t(\lambda) - \rho_s L_s(\lambda) \tag{2}$$

where L_{w+} is the water-leaving radiance just above the sea surface and ρ_s is the reflectance of sky radiance at the air-water interface, which depends on solar azimuth angle, viewing geometry, wind speed, cloud and surface roughness (Mobley, 1999; Ruddick *et al.*, 2006; Mobley, 2015). Here ρ_s was determined using the fingerprint method (Simis & Olsson, 2013), a spectral optimization technique that minimizes the propagation of atmospheric absorption features to R_{rs} and flag observations that do not resolve to a smooth R_{rs} spectrum.

The shipborne reflectance underwent a secondary screening procedure to eliminate spurious observations based on assumptions of the spectral shapes of reflectance in the highly absorbing and weakly scattering waters of the Baltic Sea. The following threshold criteria were used: (1) the average R_{rs} in the ultraviolet range (350–400 nm) and near infra-red (800–900 nm) should not be significantly negative, i.e. $R_{rs}(350\text{–}400) \geq -0.0005$ sr^{-1} and $R_{rs}(800\text{–}900) \geq -0.0005$ sr^{-1}. (2) The maximum reflectance value was limited to $R_{rs}(\lambda) < 0.015$ sr^{-1}, which removed spectra strongly affected by sun glint, whitecaps, or spray. (3) Spectra were only considered valid if they retained a green reflectance peak, following the criterion $1.5R_{rs}(400) < R_{rs}(580) > 2R_{rs}(800)$. This shape of the spectrum is expected in CDOM-rich waters with minor contribution to scattering from mineral particles, such that CDOM and pure water absorption dominate the blue and near infra-red reflectance, respectively. (4) Following the same assumption, CDOM absorption increases towards shorter wavelengths, spectra were validated with the criterion $R_{rs}(412) < R_{rs}(443)$, which removed spectra affected by incomplete removal of reflected sky light causing a rise of reflectance in the blue. (5) Removal of spectra where the difference between the maximum and minimum R_{rs} in the 760–770 nm wavelengths was larger than

10% of the maximum R_{rs}(560–600), i.e. clearly showing an effect of the oxygen absorption peak. This set of filtering criteria applies specifically to conditions in the Baltic Sea and should be revised for other water bodies. Shipborne collection of $R_{rs}(\lambda)$ should be significantly less challenging in more turbid coastal waters with a higher amplitude of reflectance and lower errors associated with the removal of reflected sky radiance. Following this screening procedure, the $R_{rs}(\lambda)$ spectra are given in Figure 2.

The NIR reflectance is expected to be close to zero in waters with low particle scattering (Hooker *et al*., 2002). The NIR reflectance measured in the Baltic Sea may depart significantly from zero near to river plumes or when there is an accumulation of near-surface phytoplankton. In most cases, however, an offset from zero in the NIR will be primarily attributed to residual surface water effects (spray, sun glint, whitecaps, and sky radiance including scattered cloud reflected on waves). Removal of the offset in the NIR reflectance minimizes additional contamination in the signal and leads to a better correlation with the satellite signal. The fingerprint method to resolve $R_{rs}(\lambda)$ per definition accounts for direct and diffuse contributions to sky radiance reflected at the water surface. Any offset observed in the NIR that is not due to high particle scatter is expected to be spectrally neutral and can be compensated for, by subtracting this signal from the R_{rs} (λ). In theory, the validity of this assumption can be easily checked by evaluating the shape of the NIR signal. For high particle scattering, this shape should reflect the spectral dependence of water absorption. When this is not the case, high particle scattering cannot account for the NIR offset and may thus be subtracted. The shape of the NIR signal did not generally show a spectral dependence of water absorption in the Baltic Sea (results not shown). NIR offset-corrected $R_{rs}(\lambda)$ is here defined as $R_{rs}(\lambda)$ from which the average R_{rs} in the near infrared region (850–900 nm) is subtracted. The difference between performing and not performing offset correction was compared (Table 3).

2.3 AERONET-OC

AERONET-OC is a standardized measurement system installed on fixed platforms at a range of coastal locations to collect marine radiometric measurements coincident with aerosol measurements for retrieving aerosol optical properties (Zibordi *et al*., 2009b). Measurements from two AERONET-OC sites were used from April 2005 to October 2011: the Gustaf Dalén Lighthouse Tower (GDLT) in the northern Baltic Proper and the Helsinki Lighthouse Tower (HLT) in the Gulf of Finland (Fig. 1). AERONET-OC measures the radiances of sun, sky and sea water at 412–1020 nm using the modified CIMEL Electronique CE-318 autonomous sun photometers, known as Sea-Viewing Wide Field-of-View Sensor (SeaWiFS) Photometer Revision for Incident

Surface Measurements (SeaPRISM). The system adopts a sea-viewing zenith angle of 140° and relative azimuth of 90° with respect to the sun in the successive observations at each waveband. Radiometric measurements in the first six wavebands (412–667 nm) are used to obtain water leaving radiance, while bands at 870 nm and 1020 nm are used for quality checks and turbid water flagging for the application of alternative above-water methods (Zibordi *et al.*, 2009b).

The AERONET-OC data are processed at three levels (Level 1.0, 1.5 and 2.0) based on different quality assurances, in which Level 2.0 is fully quality-controlled including pre- and post-field calibration with differences smaller than 5%, automatic cloud removal, and manual inspection. AERONET-OC Level 2.0 data at GDLT and HLT were obtained from http://aeronet.gsfc.nasa.gov. For the present validation, the normalized water-leaving radiances ($L_{WN\text{-}f/Q}$) corrected for viewing angle dependence and for the effects of the non-isotropic distribution of the in-water radiance field, included in the AERONET-OC Level 2.0 data products, were selected (Fig. 2).

The AERONET-OC wavebands were designed for SeaWiFS which are slightly different to waveband centers for MERIS. The AERONET-OC waveband centers are 413, 441, 491, 555, 668 and 870 nm in HLT, and 412, 439, 500, 554, 675 and 870 nm in GDLT; while the related MERIS bands are centered at 412, 443, 490, 560, 665 and 865 nm. $L_{wn\text{-}f/Q}$ was band shift corrected based on regional bio-optical algorithms to reduce inter-band uncertainties. Further details of the methods are given in Zibordi *et al.* (2009a), where $L_{WN\text{-}f/Q}$ is a function of the ratio of total backscattering and absorption coefficients, and of the extra-atmospheric solar irradiance. The calculation of R_{rs} is subsequently derived from $L_{WN\text{-}f/Q}$ after band-shifting, as follows:

$$R_{rs}(\lambda) = \frac{L_{WN-f/Q}(\lambda)}{F_0(\lambda)} \tag{3}$$

where F_0 is the extra-atmospheric solar irradiance for each waveband (Thuillier *et al*, 2003).

2.4 MERIS AC processors.

MERIS full resolution level 1b products (3rd reprocessing) segmented into 0.5° × 0.5° tiles around *in situ* measurements were processed using the following atmospheric correction schemes: CC (v1.8.3), C2R-Lakes (v1.6), C2R-CC (v 0.15), FUB (v 2.2), MEGS (v 8.1) and POLYMER (v 3.5). The first four atmospheric correction processors are based on artificial neural network algorithms to derive the atmospherically corrected water-leaving reflectance from TOA radiances. Ancillary data with actual sea surface pressure and total ozone content values are utilized to calculate reflectance at the TOA. Water-leaving reflectance was estimated using the forward artificial neural network. The main differences between these four processors are the range of water constituents and inherent optical properties used in the datasets to train their respective neural networks.

The CC processor employed a wider range of optical properties in the training data (Doerffer and Schiller, 2007), and was developed for application in optically-complex coastal waters. C2R (Doerffer & Schiller, 2007) was intended as a generic AC processor for complex Case 2 waters, and includes two plugins for the inland water constituent retrieval optimized for boreal and eutrophic lakes (Doerffer and Schiller, 2008). The training data set was produced through the ocean-atmosphere Monte Carlo photon tracing model. The atmospheric component of the model used a standard atmosphere (1013.2 hPa atmospheric pressure and 350 Dobson units of ozone) with different aerosol models, cirrus cloud particles and a rough, wind dependent sea surface with reflectance. The atmospheric correction for these two plugins is identical and hereafter we refer to them as C2R-Lakes. The atmospheric model for C2R-Lakes, in turn, was developed for optically complex inland and coastal waters using a calibration dataset specific to these environments. Similar to CC, a Monte Carlo radiative transfer model was used to simulate the TOA radiance, which contained four aerosols models (continental, maritime, urban / industrial and stratospheric). C2R-CC is the latest in the evolution of these processors, and employs artificial neural networks for atmospheric correction using a large training database obtained by radiative transfer simulations (Brockmann *et al*., 2016). C2R-CC used a coastal aerosol model derived from coastal AERONET measurements (Aznay & Santer, 2009), and the atmospheric radiative transfer was calculated through a parameterised version of the successive order of scattering technique (Lenoble *et al*, 2007). A version of C2R-CC specifically trained for extreme combinations of inherent optical properties is also included, but has not been considered here. FUB was designed for European coastal waters and integrates the entire AC process in a single neural network to retrieve water leaving reflectance from the TOA radiances. The data set used to train the neural network was generated by the matrix operator method, using a mixture of maritime and continental aerosol models as well as an US standard atmosphere (Schroeder *et al*, 2007). The atmospheric correction scheme for FUB is divided into a Rayleigh-ozone correction and an atmospheric correction network. Water constituents and atmospheric properties are retrieved simultaneously from the TOA radiance, whereas the other processors firstly derive the reflectance, then calculate in-water parameters from the reflectance (Schroeder *et al.*, 2007). FUB provides the water-leaving reflectance at a subset of eight MERIS wavebands (412–665 nm and 709 nm).

MEGS was developed specifically for MERIS and has been regularly improved and updated, following vicarious calibrations of MERIS. It performs the black-pixel atmospheric correction for open oceanic waters with the low NIR R_{rs}, and uses the bright-pixel atmospheric correction for turbid waters based on the NIR R_{rs} with a fixed spectral shape (Antoine and Morel, 2011; Moore and Lavender, 2011). The optical properties of atmospheric aerosol are inferred from the near-infrared wavebands and the atmospheric contribution to the TOA signal is then extrapolated to the visible part of the spectrum. MEGS uses the spectra at near infrared wavelengths (778 and 865 nm) to calculate the aerosol radiance ratio assuming that the reflectance is null at the wavelength beyond 700 nm. The path radiance and its spectral shape in the visible wavebands is then determined by iterating the different aerosol models and then validated using water-leaving reflectance at 510 nm assuming an priori known constant for $R_{rs}(510)$ (Nobileau and Antoine, 2005).

POLYMER is a spectral optimization method using a polynomial atmospheric model and a bio-optical ocean water reflectance model. The atmospheric model simultaneously fits three components ranging from spectrally neutral (e.g. residual sun glint) to weak (λ^{-1}, aerosols) and strong (λ^{-4}, e.g. Rayleigh scatter) wavelength dependence. The bio-optical model only relies on Chl *a* concentration and the backscattering coefficient of non-covarying particles (newer versions of POLYMER also include a mineral absorption component, which is not considered relevant to the current data set). These five parameters are optimized to obtain the best approximation of the measurements in a configurable range of spectral bands. Version 3.5 of POLYMER was not specifically designed to handle optically complex coastal waters but includes a Case 2 water switch. The initial conditions for the Case-1 bio-optical model were changed to Chl *a* = 1 mg m^{-3} and total suspended matter = 1 g m^{-3} to avoid solutions designed for oceanic waters. The atmospheric model uses the visible and NIR wavebands to assess sun glint and aerosol scattering properties (Steinmetz *et al.*, 2011). A version of POLYMER (v4.1), with a modified bio-optical model which includes scattering by mineral particles was trialed, but not considered to be a significant improvement for the low-mineral laden waters of the Baltic Sea.

The main output of the six AC processors was the reflectance $\rho_w(\lambda)$, which was converted to $R_{rs}(\lambda)$ following:

$$R_{rs}(\lambda) = \frac{\rho_w(\lambda)}{\pi} \qquad (4)$$

A series of quality flags included with the output of each processor were used to define the validity of a pixel either according to the input L1B data or as processor-specific conditions. Invalid pixels were masked based on land, haze, whitecaps, cloud or sun glint contamination flags based on processing the L1B data with

Idepix v2.2.10. Processor specific flags included: poor fits to aerosol models; TOA radiances outside of the training or application range; and results surpassing the minimum or maximum concentration bounds. The flag combinations used with each processor are listed in Table 1.

Table 1. Quality flags for pixel exclusion criteria. l1 is level 1, l2 is level 2, l2r is level 2 reflectance, agc is atmospheric sun glint correction, aot560 is aerosol optical thickness at 560 nm, oor is out of range, toa is top of atmosphere, tosa is top of standard atmsophere, oos is out of scope, ooadb is aerosol model is out of aerosol model database, rtosa is reflectance at top of standard atmosphere, atm_in is atmospheric correction failure in input, atm_out atmospheric correction failure in output, pcd_1_13 is product confidence flag in bands 1 to 13, negative _bb is negative backscatter.

processor	flags	Names
	l1_flags	suspect, land_ocean, bright, coastline, invalid
CC	l2r_flags	aot560_oor, toa_oor, tosa_oor, tosa_oos
C2R_Lake	agc_flags	atc_oor, toa_oor, tosa_oor
C2R-CC	l2_flags	rtosa_oor, rtosa_oos
FUB	result_flags	atm_in, atm_out
MEGS	l2_flags	ooadb, pcd_1_13
POLYMER	bitmask	negative _bb, out_of_bounds, exception

2.5 Match-up procedure

Match-ups between *in situ* and MERIS retrieved R_{rs} were selected based on location and overpass time, as well as a spatial homogeneity criterion following Bailey and Werdell (2006), as outlined below.

Match-ups within ±12 hours between *in situ* shipborne observations and MERIS over-pass were extracted from the processed imagery in 3 × 3 pixel boxes using the nearest neighbour approach. Subsequently match-up time-windows of ±0.5 h to ±12 h were compared (Table 3) to obtain the best balance between the highest number of match-ups and reducing artefacts such as water mass and particle dynamics (including phytoplankton mobility). Due to the high sampling frequency from the ship, MERIS match-up pixels could correspond to multiple shipborne observations. In these cases the mean (μ) and standard deviation (σ) spectrum of any valid *in situ* observations was calculated. *In situ* observations which exceeded $\mu \pm 1.5\sigma$ were discarded to decrease the effects of the horizontal (and to an extent, temporal) non-homogeneity. The mean spectrum of the remaining observations matched to the same pixel was used for further analysis.

AERONET-OC Level 2.0 data were selected strictly within a ±2-h window around the satellite overpass. The shorter time window was chosen because the AERONET-OC observations did not directly acquire E_d and changing *in situ* illumination conditions could lead to invalid comparisons with R_{rs}. The AERONET-OC data were obtained from the average in observations using the procedures outlined above to filter for outliers.

The 3 × 3-pixel boxes centered on the *in situ* locations were extracted from the atmospherically corrected MERIS products. The MERIS retrieved R_{rs} were checked for spatial homogeneity to avoid the influence of severe spatial variability and abnormal values. Differences between the value of each valid pixel and their mean in the 3 × 3-pixel box were limited to twice the standard deviation to eliminate outliers. To meet the spatial homogeneity criterion (filtered standard deviation divided by the filtered mean), the coefficient of variation was set at < 0.15. If the number of remaining pixels in the 3 × 3-pixel box was less than 5, the observation was omitted. The mean of remaining pixels in the 3 × 3-pixel box was then calculated.

Approximately 12% of the shipborne *in situ* observations remained after stringent quality control, corresponding to 1947 individual MERIS pixels within the ±12-h window around the satellite overpass. The number of shipborne observations available for match-up analysis decreased further after applying specific quality flags for each AC processor. The number of match-up observations was 59 for CC, 602 for C2R-Lakes, 644 for C2R-CC, 256 for FUB, 427 for MEGS and 644 for POLYMER within the ±12-h window. From the AERONET-OC Level 2 data approximately 22% of the available data (363 observations) corresponded to the ±2-h window around the satellite overpass, which were all used in subsequent analyses.

Figure 2 gives all shipborne and AERONET-OC data meeting these criteria. Measurements ±3-h for shipborne data and ±2-h for AERONET-OC were subsequently used for accuracy assessment analysis given in Figures 3-9 and to compute the statistics given in Table 4 and Figure 10a, c, d. The number of retrievals differed for each AC processor. Table 5 and Figure 10b gives statistics using the same number of data for each AC processor using a threshold of N = 494.

2.6 Statistical indices

The differences between MERIS observations and *in situ* observations were quantified using a number of statistical metrics, including the coefficient of determination (R^2), the average absolute percentage difference (ψ), the root-mean-square difference (Δ) and the bias (δ) between MERIS and *in situ* match-ups, calculated as follows:

$$R^2 = \frac{\left(\sum\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)\right)^2}{\sum\left(x_i - \bar{x}\right)^2 \sum\left(y_i - \bar{y}\right)^2} \tag{5}$$

$$\psi = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - x_i}{x_i}\right| \times 100\% \tag{6}$$

$$\Delta = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i - x_i\right)^2} \tag{7}$$

$$\delta = \frac{1}{N}\sum_{i=1}^{N}\frac{y_i - x_i}{x_i} \tag{8}$$

where x_i is the i-th *in situ* observation, y_i is the i-th MERIS observation, and N is the number of match-ups.

R^2 is equal to the square of the correlation coefficient, representing a linear consistency between the *in situ* and MERIS observations, and the proportion of the variation that explained by the linear regression. Higher R^2 indicates a higher degree of correlation, whereas R^2 is sensitive to both outliers and narrow data distributions. Statistical significance of the correlation coefficient is tested using the student's distribution. The smaller the probability level of significance (p), the more significant the linear relationship between *in situ* and MERIS observations. Δ and ψ measures the accuracy of match-ups. ψ is the relative difference which is sensitive to small values while Δ is the absolute difference which is sensitive to outliers. Values of ψ and Δ close to zero indicate that MERIS observations compare well with the *in situ* observations. Bias δ is used to determine the underestimation or overestimation of MERIS products compared to the *in situ* data, with a value near zero indicating no systematic under- or over-estimation.

Type-2 linear regression was used to fit the *in situ* and MERIS observations for their independent randomness (Glover *et al.*, 2011; Brewin *et al.*, 2015). The slope (S) close to one and intercept (I) close to zero indicate that the MERIS observations fit well against the *in situ* observations.

2.7. AC processor ranking

A scoring scheme based on Brewin *et al.* (2015) and Müller *et al.* (2015) was employed to rank the relative performance of the AC processors. The score was obtained by comparing all statistical metrics (R^2, ψ, δ, Δ, S and I) for each waveband of each processor. The average score of all processors was compared against each individual processor. A score of <1 or >1 indicates significantly worse or better performance respectively.

A score ranging from zero to two for each statistical metric was assigned as follows:

(1) Zero points were assigned when: (i) R^2 was less than the mean of the lower 90% confidence intervals of all processors; (ii) each of ψ and Δ was higher than the mean of the upper 90% confidence interval; (iii)

each of δ and I overlapped with neither the mean 90% confidence interval nor zero ± twice the mean standard deviation; (iv) S overlapped with neither the mean 90% confidence interval nor one ± twice the mean standard deviation.

(2) One point was assigned when: (i) each of R^2, ψ and Δ overlapped with the mean 90% confidence interval; (ii) each of δ and I overlapped with either the mean 90% confidence interval or zero ± twice the mean standard deviation, but not both; (iii) S overlapped with either the mean 90% confidence interval or one ± twice the mean standard deviation for all processors, but not both.

(3) Two points were assigned when: (i) R^2 exceeded the upper limit of the mean 90% confidence interval; (ii) each of ψ and Δ was less than the lower limit of the mean 90% confidence interval; (iii) each of δ and I overlapped with both the mean 90% confidence interval and zero ± twice the mean standard deviation; (iv) S overlapped with both the mean 90% confidence interval and one ± twice the mean standard deviation.

For each waveband, a maximum of 12 points could be scored. Considering the varying numbers of wavebands returned by the six processors, the score was standardized to the sum of points by dividing over the number of wavebands. The final score for each processor was then obtained from the individual scores divided by the average score of all processors. The Monte Carlo method (Robert and Casella, 2013) was used over 1000 repetitions, with the size of each re-sampled subset 0.75 times the size of the original dataset, to reduce the sensitivity of the scores to the size of the matched dataset available with each processor. Monte Carlo resampling resulted in a confidence range of the score for each AC processor.

3. Results

3.1. Offset correction for the shipborne observations

Fig. 2 presents the *in situ* R_{rs} spectra observed from the AERONET-OC and shipborne platforms. The two data sources may differ slightly since the AERONET-OC observations are taken from a stationary tower where the sensors are located some ~25 m from the sea surface with a field of view of 1º. The shipborne observations were located ~7 m from the sea surface with the field of view of 7º. The spectra measured from the shipborne platform are given in Fig. 2H and $R_{rs}(865)$ was > 0.0020 sr^{-1} for many spectra. These values are significantly higher than those reported in Ficek *et al*. (2011) and in most cases greater than the ranges obtained from the MERIS atmospheric correction processors for $R_{rs}(865)$ (Figs. 3–8). Spectra in panels A-G have been offset-corrected by subtracting the average R_{rs} at 850–900 nm from each reflectance spectrum.

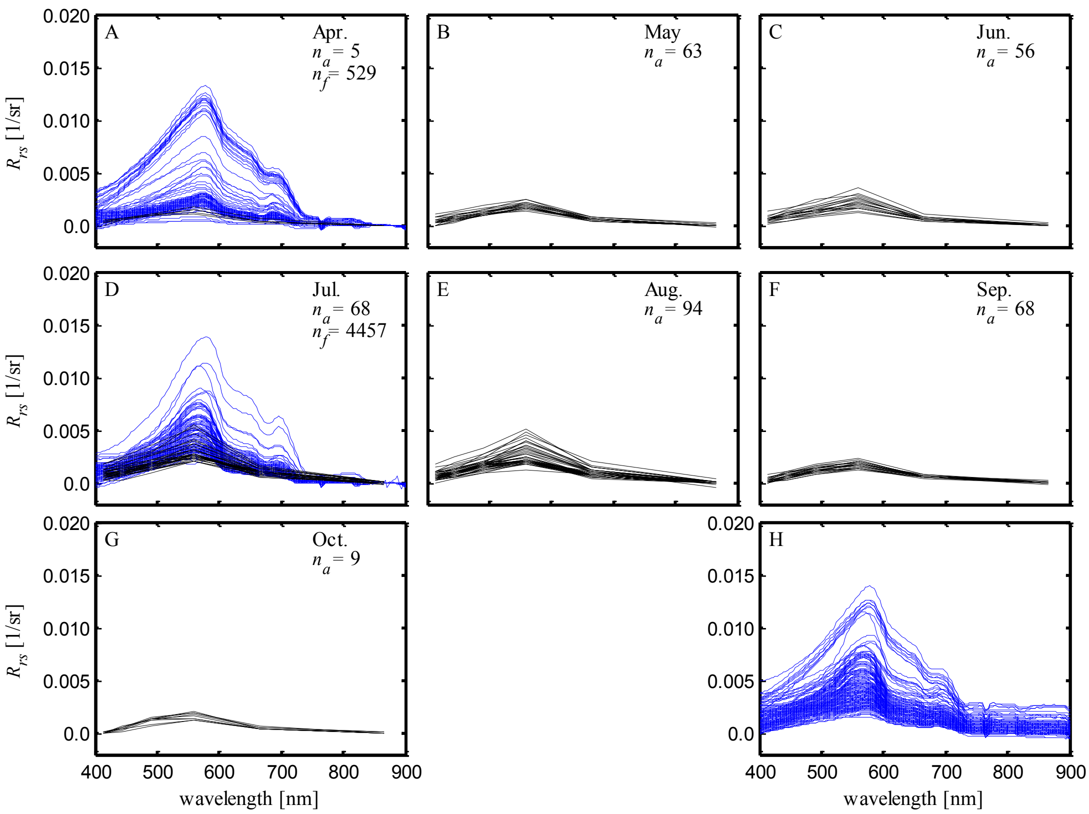

Fig. 2．Spectra of remote-sensing reflectance $R_{rs}(\lambda)$ observed from the ship (blue curves) and AERONET-OC (black curves), separated by month in panels A-G. Panel H shows the shipborne R_{rs} spectra before applying a near infra-red offset correction. The number of observations are marked n_a and n_f for the AERONET-OC and shipborne R_{rs} (following quality control), respectively.

Basic match-up statistics between MERIS-derived $R_{rs}(\lambda)$ and both offset-corrected and uncorrected shipborne observations are given in Table 2. For the match-ups between MERIS and shipborne observations, Δ of the non-offset corrected R_{rs} at selected MERIS wave bands varied from 0.0009 sr^{-1} to 0.0078 sr^{-1} for CC, while the range was 0.0004–0.0011 sr^{-1} for other processors. Using the offset corrected data, the difference was greater for CC (Δ = 0.0011–0.0080 sr^{-1}), but lower for all other processors (Δ = 0.0002–0.0011 sr^{-1}). The determination coefficients R^2 increased and the correlation was improved for each processor following the offset correction. From hereon, the *in situ* shipborne R_{rs} are reported exclusively using the offset correction. We note that the use of a spectrally neutral offset correction is suitable in combination with the fingerprint method used to calculate shipborne $R_{rs}(\lambda)$, which is discussed further in section 4.1.

Table 2

The root mean square difference (Δ, units sr^{-1}) and the coefficient of determination (R^2) between R_{rs} derived from MERIS and *in situ* shipborne observations using a match-up time window of ±3h, using no offset and offset correction.

	Δ (R^2)

No offset correction	443 nm	490 nm	560 nm	665 nm	709 nm
CC	0.0054 (0.00)	0.0068 (0.55)	0.0078 (0.63)	0.0018 (0.01)	0.0009 (0.01)
C2R-Lakes	0.0005 (0.40)	0.0006 (0.62)	0.0006 (0.88)	0.0004 (0.58)	0.0004 (0.40)
C2R-CC	0.0007 (0.20)	0.0007 (0.57)	0.0009 (0.79)	0.0004 (0.56)	0.0004 (0.44)
FUB	0.0006 (0.40)	0.0006 (0.79)	0.0011 (0.86)	0.0008 (0.35)	0.0006 (0.01)
MEGS	0.0011 (0.25)	0.0009 (0.61)	0.0008 (0.83)	0.0006 (0.44)	0.0005 (0.36)
POLYMER	0.0007 (0.44)	0.0005 (0.79)	0.0007 (0.88)	0.0004 (0.50)	0.0006 (0.31)
Offset correction	443 nm	490 nm	560 nm	665 nm	709 nm
CC	0.0056 (0.30)	0.0070 (0.47)	0.0080 (0.53)	0.0021 (0.64)	0.0011 (0.73)
C2R-Lakes	0.0005 (0.40)	0.0007 (0.65)	0.0005 (0.91)	0.0003 (0.78)	0.0002 (0.66)
C2R-CC	0.0008 (0.23)	0.0008 (0.62)	0.0008 (0.80)	0.0003 (0.71)	0.0003 (0.63)
FUB	0.0004 (0.51)	0.0003 (0.87)	0.0008 (0.88)	0.0004 (0.76)	0.0003 (0.61)
MEGS	0.0011 (0.22)	0.0009 (0.62)	0.0008 (0.85)	0.0004 (0.62)	0.0004 (0.62)
POLYMER	0.0009 (0.54)	0.0006 (0.88)	0.0006 (0.91)	0.0003 (0.80)	0.0003 (0.67)

3.2. *In situ* R_{rs}

The monthly AERONET-OC spectral reflectance over the visible and near-infrared domains exhibited a high degree of similarity (Fig. 2 A–G). The hyperspectral R_{rs} collected from the shipborne measurements covered a wider R_{rs} range, but were largely restricted to observations in April and July when the research cruises primarily targeted the period of highest chlorophyll *a*.

A dominant peak in the reflectance between 500 to 600 nm is seen in all AERONET-OC spectra. The amplitude of reflectance was consistently low with the maxima < 0.006 sr^{-1} around 550 nm and the minima approaching zero in the blue waveband at 412 nm, indicating high absorption by CDOM. Monthly average R_{rs}(550) changed seasonally with the lower values of 0.0017 sr^{-1} in May and September, and with the higher values of 0.0029 sr^{-1} in July and August.

Table 3

Statistical results of R_{rs}(560) between MERIS and shipborne observations for six match-up time windows of ± 0.5 h, ± 2 h, ± 3 h, ± 4 h, ± 6 h and ± 12 h, including the number of match-ups (N), the determination coefficient (R^2), the average absolute percentage difference (ψ), the root mean square difference (Δ) and the bias (δ).

		Time window					
		± 0.5 h	± 2 h	± 3 h	± 4 h	± 6 h	± 12 h
CC	N	4	26	40	40	52	59
	R^2	1.00	0.56	0.53	0.53	0.13	0.08
	ψ (%)	240.95	215.56	246.10	246.10	219.57	211.65
	Δ	0.0077	0.0073	0.0080	0.0080	0.0076	0.0074

	δ	2.41	2.13	2.45	2.45	2.14	2.07
C2R-lakes	N	81	245	420	490	544	602
	R^2	0.91	0.92	0.91	0.87	0.86	0.85
	ψ (%)	8.85	9.77	9.03	9.70	10.46	10.44
	Δ	0.0004	0.0005	0.0005	0.0006	0.0006	0.0006
	δ	0.05	0.07	0.05	0.05	0.05	0.04
C2R-CC	N	86	265	453	534	578	644
	R^2	0.91	0.87	0.80	0.75	0.70	0.69
	ψ (%)	9.79	14.07	14.51	16.49	16.56	17.39
	Δ	0.0004	0.0007	0.0008	0.0008	0.0009	0.0009
	δ	-0.03	-0.07	-0.03	-0.04	-0.02	-0.03
FUB	N	43	149	221	255	255	256
	R^2	0.61	0.85	0.88	0.87	0.87	0.86
	ψ (%)	22.36	21.36	18.20	18.40	18.41	18.56
	Δ	0.0009	0.0008	0.0008	0.0009	0.0009	0.0009
	δ	-0.22	-0.17	-0.14	-0.15	-0.15	-0.15
MEGS	N	74	201	306	364	377	427
	R^2	0.92	0.88	0.85	0.77	0.76	0.76
	ψ (%)	7.11	12.28	12.05	12.89	13.09	14.79
	Δ	0.0004	0.0007	0.0008	0.0009	0.0009	0.0009
	δ	0.02	-0.03	-0.01	-0.02	-0.02	-0.05
POLYMER	N	95	281	453	518	573	644
	R^2	0.98	0.93	0.91	0.87	0.85	0.83
	ψ (%)	5.47	9.19	8.63	9.10	9.48	11.51
	Δ	0.0002	0.0005	0.0006	0.0006	0.0007	0.0008
	δ	-0.03	-0.05	-0.03	-0.03	-0.03	-0.06

The shipborne hyperspectral R_{rs} observation showed a similar spectral shape, with the green peak located near 580 nm, with a maximum < 0.015 sr^{-1}. A local minimum at 660 nm and maximum at 680 nm were consistently observed in the shipborne hyperspectral R_{rs}, corresponding to the absorption of Chl *a* in the red waveband (675 nm) and sun-induced fluorescence of Chl *a*, respectively. Spectra in July and August also had the highest absorption at red wavebands when Chl *a* concentrations reached up to 15 mg m^{-3} (Simis & Olsson, 2013).

3.3. *Match-up time window of the shipborne observations*

We analyzed various time windows (±12 h, ±6 h, ±4 h, ±3 h, ±2 h and ±0.5 h) between the shipborne data and MERIS over-pass to assess the effect on the match-up results, which are given in Table 3 for R_{rs}(560).

Compared to the ±12-h window, the number of match-ups decreased to 90% for the ±6-h window, 83 %

for the ±4-h window, 73% for the ±3-h window, 46% for the ±2-h window and 14% for the ±0.5-h window. The ψ values of $R_{rs}(560)$ by POLYMER ranged from 5.5% (±0.5-h window) up to 11.5% (±12-h window) and Δ were from 0.0002 sr^{-1} to 0.0008 sr^{-1}. Analogous results were observed for C2R-Lakes, C2R-CC and MEGS. The number of match-ups was lower for these processors and MERIS $R_{rs}(560)$ showed a lower difference with the *in situ* R_{rs} when using the shorter time windows. For the shorter match-up windows, the coefficient of determination improved for most processors except for CC and FUB, while the bias varied slightly for all processors. $R_{rs}(560)$, irrespective of AC processor, had the lowest deviation when using the ±3-h match-up window. Similar performance was observed for the other wavebands. The time window of ±3 h was selected to report further results, providing the best balance between match-up volume and statistical match-up performance.

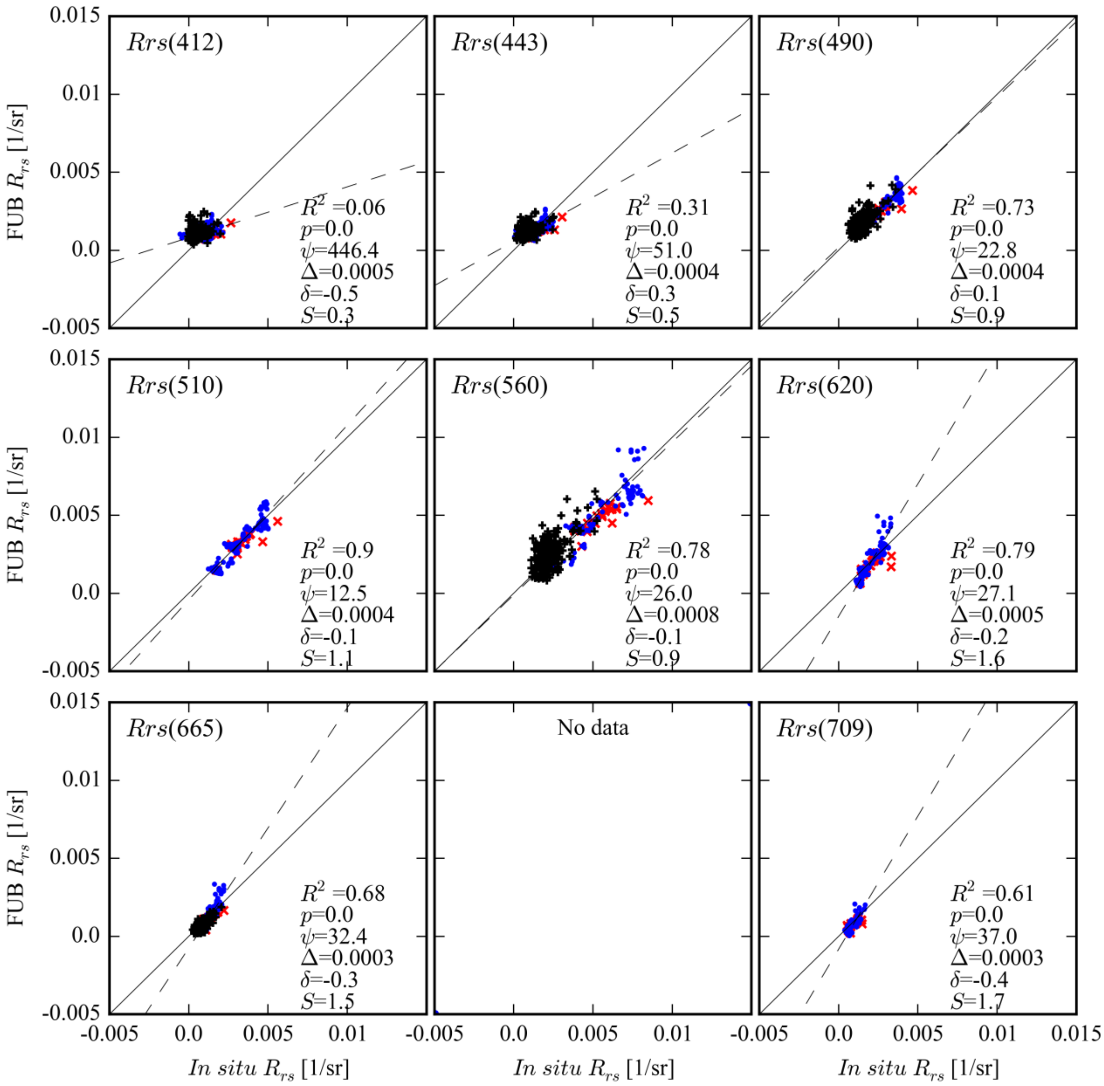

Fig. 3. Scatter plots of MERIS R_{rs} retrieved by FUB versus *in situ* R_{rs}, for MERIS bands as indicated at the top of each panel. The number of observations are n_a = 176 for the AERONET-OC and n_f = 221 for shipborne R_{rs}. Blue points represent match-ups with shipborne data, red crosses are shipborne observations where R_{rs} was negative in the near infra-red (before offset correction), and black plusses are match-ups with AERONET-OC. The solid line represents unity and the dashed line is the best fit of Type-2 linear least-squares regression through the combined data sets. R^2 is the coefficient of determination, p is the probability level of significance, ψ is the average absolute percentage difference, Δ is the root mean square difference and δ is the bias between MERIS and *in situ* match-ups, S is the slope of the Type-2 linear regression.

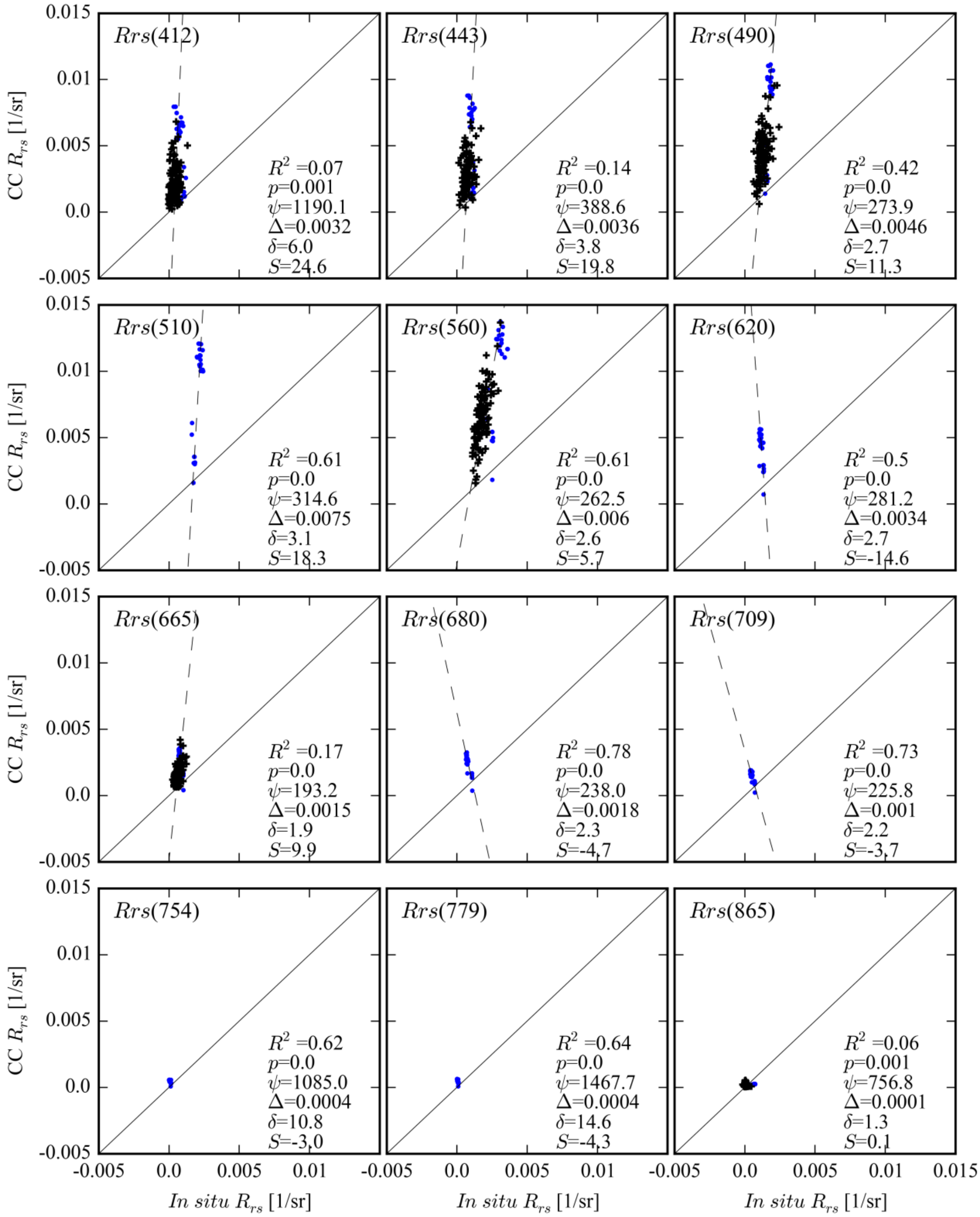

Fig. 4. Scatter plots of R_{rs} retrieved by CC versus *in situ* R_{rs}. The number of observations are n_a = 110 for the AERONET-OC and n_f

= 40 for shipborne R_{rs}. Markers and symbols are as described for Fig. 3.

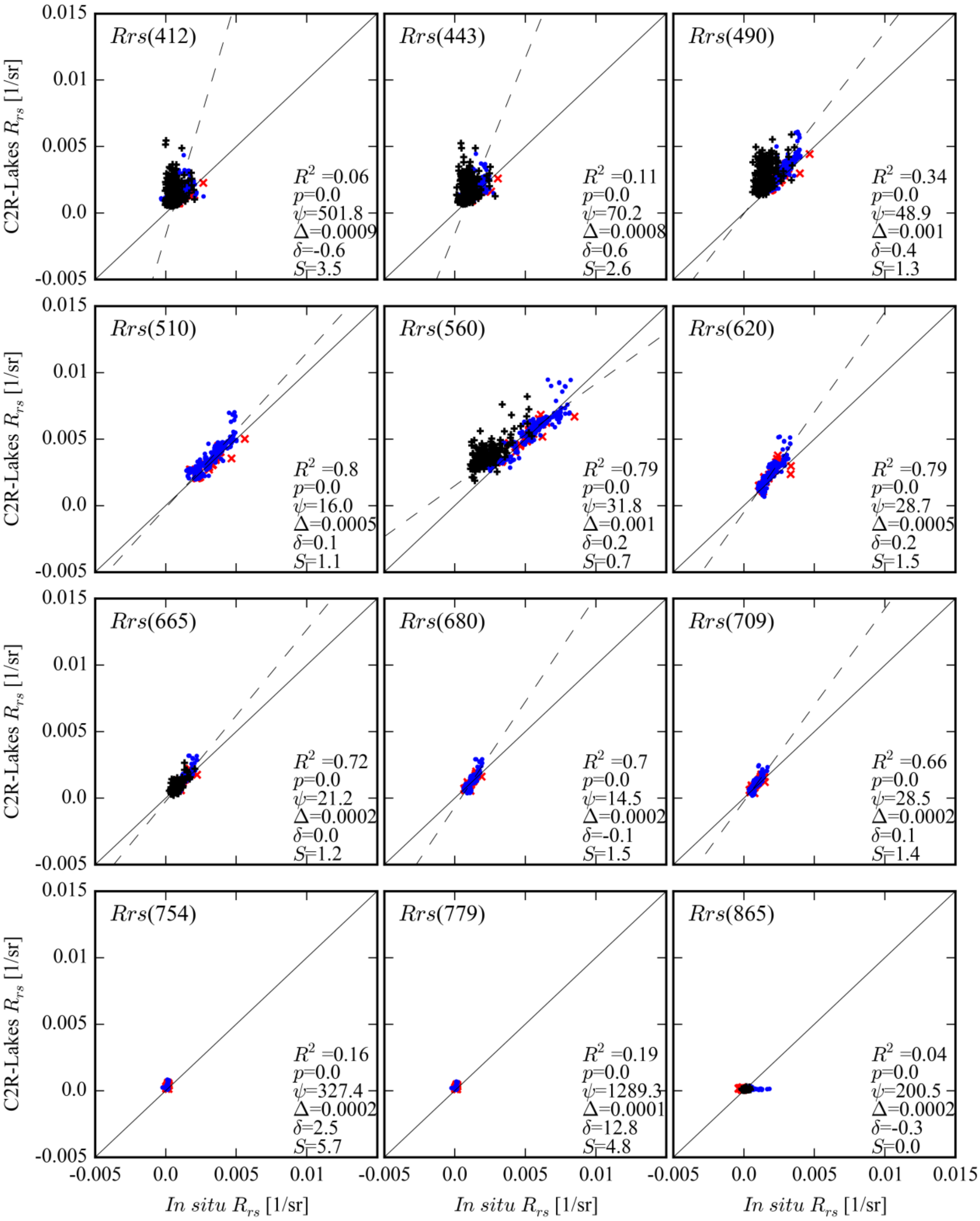

Fig. 5. Scatter plots of R_{rs} retrieved by C2R-Lakes versus *in situ* R_{rs}. The number of observations are n_a = 213 for the AERONET-OC and n_f = 420 for shipborne R_{rs}. Markers and symbols are as described for Fig. 3.

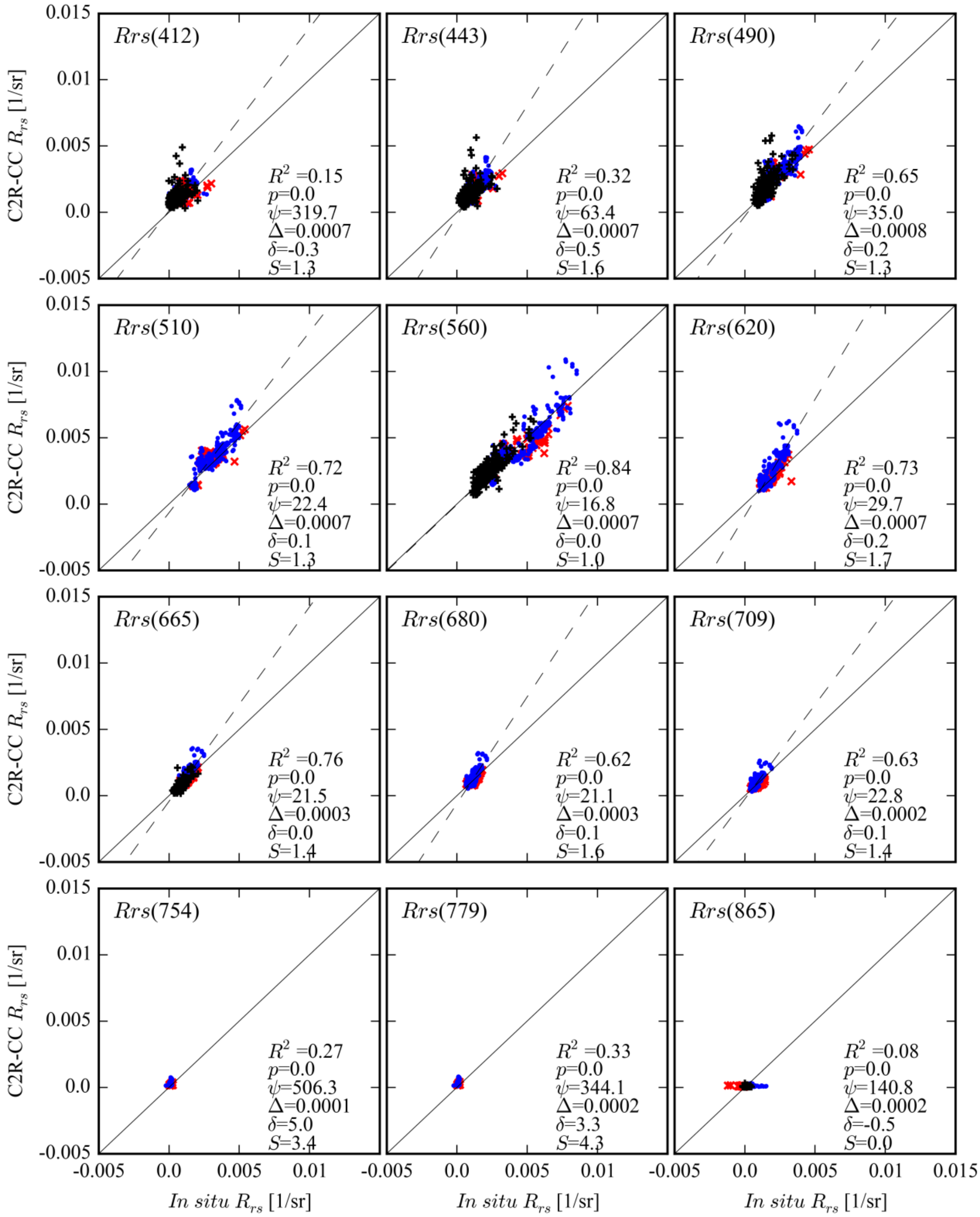

Fig. 6. Scatter plots of R_{rs} retrieved by C2R-CC versus *in situ* R_{rs}. The number of observations are n_a = 214 for the AERONET-OC and n_f = 453 for shipborne R_{rs}. Markers and symbols are as described for Fig. 3.

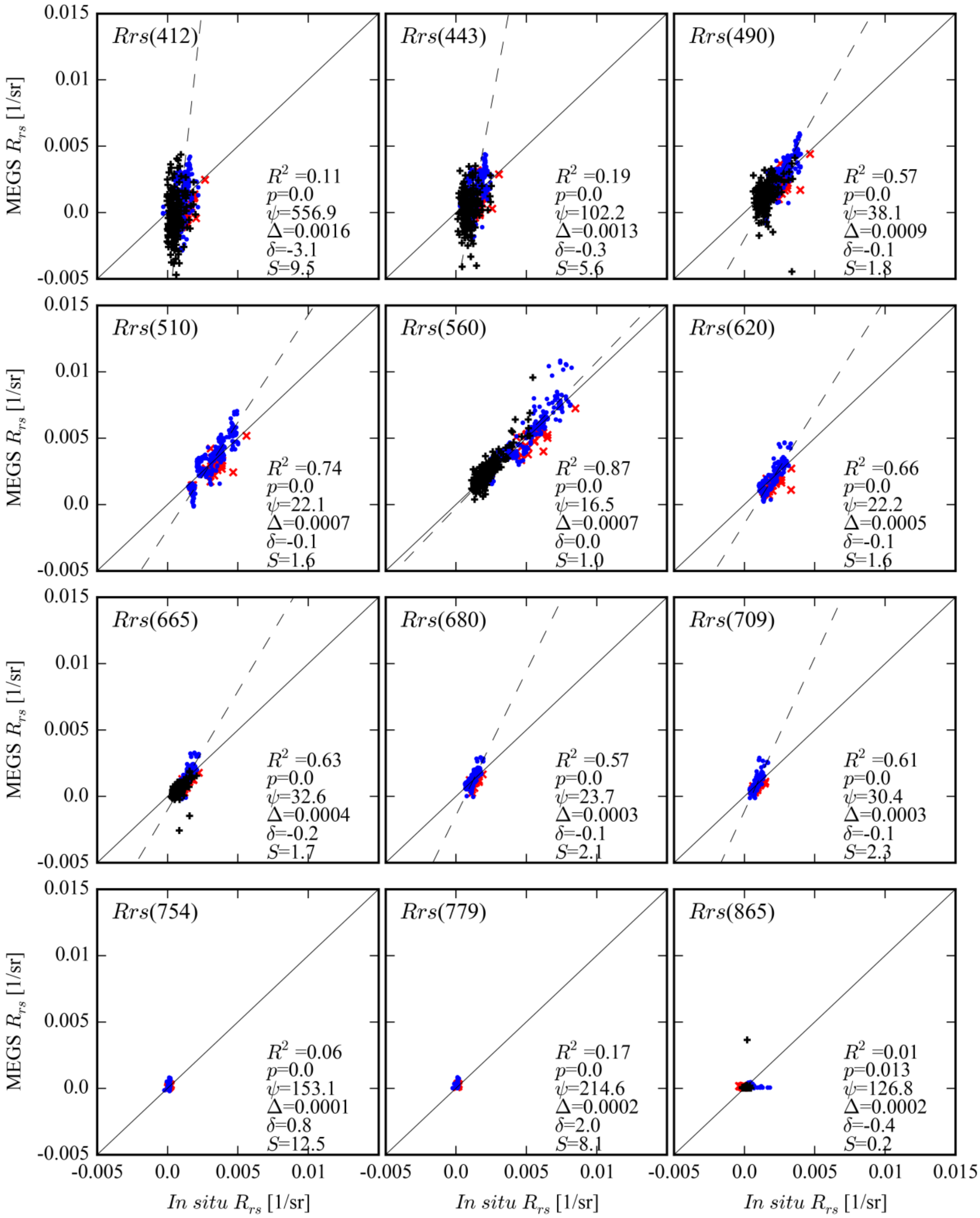

Fig. 7. Scatter plots of R_{rs} retrieved by MEGS versus *in situ* R_{rs}. The number of observations are n_a = 187 for the AERONET-OC and n_f = 306 for shipborne R_{rs}. Markers and symbols are as described for Fig. 3.

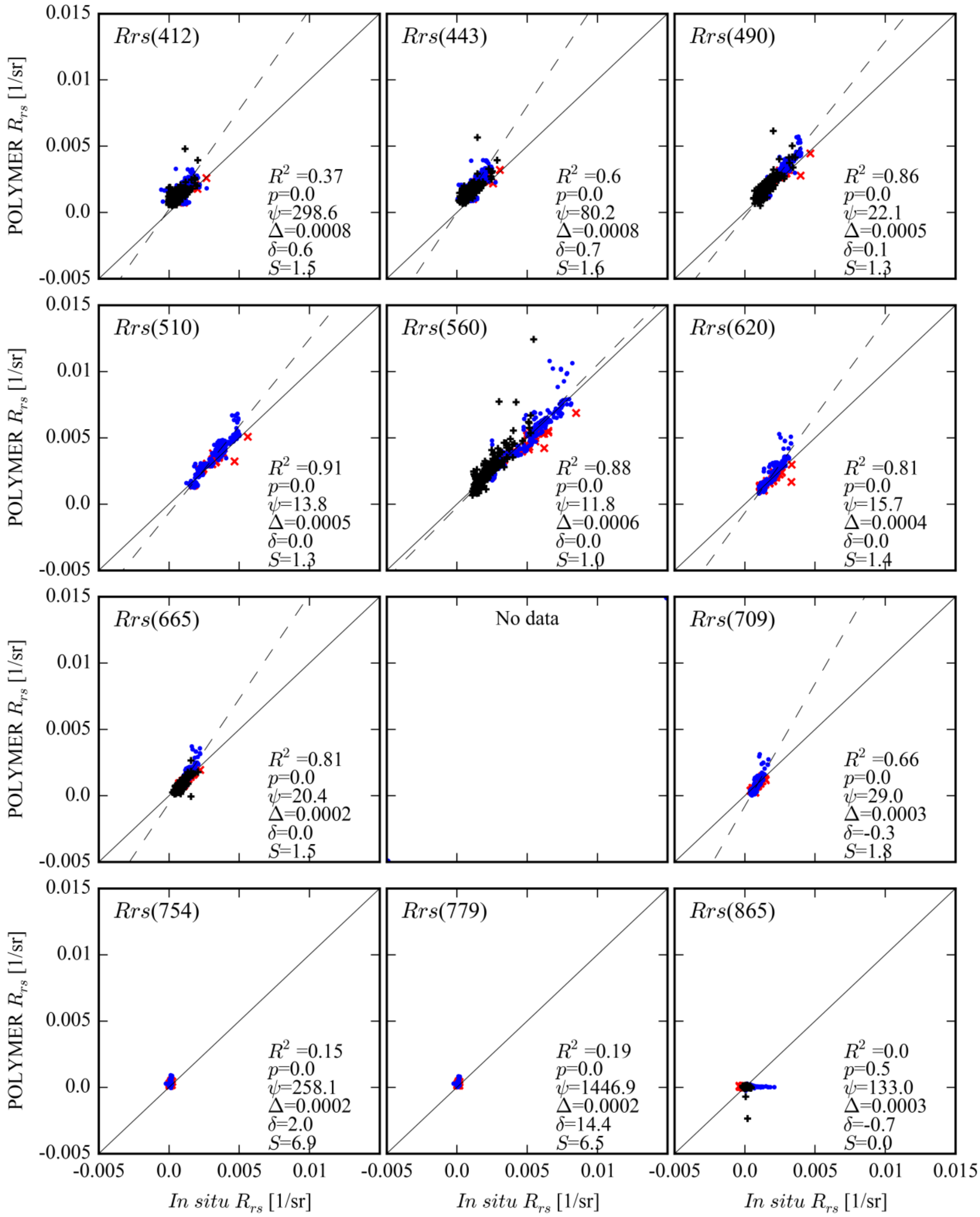

Fig. 8. Scatter plots of R_{rs} retrieved by POLYMER versus *in situ* R_{rs}. The number of observations are n_a = 211 for the AERONET-OC and n_f = 453 for shipborne R_{rs}. Markers and symbols are as described for Fig. 3.

3.4. Accuracy assessment of AC processors

Firstly, all valid match-up observations within ±3-h for shipborne data and ±2-h for AERONET-OC were considered for each of the processors. All six AC processors showed a low correlation at 412 nm ($R^2 < 0.37$), significant probability of the regression fit ($p < 0.001$) and large deviations ($\psi > 200\%$) with *in situ* R_{rs}

match-ups (Figs. 3–8). It is noted that the range in *in situ* R_{rs} at blue bands was small, which influences these regression results. For C2R-Lakes, C2R-CC, FUB, MEGS and POLYMER $R_{rs}(443)$, there was a slightly higher correlation (R^2 ranging from 0.11 to 0.60) and lower differences (ψ ranging from 51% to 102%) compared to *in situ* $R_{rs}(443)$, except for CC. The highest relative differences of all processors were observed in the near infrared at 754 and 779 nm with $\psi > 150\%$. For the other visible wavebands (490–709 nm), the performance of all processors improved, especially FUB, C2R-Lakes, C2R-CC, MEGS and POLYMER (Figs. 3, 5, 6, 7 and 8). The FUB processor performed well at 490–709 nm with $R^2 > 0.61$ and low ψ of 12–37% (Fig. 3). Compared to all *in situ* $R_{rs}(\lambda)$, the CC processor over-estimated R_{rs} with a high positive bias ($\delta > 1.9$), which resulted in the highest differences ($\psi > 190\%$) in visible wavebands (Fig. 4). The C2R-Lakes processor showed good agreement with *in situ* $R_{rs}(\lambda)$ for most bands with $\psi < 30$ % and $R^2 = 0.66$–0.80, but exhibited high differences at 490 nm ($\psi = 49\%$) and 560 nm ($\psi = 32\%$). C2R-CC also performed well and had low ψ at < 35% in bands 490–709 nm and a moderate coefficient of determination (R^2 ranging from 0.62 to 0.84). MEGS had a low correlation at most wavebands ($R^2 < 0.74$) and similar deviations with $\psi = 22$–38%, except for $R_{rs}(560)$ with a better performance ($\psi = 16\%$ and $R^2 = 0.87$). POLYMER was the most accurate processor with lowest ψ (12% to 22%) and highest consistency ($R^2 > 0.81$), except for $R_{rs}(709)$ with lower accuracy ($\psi =$ 29%, $R^2 = 0.66$).

Table 4
The numbers of observations shared between any two AC processors.

	CC	FUB	C2R-lakes	C2R-CC	MEGS	POLYMER
CC	150					
FUB	101	397				
C2R-lakes	150	355	633			
C2R-CC	150	396	622	667		
MEGS	118	336	495	494	495	
POLYMER	150	397	632	664	495	664

There was a large variation in the number of valid match-ups between processors with 150 for CC, 633 for C2R-Lakes, 667 for C2R-CC, 397 for FUB, 495 for MEGS and 664 for POLYMER. We therefore also compared performance over the set of match-ups shared by the processors to reduce the effect of processor-specific quality flags. Table 4 gives an overview of the number of observations shared between any two AC processors within the ±3-h window. CC and FUB had the lowest number of valid observations, which indicates that these two processors were often operating out of their scope and may not be applicable to the

Baltic Sea. When CC and FUB are not considered, the shared subset of match-ups for C2R-Lakes, C2R-CC, MEGS and POLYMER was 494 and the statistical results at R_{rs} 490, 560, 620, 665 and 709 nm is given in Table 5.

Table 5

Statistical results of R_{rs} match-ups based on 494 shared observations within a time window of ±3 h, including the coefficient of determination (R^2), the average absolute percentage difference (ψ), the root mean square difference (Δ), and bias (δ), slope (S) and intercept (I) of type-2 linear regression between MERIS and *in situ* match-ups.

processor	λ (nm)	R^2	ψ (%)	Δ (sr^{-1})	δ	S	I (sr^{-1})
C2R-Lakes	490	0.35	53.75	0.0012	0.52	1.21	0.0004
	510	0.78	17.52	0.0006	0.15	1.09	0.0001
	560	0.81	35.25	0.0011	0.33	0.75	0.0017
	620	0.78	28.33	0.0006	0.20	1.64	-0.0008
	665	0.71	22.99	0.0003	0.08	1.31	-0.0003
	680	0.66	16.34	0.0003	-0.02	1.95	-0.0012
	709	0.62	27.67	0.0003	0.12	1.71	-0.0005
C2R-CC	490	0.73	31.08	0.0008	0.23	1.39	-0.0004
	510	0.82	17.78	0.0007	0.09	1.47	-0.0012
	560	0.85	17.06	0.0008	0.00	1.03	-0.0002
	620	0.73	27.76	0.0008	0.23	2.08	-0.0016
	665	0.78	20.89	0.0003	0.03	1.56	-0.0006
	680	0.64	20.42	0.0004	0.07	2.11	-0.0013
	709	0.62	21.39	0.0003	0.10	1.81	-0.0006
MEGS	490	0.57	38.19	0.0009	-0.08	1.87	-0.0020
	510	0.74	22.15	0.0008	0.00	1.62	-0.0019
	560	0.87	16.54	0.0007	0.05	1.10	-0.0003
	620	0.66	22.33	0.0005	-0.08	1.69	-0.0014
	665	0.63	32.66	0.0004	-0.19	1.79	-0.0010
	680	0.57	23.80	0.0004	-0.01	2.18	-0.0015
	709	0.62	30.39	0.0004	-0.03	2.33	-0.0011
POLYMER	490	0.87	22.05	0.0006	0.17	1.31	-0.0003
	510	0.90	13.44	0.0005	0.07	1.34	-0.0008
	560	0.87	12.89	0.0007	0.02	1.08	-0.0002
	620	0.79	15.19	0.0005	0.06	1.58	-0.0010
	665	0.83	21.07	0.0003	0.00	1.60	-0.0006
	709	0.65	26.95	0.0003	-0.15	2.10	-0.0010

For this data set, C2R-Lakes tended to overestimate R_{rs} from 510–709 nm where δ varied from -0.02 to 0.33 and ψ was < 35%. R_{rs}(490) showed higher deviation with ψ = 54% and R^2 = 0.35 (Table 5). C2R-CC tended to overestimate R_{rs} with δ between 0.00 and 0.23 at 490 to 709 nm, with ψ < 31%. MEGS underestimated R_{rs} especially at 620 to 709 nm, with a moderate correlation (R^2 = 0.57–0.74) and ψ varying from 22% to 38%, except for $R_{rs}(560)$ where R^2 = 0.87 and ψ = 17%. The highest correlation with *in situ* R_{rs} was for POLYMER which gave $R^2 > 0.65$ and $\psi < 27\%$ at these wavebands.

3.5. Accuracy assessment of band ratios

Band ratios $R_{rs}(443)/R_{rs}(560)$, $R_{rs}(490)/R_{rs}(560)$ and $R_{rs}(709)/R_{rs}(665)$ are commonly used to relate the shape of reflectance to biogeochemical properties, notably phytoplankton absorption signals in the blue/green and near infrared/red part of the spectrum. The band ratios for each processor were evaluated against *in situ* band ratios (Fig. 9), to assess the potential for retrieving accurate spectral shapes and phytoplankton biomass in these CDOM rich waters. Owing to limited spectral variability in the dataset, the band ratios from all processors had relatively low correlations ($R^2 < 0.36$) with the *in situ* observations. ψ varied from 5.7% at $R_{rs}(709)/R_{rs}(665)$ by C2R-CC to 94% at $R_{rs}(443)/R_{rs}(560)$ using MEGS. $R_{rs}(490)/R_{rs}(560)$ had a relatively stable accuracy compared to other band ratios with ψ of 18–31% and Δ of 0.12–0.32 for all AC processors. $R_{rs}(490)/R_{rs}(560)$ retrieved by POLYMER had better agreement with the *in situ* values with ψ = 18% and Δ = 0.12. Compared with $R_{rs}(443)/R_{rs}(560)$ and $R_{rs}(490)/R_{rs}(560)$, the retrieval accuracy for $R_{rs}(709)/R_{rs}(665)$ was better with low ψ and Δ of 10.2% and 0.08 for CC, 13.8% and 0.1 for C2R-Lakes, 5.7% and 0.05 for C2R-CC, 17.7% and 0.14 for FUB, and 12.5% and 0.11 for MEGS. This suggests that the best retrieval of spectral shape occurs in the red to NIR domain.

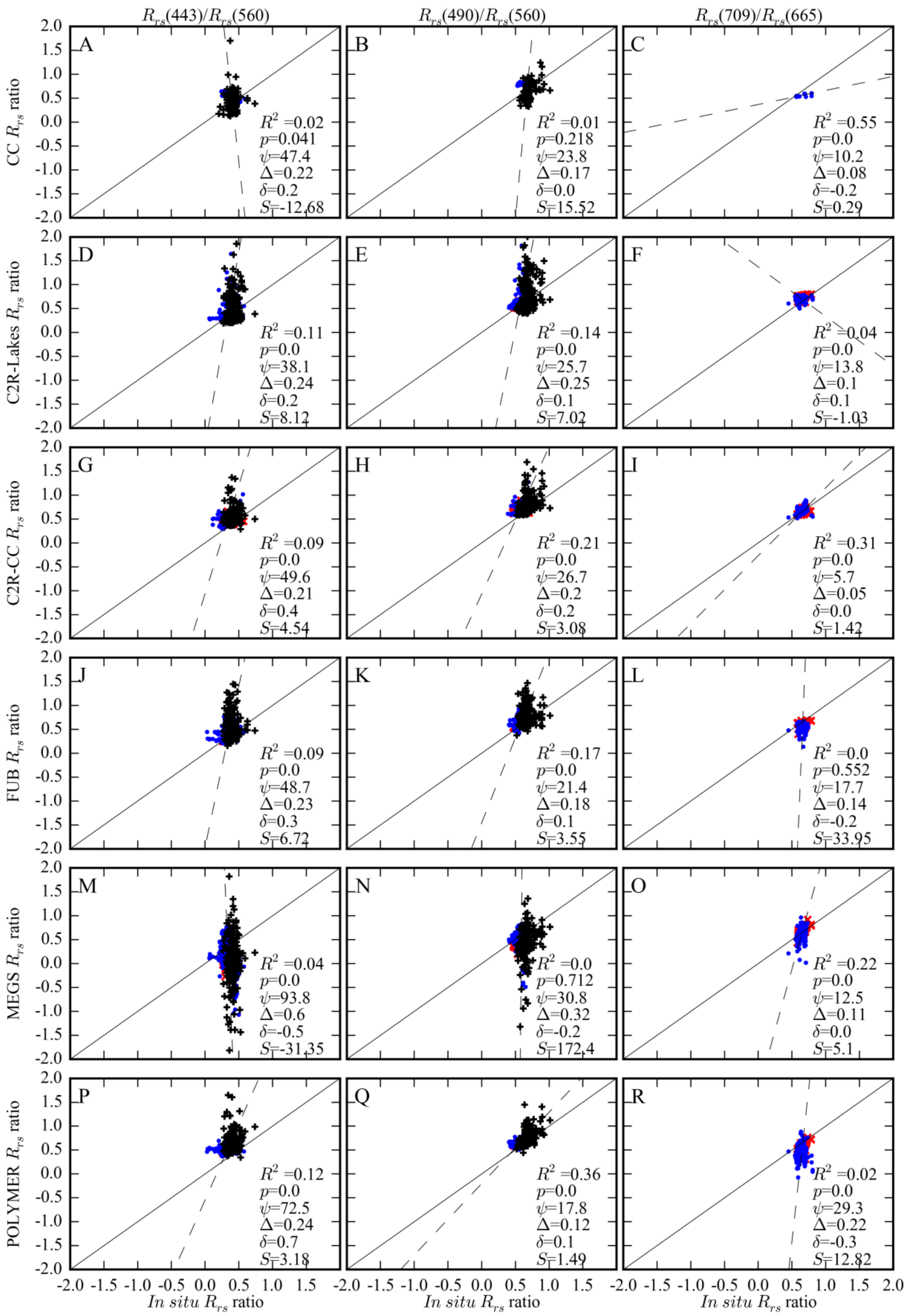

Fig. 9. Scatter plots of band ratio between MERIS-derived and *in situ* R_{rs}. Markers and symbols are as described for Fig. 3.

3.6. Statistical ranking of the accuracy of AC processors

Based on the statistical metrics given in Figs. 3–8 and Table 5 for all match-up data, the ranked scores of all processors is given in Fig. 10A and the subset of match-ups shared between C2R-Lakes, C2R-CC, MEGS and POLYMER is given in Fig. 10B.

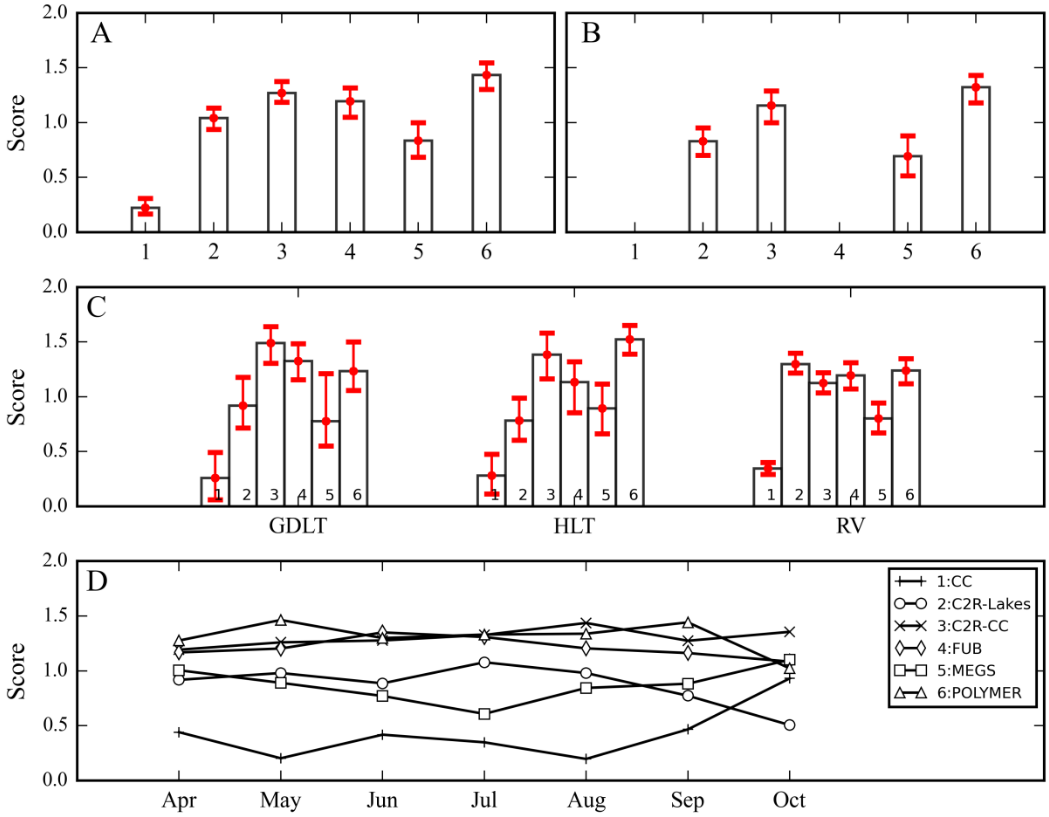

Fig. 10. Scores assigned to the R_{rs} retrieval performance of each processor (1. CC; 2. C2R-Lakes; 3. C2R-CC; 4. FUB; 5. MEGS; 6. POLYMER). (A) Scores when including all data available for each processor. (B) Scores obtained with observations shared between C2R-Lakes, C2R-CC, MEGS, and POLYMER. (C) Scores (all data and processors included) separated by data source (GDLT = Gustaf Dalen Lighthouse Tower, HLT = Helsinki Lighthouse Tower, RV = Research Vessel). (D) Scores separated by month. Error bars in panels A-C are the 2.5% and 97.5% confidence interval of the scores (see text).

For each processor, POLYMER showed the highest score of 1.43 and a 95% confidence interval of 1.31 to 1.55. C2R-CC and FUB had the next highest scores (1.18–1.37 and 1.05–1.32, respectively), and the overlapping error bars between them indicated statistical similarity (Fig. 10A). CC had the lowest score (~ 0.22), indicating that it was the least accurate processor.

For shared observations, the performance of C2R-Lakes, C2R-CC, MEGS and POLYMER was similar to those for all processors using all match-ups. POLYMER still obtained the highest score (~ 1.32 with a 1.19–1.44 at 95 % confidence interval), followed by C2R-CC (~ 1.15; 1.00–1.29), and MEGS with a score of 0.69.

Further comparisons of these ranked scores to account for differences between methods (Shipborne vs

AERONET_OC), locations (coastal AERONET-OC and open Baltic Sea) and months are given in Fig. 10C & D. For the GDLT, C2R-CC had the highest score (~ 1.49; 1.31–1.65 at 95 % confidence; Fig. 10C), followed by FUB and POLYMER with the average scores of 1.33 and 1.23, respectively. For the HLT, the highest score was obtained for POLYMER (~ 1.52), followed by C2R-CC (~ 1.38). For the shipborne observations, C2R-Lakes and POLYMER had similar mean scores (~ 1.27; 1.12–1.40 at 95% confidence). C2R-CC and FUB exhibited slightly lower scores of about 1.15, and CC had consistently the lowest score (~ 0.35).

The monthly scores of each processor are shown in Fig. 10 D based on the match-ups between MERIS and AERONET-OC observations. The ranges of the average scores separated by month are 0.19–0.94 for CC, 0.50–1.08 for C2R-Lakes, 1.09–1.35 for FUB, 0.61–1.11 for MEGS, 1.02–1.46 for POLYMER and 1.19–1.43 for C2R-CC. The scores of POLYMER, C2R-CC and FUB were consistently > 1.0, indicating better than average performance throughout the seasons. C2R-Lakes scored highest in July (~1.07), and CC always scored lowest.

4. Discussion

4.1. Offset correction for shipborne observations

Shipborne R_{rs} in 750–900 nm bands was high compared to previous observations in the Baltic Sea (Ficek *et al*, 2011). The sources of these differences were investigated. The NIR spectrum is largely determined by the absorption of pure water except in optically turbid waters, in which case NIR reflectance ratios approach constant values, a phenomenon known as the 'NIR similarity spectrum' (Ruddick *et al,* 2006). The shipborne observations (before offset correction) showed a linear regression between $R_{rs}(779)$ and $R_{rs}(865)$ of $R_{rs}(779)=1.0054R_{rs}(865)+0.0001$ with a high correlation ($R^2=0.99$; Fig. 11) and slope near unity. For the turbid waters of the North Sea it has been reported that this value should approach 1.82 (Ruddick *et al*, 2006). This suggests that the NIR signal of the Baltic Sea shipborne observations does not represent significant particle scattering as no discernable variation due to the absorption characteristics of pure water are observed. The high reflectance from 750–900 nm in the Baltic Sea is therefore likely caused by residual effects of surface contamination effects from waves, ship movement, spray, or whitecaps. It may be assumed that this effect is spectrally neutral because the fingerprint method (Simis and Olsson 2013) already accounts for the combined effect of diffuse and specular reflection at the water surface. Sun glint effects are likely minor due to the use of a sun-tracking platform measuring water-leaving radiance at an azimuth angle close to 135° from the solar azimuth. Nevertheless, due to the low amplitude of water-leaving radiance in the highly absorbing waters of the Baltic Sea, the residual offset can become significant with respect to the amplitude of reflectance. It is likely that a similar correction is not needed in more turbid water bodies. Since the AERONET-OC observations are made from a fixed platform, they are less prone to sea spray, tilt and roll that can affect the shipborne observations, hence there is no or little residual offset in these data. Following offset correction, the shipborne reflectance spectra and AERONET-OC observation produced a continuous pattern compared to the MERIS-derived reflectance bands.

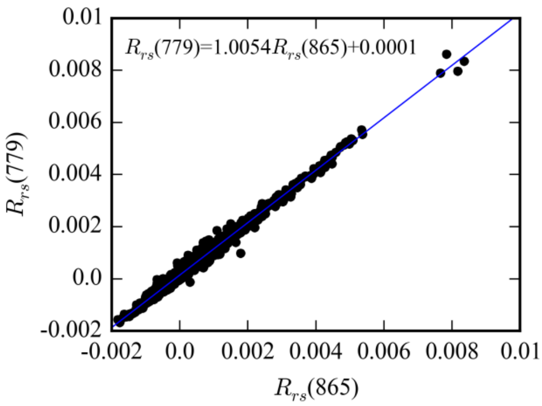

Fig. 11. Relationship between $R_{rs}(779)$ and $R_{rs}(865)$ in shipborne observations.

4.2. Match-up time window

The time window between *in situ* collection and satellite overpass was a compromise between reducing the effects of temporal variability in the *in situ* data and obtaining a large volume of match-up observations. It has been recommended to restrict match-up windows to ±3 h in Case 1 waters and no more than ±0.5 h in Case 2 waters (Bailey & Werdell, 2006). However, the movement of water masses, rate of vertical mixing, and the motility of phytoplankton ultimately determine how fast optical conditions change. Tidal currents in the Baltic Sea are slight due to limited connectivity with the Atlantic Ocean in the Baltic. The movement of water masses in the Baltic Sea resembles a quasi-enclosed estuary supplied with fresh water from river runoff. The basins are normally well-mixed within the visible surface layer, except during some phytoplankton bloom periods (Drozdowska, 2007). Comparison results within various time windows (± 0.5–12 h) between the shipborne observations and MERIS over-pass suggest that a ±3-h window yielded a useful number of match-ups and close to optimal statistical match-up performance.

4.3. Accuracy assessment of AC processors

Our results showed a variable number of match-ups between the six AC processors. For the neural network based processors, the numbers were dependent on the range of the training datasets. CC developed for Case 2 waters had the lowest number of match-up pairs, indicated that more pixels retrieved by CC were out of the training range and the range of CC was not available to the Baltic Sea. C2R-CC, in contrast, showed the largest number of match-ups due to the increased range in the training dataset of the neural network. POLYMER also obtained a higher number of match-ups as it applies less stringent flagging of the processor output.

The radiometric validation results illustrated that the six AC processors had the lowest accuracy at shorter wavebands (412 and 443 nm). The accuracies improved from 490 to 560 nm, but the deviations increased again at longer wavebands (> 709 nm), corresponding to varying amplitude of Baltic Sea reflectance between these bands, which are similar to previous studies (Beltrán-Abaunza *et al.*, 2014; Attila *et al.* 2013; Zibordi *et al*., 2009a; Zibordi *et al*., 2013). Based on the AERONET-OC data collected at the HLT and GDLT stations, Zibordi *et al*. (2013) found that MERIS L_{WN} by MEGS at the 490, 560 and 665 nm bands had lower deviation ($\psi < 24\%$) and moderate correlation ($R^2 > 0.39$) than the blue bands (412 and 443 nm). Beltrán-Abaunza *et al*. (2014) used the in-water radiometer to compare the MERIS $\rho_w(\lambda)$ obtained by the MEGS, C2R and FUB processors on the Northern Baltic Proper. Better consistency with *in situ* observations was found at 560 nm with the correlation coefficient of 0.91 for MEGS, 0.87 for C2R and 0.84 for FUB, and the worst consistency was at 412 nm for these three processors. The relatively weak R_{rs} at blue bands (412 and 443 nm) is characteristic of the optical properties of highly absorbing waters. The contribution of the reflectance at the sea surface to the top-of-atmospheric radiance is therefore low, which amplifies the errors at these wavebands. This resulted in the poor performance to retrieve R_{rs} in the blue wavebands.

The combined validation results assigned the POLYMER processor the highest overall score, better correlation, lowest deviations and highest number of match-ups compared against all other processors. This indicated that POLYMER was the most accurate processor applied to MERIS for the Baltic Sea. Owing to the flexibility of this model, POLYMER exhibited the smallest deviation and highest score in the Case 1 and Case 2 waters compared to MEGS, SeaDAS and Forward NN (Müller *et al*, 2015). POLYMER also showed the best performance and highest score in the CDOM dominated waters of the Baltic Sea, throughout the observation period. Even so, the accuracy of retrieval at blue wavelengths was worse than at longer wavelengths for POLYMER and this still requires improvement. Possible reasons for this were that the absorption of CDOM was neglected or expressed as the Chl *a* concentration in the bio-optical model. In the Baltic Sea CDOM does not co-vary with Chl *a* and significantly affects the blue to green range of the spectra.

Among the four neural network AC processors (CC, C2R-Lakes, C2R-CC and FUB), C2R-CC showed the best performance and CC the worst, which is likely to be due to the training data sets used to calibrate the neural network. This calibration also includes the effects of different aerosol types, cirrus clouds, sun and sky radiance, and the coupling between them and the air molecules. The atmospheric masses in the Baltic Sea are affected by both land and marine due to its geographical position. The average aerosol optical thickness was about 1.3 as determined at the island of Gotland in the central part of the basin (Carlund *et al*., 2005). The higher values of the aerosol optical thickness over the Baltic Sea in April may be related to the burning of agricultural waste straw in northern Europe and Russia (Zdun *et al*, 2011). The standard AC used in CC was not suited to this region, which resulted in the worst performance of all the processors tested. The mixture of maritime and continental aerosol models may account for the improved accuracy of FUB and C2R-Lakes. The coastal aerosol model used in C2R-CC is appropriate for the Baltic Sea. The maximum CDOM absorption used to generate the simulated reflectance in the training databases was 1 m^{-1} at 443nm for CC, C2R-Lakes, C2R-CC and FUB, which was sufficient for most areas of the Baltic Sea except for areas near large rivers in the north and east which are not close to the AERONET-OC or shipborne stations.

The performance of MEGS 8.1 was poor in the Baltic Sea, most likely because it was primarily designed for open ocean waters dominated by phytoplankton, but it uses the bright pixel (BP) AC in highly scattering waters. In the Baltic Sea however, the BPAC is rarely triggered and only the open ocean AC model is used in this region. The constant for $R_{rs}(510)$ was obtained from the Case 1 waters, and likely resulted in larger derivations from the actual aerosol and path radiance when used in high-CDOM absorption waters of the Baltic Sea. An over-correction of the atmospheric signal resulted in the bias (δ) being less than zero at blue and green wavebands (Fig. 7).

The use of such a comprehensive data set for the Baltic Seahas wider implications for other similar high CDOM waters and for the new generation of Copernicus Sentinels, which additional have short wave infra-red (SWIR) bands that can potentially improve the performance of AC models (Wang et al. 2007). The estuaries of the Northern most parts of the Gulf of Bothnia, in Finland and Sweden, and the Eastern most part of the Gulf of Finland are the highest absorbing CDOM waters in the region (Kowalczuk et al., 1999; Ylöstalo et al., 2016), but were not covered by the shipborne observations.

4.4 Implications for use of AC processors with band ratio algorithms

The retrieval of biogeochemical components, such as the Chl *a* concentration, from satellite sensors depends on the availability of suitable algorithms, as well as the performance of atmospheric correction to accurately retrieve both the amplitude and shape of R_{rs} at the sea surface from the TOA radiances. Band ratio algorithms are common in optically complex and productive waters, and can reduce systematic retrieval error caused by atmospheric corrections when the aerosols are not absorbing, i.e. when the error affects the bands used in the band ratio in equal measure. Low correlation coefficients between satellite and *in situ* reflectance band ratios appear to have been caused by a highly conserved shape of the R_{rs} spectrum in the Baltic Sea resulting in a narrow range of band ratio values.

For all six atmospheric correction processors, the bias between MERIS and *in situ* observations at blue wavebands was larger than at blue-green bands, which resulted in poor retrieval of $R_{rs}(443)/R_{rs}(560)$ ratios (Fig. 9) suggesting that these band ratios are not suitable to retrieve biogeochemical products in these waters. For POLYMER, the MERIS-retrieved $R_{rs}(490)/R_{rs}(560)$ had the best agreement with the *in situ* data. The retrieval of $R_{rs}(709)/R_{rs}(665)$ ratios improved for some processors, such as C2R-CC, C2R-Lakes, CC, FUB and MEGS which is relevant for retrieving Chl *a* in highly absorbing waters when the use of blue-green ratios can be erroneous.

For all processors, the blue-green ratio of $R_{rs}(443)/R_{rs}(560)$ exhibitdd the worst performance with the lowest R^2 (< 0.11) and largest ψ (38.1–72.5%). R_{rs} is low in the blue region due to the high absorption by CDOM, and the performance of $R_{rs}(490)/R_{rs}(560)$ ratios were better compared to $R_{rs}(443)/R_{rs}(560)$ ratios since the $R_{rs}(490)$ signal was stronger than $R_{rs}(443)$, which may be relevant for Chl *a* algorithms such as OC3 and OC4 when they use the $R_{rs}(490)/R_{rs}(560)$ ratios. Pitarch *et al*. (2016), however used the regional calibration of OC4v6 to map the Chl *a* concentration in the Baltic Sea, but they found that OC4v6 over-estimates Chl *a* resulting in a $R^2 = 0.43$ and bias of 0.44, suggesting that Chl *a* algorithms for the Baltic Sea, should use longer wavelengths than $R_{rs}(490)$. In their analysis, they also included data from the Kattegat and Skagerrak which proved to be more accurate with blue : green Chl *a* algorithms than for the Baltic Sea area. Considering that 600 nm was the waveband for minimum particle absorption and that pigment absorption dominated the total absorption at wavelengths longer than 510 nm, Darecki *et al*. (2003) shifted the wavelengths from $R_{rs}(490)/R_{rs}(550)$ to $R_{rs}(550)/R_{rs}(590)$ in empirical Chl *a* algorithm. Better results were obtained with $R^2 = 0.75$ and $\psi = 20\%$. Based solely on our observations of the radiometric retrieval accuracy of AC models, other Chl *a* algorithms, including NIR-red ratio algorithms and possibly algorithms based on fluorescence line height, could improve the accuracy of Chl *a* retrieval..

The band ratio $R_{rs}(709)/R_{rs}(665)$ showed the highest accuracy for C2R-CC and C2R-Lakes. Ligi *et al* (2016) recently showed, that the NIR-Red model $R_{rs}(709)/R_{rs}(665)$ is most suitable for Chl *a* concentration based on a large dataset of simulated $R_{rs}(\lambda)$ and field measurements in the Baltic Sea. The wavelength region from 620 nm to 709 nm provides essential features for Chl *a* estimation, as well as absorption diagnostic of cyanobacteria pigments at 620 nm, smaller interference of CDOM absorption, and the light scattering peak near 709 nm where absorption of water constituents is small with respect to absorption by water. Accurate retrieval of R_{rs} in the NIR-red region in general and the $R_{rs}(709)/R_{rs}(665)$ ratio in particular should therefore be considered a priority in further AC and in-water algorithm validation. Currently, three AC processors (POLYMER, C2R-CC and C2R-Lakes) exhibit promising results in this spectral domain.

The 709 nm band as well as retrieval further into the NIR also plays an essential role in the detection of surface accumulation of phytoplankton, such as cyanobacteria blooms in the Baltic Sea during calm weather in summer (Groetsch et al. 2014). During the field campaigns only small surface blooms were encountered and few co-occured during clear-sky satellite passes, so we can only focus on the systematic R_{rs} retrieval performance of the various AC schemes during relatively well mixed conditions. Reflectance retrieval over patchy, sub-pixel sized surface blooms is an enormous challenge both from the perspective of satellite AC and *in situ* data collection. Neither the AERONET (due to its limited band set) nor the shipborne (disturbance of the water mass) platforms are well suited to perform this matchup analysis. Spectra characteristic of surface blooms were therefore not included in this analysis.

5 Conclusions

The performance of six AC processors (CC, C2R-Lakes, C2R-CC, FUB, MEGS, and POLYMER) for MERIS was assessed in the Baltic Sea, against *in situ* remote sensing reflectance from AERONET-OC and shipborne measurements. All six processors showed poor performances in the blue (412 and 443 nm) and NIR wavebands (754–865 nm), but better performances at 490 to 709 nm except for CC. The CC processor exhibited the worst accuracy with $\psi > 190\%$ for all wavebands. POLYMER exhibited the best performance at MERIS bands from 490–709 nm and had the lowest deviations (ψ = 12–29%) and bias (δ = -0.3–0.1) and the highest correlation (R^2 = 0.66–0.91) when compared to the *in situ* data. C2R-CC was the second most accurate algorithm. The retrieval of $R_{rs}(709)/R_{rs}(665)$ was supported by all processors, suggesting that accurate Chl *a* concentrations for the Baltic Sea are feasible. Further improvement in POLYMER and C2R-CC at blue and NIR bands, which are both still under development, would improve their applicability for highly absorbing waters such as the Baltic Sea.

This analysis represents the largest data set used to date to test a range of AC models for the highly absorbing waters of the Baltic Sea, and is therefore relevant and applicable to other highly absorbing water bodies such as the Arctic Ocean, The Yellow Sea, the Black Sea, the River mouths of the Amazon and a large range of freshwater lakes, where ocean colour products still prove to be erroneous.

Acknowledgements

SS and GT were funded by the European Space Agency SEOM contract Extreme Case 2 waters (C2X) (4000113691/15/I-LG). PQ was funded by a scholarship from the Chinese Scholarship Program. We thank the crew of RV *Aranda* and colleagues from the Finnish Environment Institute (SYKE) for the data collected during research cruises. We are grateful to Giuseppe Zibordi from the Joint Research Center for AERONET-OC data from the Gustaf Dalen and Helsinki Lighthouse Tower and for valuable comments on a draft of the manuscript.

References

Aiken, J., &Moore, G. (2000). Case 2 (S) bright pixel atmospheric correction. MERIS ATBD, 2,6-6.

Antoine, D., & Morel, A. (2011). Atmospheric correction of the MERIS observations over ocean Case 1 waters. Tech. rep., MERIS ATBD 2.7, issue 5, revision 5. http://envisat.esa.int/instruments/meris/atbd /atbd_2.7.pdf.

Attila, J., Koponen, S., Kallio, K., *et al.* (2013). MERIS Case II water processor comparison on coastal sites of the northern Baltic Sea. Remote sensing of environment, 128, 138-149.

Aznay, O., & Santer, R. (2009). MERIS atmospheric correction over coastal waters: validation of the MERIS aerosol models using AERONET. International Journal of Remote Sensing, 30(18), 4663-4684.

Bailey, S. W., & Werdell, P. J. (2006). A multi-sensor approach for the on-orbit validation of ocean color satellite data products. Remote Sensing of Environment, 102(1), 12-23.

Beltrán-Abaunza, J. M., Kratzer, S., & Brockmann, C. (2014). Evaluation of MERIS products from Baltic Sea coastal waters rich in CDOM. Ocean Science, 10(3), 377-396.

Brewin, R. J., Sathyendranath, S., Müller, D., *et al.* (2015). The Ocean Colour Climate Change Initiative: III. A round-robin comparison on in-water bio-optical algorithms. Remote Sensing of Environment, 162, 271-294.

Brockmann, C., Doerffer, R., Peters, M., *et al*. (2016). Evolution of the C2R-CC neural network for Sentinel 1 and 3 for the retrieval of ocean colour products in normal and extreme optically complex waters, In Proceedings ESA Living Planet Symposium Prague, 09-13, May.

Callieco, F., & Dell'Acqua, F. (2011). A comparison between two radiative transfer models for atmospheric correction over a wide range of wavelengths. International Journal of Remote Sensing, 32(5), 1357-1370.

Carlund, T., Håkansson, B., & Land, P. (2005). Aerosol optical depth over the Baltic Sea derived from AERONET and SeaWiFS measurements. International Journal of Remote Sensing, 26(2), 233-245.

D'Alimonte, D., Zibordi, G., Berthon, J-F., Canuti, E., Kajiyama, T. (2012) Performance and applicability of bio-optical algorithms in different European seas. Remote Sensing of Environment, 124: 402-412.

Darecki, M., Weeks, A., Sagan, S., Kowalczuk, P., & Kaczmarek, S. (2003). Optical characteristics of two contrasting Case 2 waters and their influence on remote sensing algorithms. Continental Shelf Research, 23(3), 237-250.

Darecki, M., & Stramski, D. (2004). An evaluation of MODIS and SeaWiFS bio-optical algorithms in the Baltic Sea. Remote Sensing of Environment, 89(3), 326-350.

Doerffer, R., & Schiller, H. (2007). The MERIS Case 2 water algorithm. International Journal of Remote Sensing, 28(3-4), 517-535.

Doerffer, R., Schiller, H. (2008). MERIS regional, coastal and lake case 2 water project - atmospheric correction ATBD. GKSS Research Center, Geesthacht, Germany, version 1.0, 18 May 2008.

Doron, M., Babin, M., Hembise, O., et al. (2011) Ocean transparency from space: Validation of algorithms using MERIS, MODIS and SeaWiFS data. Remote Sensing of Environment, 115: 2986-3001.

Drozdowska, V. (2007). Seasonal and spatial variability of surface seawater fluorescence properties in the Baltic and Nordic Seas: results of lidar experiments. Oceanologia, 49(1), 59-69.

Ficek, D., Zapadka, T., & Dera, J. (2011). Remote sensing reflectance of Pomeranian lakes and the Baltic. Oceanologia, 53(4), 959-970.

Glover, D.M., Jenkins, W. J., Doney, S.C. (2011). Modeling methods for marine science: Cambridge University Press.

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.

Groetsch, P. M., Simis, S. G., Eleveld, M. A., & Peters, S. W. (2014). Cyanobacterial bloom detection based on coherence between ferrybox observations. Journal of Marine Systems, 140, 50-58.

Harvey, E. T., Kratzer, S., Philipson, P. (2015) Satellite-based water quality monitoring for improved spatial and temporal retrieval of chlorophyll-a in coastal waters. Remote Sensing of Environment, 158: 417-430.

He, X., Bai, Y., Pan, D., Tang, J., & Wang, D. (2012). Atmospheric correction of satellite ocean color imagery using the ultraviolet wavelength for highly turbid waters. Optics express, 20(18), 20754-20770.

Hooker, S. B., Lazin, G., Zibordi, G., & McLean, S. (2002). An evaluation of above-and in-water methods for determining water-leaving radiances. Journal of Atmospheric and Oceanic Technology, 19(4), 486-515.

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.

IOCCG (2000), Remote sensing of ocean colour in coastal, and other optically-complex, waters, In S. Sathyendranath (Ed.), Technical report: Reports of the International Ocean-Colour Coordinating Group, No. 10, Dartmouth, Canada: IOCCG.

IOCCG (2010). Atmospheric Correction for Remotely-Sensed Ocean-Colour Products. In M. Wang (Ed.), Technical report: Reports of the International Ocean-Colour Coordinating Group, No. 10, Dartmouth, Canada: IOCCG.

Kahru, M., Elmgren, R., & Savchuk, O. P. (2015). Changing seasonality of the Baltic Sea. Biogeosciences Discussions, 12(22), 18855-18882.

Knaeps, E., Dogliotti, A. I., Raymaekers, D., Ruddick, K., & Sterckx, S. (2012). *In situ* evidence of non-zero reflectance in the OLCI 1020nm band for a turbid estuary. Remote Sensing of Environment, 120, 133-144.Kratzer, S., Brockmann, C., Moore, G. (2008) Using MERIS full resolution data to monitor coastal waters - A case study from Himmerfjarden, a fjord-like bay in the north western Baltic Sea. Remote Sensing of Environment, 112: 2284-2300.

Koponen, S., Attila, J., Pulliainen, J., Kallio, K., Pyhälahti, T., Lindfors, A., Rasmus, K., & Hallikainen, M. (2007). A case study of airborne and satellite remote sensing of a spring bloom event in the Gulf of Finland. Continental Shelf Research, 27(2), 228-244.

Kowalczuk, P. (1999). Seasonal variability of yellow substance absorption in the surface layer of the Baltic Sea. Journal of Geophysical Research-Oceans, 104, C12, 30047-30058.

Krawczyk, H., Neumann, A., Walzel, T., Hetscher, M., Siegel, H. (1997). Application of multispectral interpretation algorithm to remote sensing data over the Baltic Sea. Ocean Optics XIII, Proceedings of the Society of Photo-optical Instrumentation Engineers (SPIE). 2963: 234-239.

Lenoble, J., Herman, M., Deuzé, J. L., Lafrance, B., Santer, R., & Tanré, D. (2007). A successive order of scattering code for solving the vector equation of transfer in the earth's atmosphere with aerosols. Journal of Quantitative Spectroscopy and Radiative Transfer, 107(3), 479-507.

Leppäranta, M., & Myrberg, K. (2009). Physical oceanography of the Baltic Sea. Springer Science & Business Media.

Ligi, M., Kutser, T., Kallio, K., *et al*. (2016). Testing the performance of empirical remote sensing algorithms in the Baltic Sea waters with modelled and *in situ* reflectance data. Oceanologia.

Matthews, M.W. (2011). A current review of empirical procedures of remote sensing in inland and near coastal transitional waters. International Journal of Remote Sensing, 32(21), 6855-6899.

Melin, F., Zibordi, G., Berthon, J-F. (2007) Assessment of satellite ocean color products at a coastal site. Remote Sensing of Environment, 110: 192-215.

Moore, G. F., & Lavender, S. (2011). MERIS ATBD 2.6. Case II. S Bright Pixel Atmospheric Correction. URL. https://earth.esa.int/documents/700255/2042855/MERIS_ATBD_2.6_v5.0+-+2011.pdf

Mobley, C. D. (1999). Estimation of the remote-sensing reflectance from above-surface measurements. Applied Optics, 38(36), 7442-7455.

Mobley, C. D. (2015). Polarized reflectance and transmittance properties of windblown sea surfaces. Applied optics, 54(15), 4828-4849.

Müller, D., Krasemann, H., Brewin, R. J., *et al.* (2015). The Ocean Colour Climate Change Initiative: I. A methodology for assessing atmospheric correction processors based on in-situ measurements. Remote Sensing of Environment, 162, 242-256.

Nobileau, D., & Antoine, D. (2005). Detection of blue-absorbing aerosols using near infrared and visible (ocean color) remote sensing observations. Remote Sensing of Environment, 95(3), 368-387.Ohde, T; Sturm, B; Siegel, H. (2002) Derivation of SeaWiFS vicarious calibration coefficients using in situ measurements in Case 2 water of the Baltic Sea. Remote Sensing of Environment, 80: 248-255.

Omstedt, A., Elken, J., Lehmann, A., & Piechura, J. (2004). Knowledge of the Baltic Sea physics gained during the BALTEX and related programmes. Progress in Oceanography, 63(1), 1-28.

Pierson, D. C., Kratzer, S., Strombeck, N. (2008) Relationship between the attenuation of downwelling irradiance at 490 nm with the attenuation of PAR (400 nm-700 nm) in the Baltic Sea. Remote Sensing of Environment, 112: 668-680.

Pitarch, J., Volpe, G., Colella, S., Krasemann, H., & Santoleri, R. (2016). Remote sensing of chlorophyll in the Baltic Sea at basin scale from 1997 to 2012 using merged multi-sensor data. Ocean Science, 12(2), 379-389.

Reinart, A., Kutser, T. (2006) Comparison of different satellite sensors in detecting cyanobacterial bloom events in the Baltic Sea. Remote Sensing of Environment, 102: 74-85.

Robert, C., & Casella, G. (2013). Monte Carlo statistical methods. Springer Science & Business Media.

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.

Ruddick, K. G., De Cauwer, V., Park, Y. J., & Moore, G. (2006). Seaborne measurements of near infrared water-leaving reflectance: The similarity spectrum for turbid waters, Limnol. Oceanogr., 51(2), 1167–1179.

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., Schaale, M., & Fischer, J. (2007). Retrieval of atmospheric and oceanic properties from MERIS measurements: A new Case 2 water processor for BEAM. International Journal of Remote Sensing, 28(24), 5627-5632.

Siegel, D. A., Wang, M., Maritorena, S., & Robinson, W. (2000). Atmospheric correction of satellite ocean color imagery: the black pixel assumption. Applied optics, 39(21), 3582-3591.

Simis, S. G. H., & Olsson, J. (2013). Unattended processing of shipborne hyperspectral reflectance measurements. Remote Sensing of Environment, 135, 202-212.

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

Stramska, M., Swirgon, M. (2014) Influence of atmospheric forcing and freshwater discharge on interannual variability of the vertical diffuse attenuation coefficient at 490 nm in the Baltic Sea. Remote Sensing of Environment, 40: 155-164.

Thuillier, G., Hersé, M., Foujols, T., *et al.* (2003). The solar spectral irradiance from 200 to 2400 nm as measured by the SOLSPEC spectrometer from the ATLAS and EURECA missions. Solar Physics, 214(1), 1-22.Voss, K. J., Morel, A., Antoine, D. (2007) Detailed validation of the bidirectional effect in various Case 1 waters for application to ocean color imagery. Biogeosciences, 4, 781-789.

Wang, M., & Shi, W. (2007). The NIR-SWIR combined atmospheric correction approach for MODIS ocean color data processing. Optics Express, 15(24), 15722-15733.

Woźniak, M., Bradtke, K. M., & Krężel, A. (2014). Comparison of satellite chlorophyll a algorithms for the Baltic Sea. Journal of Applied Remote Sensing, 8(1), 083605.

Ylöstalo, P., Seppälä, J., Kaitala, S., Maunula, P., & Simis, S. (2016). Loadings of dissolved organic matter and nutrients from the Neva River into the Gulf of Finland–Biogeochemical composition and spatial distribution within the salinity gradient. Marine Chemistry, 186, 58-71.

Zibordi, G., Berthon, J. F., Mélin, F., D'Alimonte, D., & Kaitala, S. (2009a). Validation of satellite ocean color primary products at optically complex coastal sites: Northern Adriatic Sea, Northern Baltic Proper and Gulf of Finland. Remote Sensing of Environment, 113(12), 2574-2591.

Zibordi, G., Mélin, F., Berthon, J.-F., *et al.* (2009b). AERONET-OC: A Network for the Validation of Ocean Color Primary Products. Journal of Atmospheric and Oceanic Technology, 26(8), 1634-1651.

Zibordi, G., Mélin, F., Berthon, J. F., & Canuti, E. (2013). Assessment of MERIS ocean color data products for European seas. Ocean Science, 9(3), 521-533.

Zdun, A., Rozwadowska, A., & Kratzer, S. (2011). Seasonal variability in the optical properties of Baltic aerosols. Oceanologia, 53(1), 7-34.