Hyperspectral remote sensing of shallow waters: considering environmental noise and bottom intra-class variability for modeling and inversion of water reflectance

Sylvain Jay[a,*], Mireille Guillaume[a], Audrey Minghelli[b], Yannick Deville[c], Malik Chami[d,e], Bruno Lafrance[f], Véronique Serfaty[g]

[a] *Aix Marseille Univ, CNRS, Centrale Marseille, Institut Fresnel, F-13013 Marseille, France*
[b] *University of Toulon, CNRS, SeaTech, LSIS laboratory, UMR 7296, 83041 Toulon, France*
[c] *Institut de Recherche en Astrophysique et Planétologie (IRAP), Observatoire Midi-Pyrénées, Université de Toulouse, UPS-CNRS-OMP, 31400 Toulouse, France*
[d] *Sorbonne Universités, UPMC Univ Paris 06, INSU-CNRS, Laboratoire Atmosphères Milieux Observations Spatiales (LATMOS), 06230 Villefranche sur Mer, France*
[e] *Institut Universitaire de France, 75231 Paris Cedex 05, France*
[f] *CS Systemes d'Information, 31506 Toulouse Cedex 05, France*
[g] *DGA/DS/MRIS, 75509 Paris Cedex 15, France*

Abstract

Hyperspectral remote sensing is now an established tool to determine shallow water properties over large areas, usually by inverting a semi-analytical model of water reflectance. However, various sources of error may make the observed subsurface remote-sensing reflectance deviate from the model, resulting in an increased retrieval error when inverting the model based on classical least-squares fitting. In this paper, we propose a probabilistic forward model of shallow water reflectance variability that describes two of the main sources of error, namely, (1) the environmental noise that includes every source of above-water variability (e.g., sensor noise and rough water surface), and (2) the potentially complex inherent spectral variability of each benthic class through their associated spectral covariance matrix. Based on this probabilistic model, we derive two inversion approaches, namely, MILE (MaxImum Likelihood estimation including Environmental noise) and MILEBI (MaxImum Likelihood estimation including Environmental noise and Bottom Intra-class variability) that utilize the information contained in the proposed covariance matrices to further constrain the inversion while allowing the observation to differ from the model in the less reliable wavebands. In this paper, MILE and MILEBI are compared with the widely used least-squares (LS) criterion in terms of depth, water clarity and benthic cover retrievals. For these three approaches, we also assess the influence of constraining bottom mixture coefficients to sum to one on

* *Corresponding author.*

estimation results.
The results show that the proposed probabilistic model is a valuable tool to investigate the influence of bottom intra-class variability on subsurface reflectance, e.g., as a function of optical depth or environmental noise. As expected, this influence is critical in very optically shallow waters, and decreases with increasing optical depth. The inversion results obtained from synthetic and airborne data of Quiberon Peninsula, France, show that MILE and MILEBI generally provide better performances than LS. For example, in the case of airborne data with depth ranging from 0.44 to 12.00 m, the bathymetry estimation error decreases by about 32% when using MILE and MILEBI instead of LS. Estimated maps of bottom cover are also more consistent when derived using sum-to-one constrained versions of MILE and MILEBI. MILE is shown to be a simple but powerful method to map simple benthic habitats with negligible influence of intra-class variability. Alternatively, MILEBI is to be preferred if this variability cannot be neglected, since taking bottom covariance matrices into account concurrently with mean reflectance spectra may help the bottom discrimination, e.g., in the presence of overlapping classes. This study thus shows that taking potential sources of error into account through appropriate paramerizations of spectral covariance may be critical to improve the remote sensing of shallow waters, hence making MILE and MILEBI interesting alternatives to LS.
Keywords: Bottom intra-class variability, Environmental noise, Maximum likelihood estimation, Radiative transfer model inversion, Shallow water hyperspectral remote sensing, Spectral covariance

---

1. Introduction

Optical remote sensing provides an outstanding opportunity to monitor aquatic environments from local to global scales, potentially offering high temporal and spatial resolutions, e.g., as allowed by recent advances in unmanned aerial vehicles or by the Sentinel-2 mission developed by the European Space Agency within the "Copernicus" program (Aschbacher & Milagro-Pérez, 2012; Drusch et al., 2012). The use of such high spatial resolution data (i.e., less than a few dozen meters) is particularly critical for coastal and inland waters, e.g., to map heterogeneous benthic habitats (Mishra et al., 2006; Hedley et al., 2012b), to detect coral bleaching (Andréfouët et al., 2002; Hedley et al., 2012a) or to monitor small lakes and rivers (Joshi & D'Sa, 2015). As compared with the open ocean, coastal and inland waters are generally more complex environments, whose remotely-sensed reflectance may be highly variable due to simultaneous changes in bathymetry, water quality, bottom type, water surface and atmospheric conditions. In shallow waters, the decoupling of these effects has been shown to be more accurate when using hyperspectral data instead of multispectral data (Lee & Carder, 2002; Lee et al., 2013). Indeed, a higher number of spectral bands as well as an increased spectral resolution allow reducing confounding effects between optically-active parameters, e.g., by detecting the subtle changes in reflectance that originate from narrow absorption regions potentially present in bottom albedo (Kutser et al., 2003; Hochberg & Atkinson, 2003; Hedley et al., 2012a; Botha et al., 2013).

In coastal environments, hyperspectral remote sensing methods that allow the simultaneous retrieval of bathymetry, water quality and benthic cover are usually based on a radiative transfer model that describes how light propagates in water (Mobley, 1994). This inverse problem is generally solved using either look-up tables (LUTs) or iterative optimization (Dekker et al., 2011). In the first case, a spectral library corresponding to different combinations of depth, water quality and benthic cover is pre-computed using an exact (Mobley, 1994) or approximated (Lee et al., 1998) radiative transfer model. For each image pixel, the measured reflectance is then matched with the closest simulated spectrum in the LUT. CRISTAL (Comprehensive Reflectance Inversion based on Spectrum matching and TAble Lookup) (Mobley et al., 2005) and ALLUT (Adaptive Linearized Look-Up Trees) (Hedley et al., 2009) as denoted by Dekker et al. (2011) are examples of such approaches. The inverse problem can also be solved by numerically optimizing a cost function that relates measured and simulated reflectance spectra. In this case, the forward model used for simulation has to be sufficiently fast to permit multiple runs for each image pixel. To this end, a number of analytical and semi-analytical models have been developed under various assumptions and water types (Maritorena et al., 1994; Lee et al., 1998; Albert & Mobley, 2003). These models approximate the radiative transfer equation and generally simulate the reflectance of shallow waters as a function of sun-sensor geometry, depth, bottom albedo and water-column inherent optical properties (i.e., absorption and scattering properties of the water column). Note that, whenever possible, the latter can further be related to specific inherent optical properties and concentrations of optically-active water constituents (Brando et al., 2009).

Due to its accurate performance and simplicity, the Euclidean distance has generally been used to assess the goodness-of-fit between the observation and the model, either when using LUTs (Mobley et al., 2005; Hedley et al., 2009, 2012a) or iterative optimization (Lee et al., 1999, 2001; Lee & Carder, 2002; Albert & Gege, 2006; Klonowski et al., 2007; Dekker et al., 2011; Jay et al., 2012; Giardino et al., 2012; Garcia et al., 2014a; McKinna et al., 2015; Jay & Guillaume, 2016). Note that in the case of iterative optimization, the use of Euclidean distance for model inversion corresponds to nonlinear unweighted least-squares fitting. However, this cost function does not fully consider the information contained in the reflectance data. In particular, it does not utilize spectral covariance (i.e., covariance between wavebands), yet such knowledge of the data structure may be useful to improve the retrieval accuracy due to the non-negligible correlation between hyperspectral bands (Gillis et al., 2013).

Importantly, as the least-squares method tries to find the best possible fit between the observation and the model, it is not designed to handle possible deviations between them. For example, the "environmental noise equivalent reflectance difference" (Brando & Dekker, 2003) (hereafter called environmental noise and denoted $\mathrm{NE}\Delta r_E$) may lead the measured subsurface reflectance to strongly differ from the modeled one. For a given spectral band, $\mathrm{NE}\Delta r_E$ corresponds to the reflectance standard deviation as estimated over an "as homogeneous as possible" water area. As a result, it not only takes into account the sensor noise, but also scene-specific above-water variability, including atmospheric variability, effects related to the rough water surface, refractions of diffuse and direct sunlight, and residuals from imperfect atmospheric, air-water interface and sun glint corrections (Brando & Dekker, 2003; Brando et al., 2009; Botha et al., 2013). To consider such errors within model inversion, Brando et al. (2009) and Botha et al. (2013) have weighted the contribution of each waveband according to the inverse of $\mathrm{NE}\Delta r_E$. In doing so, the influence of the noisiest and least accurate spectral bands is reduced, which lowers the estimation variance.

Another important source of error between the measured and simulated spectra is the inherent spectral variability of each considered benthic class. Based on PlanarRad simulations and a comprehensive bottom spectral library, Hedley et al. (2012b) have actually demonstrated that this is one of the primary limiting factors for benthic mapping purposes (whereas sensor noise is only a minor factor). Indeed, while a single mean reflectance spectrum is generally used to characterize the spectral response of each benthic class, many authors show that such intrinsic variability may sometimes be greater than the mean reflectance itself, either at the local or global scales (Hochberg et al., 2003; Mobley et al., 2005; Hedley et al., 2012b; Petit et al., 2017). Therefore, this variability may strongly affect the retrieval accuracy if it is not (or not properly) taken into account during the inversion process. To this end, assuming that the bottom reflectance spectrum only varies according to a single multiplicative factor across all the wavebands, several authors have proposed to estimate this factor for each possible substrate (Lee et al., 1999; Fearns et al., 2011; Garcia et al., 2014b; Petit et al., 2017). Under the same assumption, using the Spectral Angle Mapper (SAM) as a cost function may also decrease the detrimental influence of bottom intra-class variability, since the SAM is insensitive to variations in the global reflectance magnitude (Brando et al., 2009; Botha et al., 2013; Petit et al., 2017). However, this spectral variability cannot always be reliably represented using a single multiplicative factor (Hochberg et al., 2003; Hedley et al., 2012b), thus making the development of alternative inversion methods highly desirable.

In this study, we first propose a realistic probabilistic model of shallow water reflectance variability based on the semi-analytical model of Lee et al. (1998) and that fully describes the influences of environmental noise and bottom intra-class variability. Both sources of error are considered to be Gaussian and characterized by a mean vector and a spectral covariance matrix. Then, using this modeling, we develop two new inversion approaches based on maximum likelihood estimation that enable a pixelwise retrieval of all optically-active parameters, i.e., bathymetry, water clarity parameters and benthic cover. These two approaches are compared with the classical least-squares method using both simulated and airborne data.

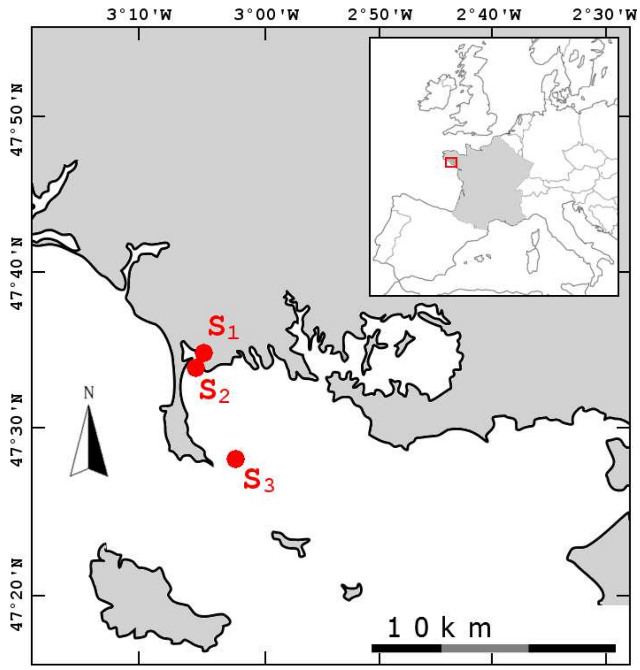

Figure 1: Location of the three study sites S_1, S_2 and S_3.

2. Data

2.1. Study area

As shown in Fig. 1, the overall study area is located in the Quiberon Bay on the French west coast (around 47°31'N, 3°05'W). Three sites (hereafter denoted S_1, S_2 and S_3) were chosen in order to include a large bathymetric range and various bottom covers. Site S_1 and Site S_2 are located near the shore in the Bay of Plouharnel (47°34'46"N, 3°06'24"W), and are characterized by relatively shallow waters (less than 5 m at the time of acquisitions) and heterogeneous bottom covers including sand, brown and green algae, seagrasses and oyster farming structures. Site S_3 is located a few kilometers away from the Quiberon peninsula (47°28'11"N, 3°02'18"W) and is characterized by a large bathymetric range (from 4 to 12 m at the time of acquisitions) and a nearly uniform sandy bottom.

2.2. Image acquisition and preprocessing

Eight hyperspectral images were acquired on September 14-18, 2010 around solar noon (the solar zenith angle being close to 50^o) using an airborne Hyspex VNIR-1600 push-broom camera (Norsk Elektro Optikk, Norway). The flight altitude was 650 m, resulting in a 0.5 m spatial resolution. The camera acquired successive lines of 1600 pixels and 160 spectral bands

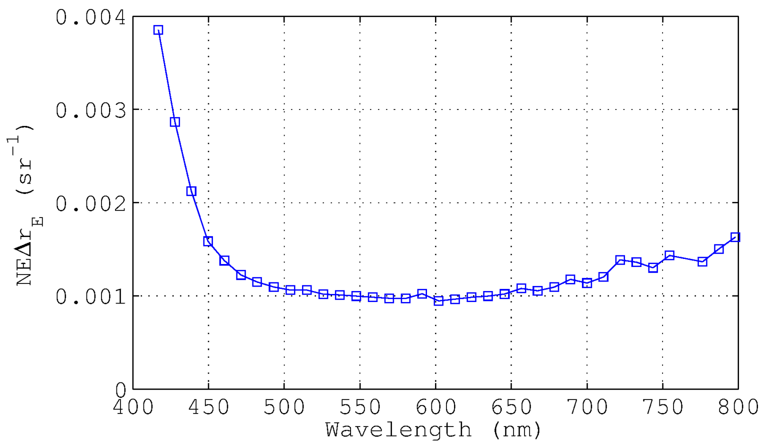

Figure 2: Environmental noise as measured on September 18, 2010.

ranging from 410 to 987 nm. The spectral sampling interval and full width at half maximum were 3.7 nm and 4.5 nm respectively. Only 105 bands in the 410-800 nm domain were kept when removing the strong water and oxygen absorption regions. Further, a three-band aggregate was performed similarly to the PRISM instrument developed by the Jet Propulsion Laboratory (Mouroulis et al., 2014), therefore leading to a 11 nm sampling interval (35 bands). This allows us to enhance the signal-to-noise ratio while keeping similar estimation results (Hochberg & Atkinson, 2003; Garcia et al., 2015).

The at-sensor radiance images were geometrically corrected, geolocated, and converted into above-surface reflectance using the ATCOR atmospheric correction (Richter, 2012) (for further details about these corrections, please see Jay & Guillaume (2016)). Sun glint (Hedley et al., 2005) and the air/water interface (Lee et al., 1999) were corrected in order to finally obtain the subsurface remote-sensing reflectance $r(\lambda)$ (in sr^{-1}). For each day of acquisition, the environmental noise $\mathrm{NE}\Delta r_E$ (in sr^{-1}) (Brando & Dekker, 2003) was estimated over optically deep waters according to the methodology proposed by Wettle et al. (2004). As shown in Fig. 2, its spectral shape is similar to those obtained in previous studies (Brando et al., 2009; Wettle et al., 2004), i.e., $\mathrm{NE}\Delta r_E$ is nearly constant across all wavebands and mainly increases in the blue domain, where the sensitivity of the CCD sensor is the lowest and spectral variations in incident light are the strongest.

2.3. Data used for depth and phytoplankton concentration estimations

The eight hyperspectral images were used to evaluate the accuracy of bathymetry retrieval. For each image, the depth was only known in a few 6×6 m^2 flat sandy-bottom areas thanks to sonar measurements and a tide model. A total of 14 validation points (depth ranging from 0.44 to 12 m) were therefore available to assess the accuracy of bathymetry estimation.

In addition, phytoplankton concentration was also measured concurrently with most airborne acquisitions in Site S_3. To do so, water samples were collected at the surface and bottom (whose depth ranged from 4.70 to 12 m) levels to better account for a possible vertical gradient in phytoplankton concentration. Chlorophyll-a and pheopigment concentrations were measured according to the French standard NF T 90117 (AFNOR, December 1999). Surface and bottom phytoplankton concentrations were then given by the sum of chlorophyll-a and pheopigment concentrations, and averaged so as to obtain a single measurement for each sampled area. These mean values were finally used to derive the absorption coefficient of phytoplankton at 440 nm (denoted P, in m^{-1}) similarly to Lee et al. (1999). In total, 8 validation points (phytoplankton concentration ranging from 1.25 to 1.95 μg.L^{-1}, corresponding to P ranging from 0.069 to 0.093 m^{-1}) were available (still over 6×6 m^2 flat sandy-bottom areas within which P was assumed to be homogeneous).

Note that no data were available to assess the retrievals of the other optically-active water constituents, namely, colored dissolved organic and detrital matter as well as suspended matter (see Section 3.1.1).

2.4. Data used for bottom cover estimation

The above eight images were also used to assess bottom cover estimation over the 14 6×6 m^2 flat sandy-bottom areas of known depth. In addition, one of these images was used to assess the tested methods over more complex bottom covers (Fig. 3). This image was acquired over a 0.22 km^2 area located in site S_2. This shallow area was part of a large oyster farming area and was thus relatively heterogeneous, both in terms of bottom cover and bathymetry (the depth ranged from about 1 m in the left-hand part to 5 m in the top-right part, with locally sharp changes in bathymetry due to the presence of oyster racks). Various

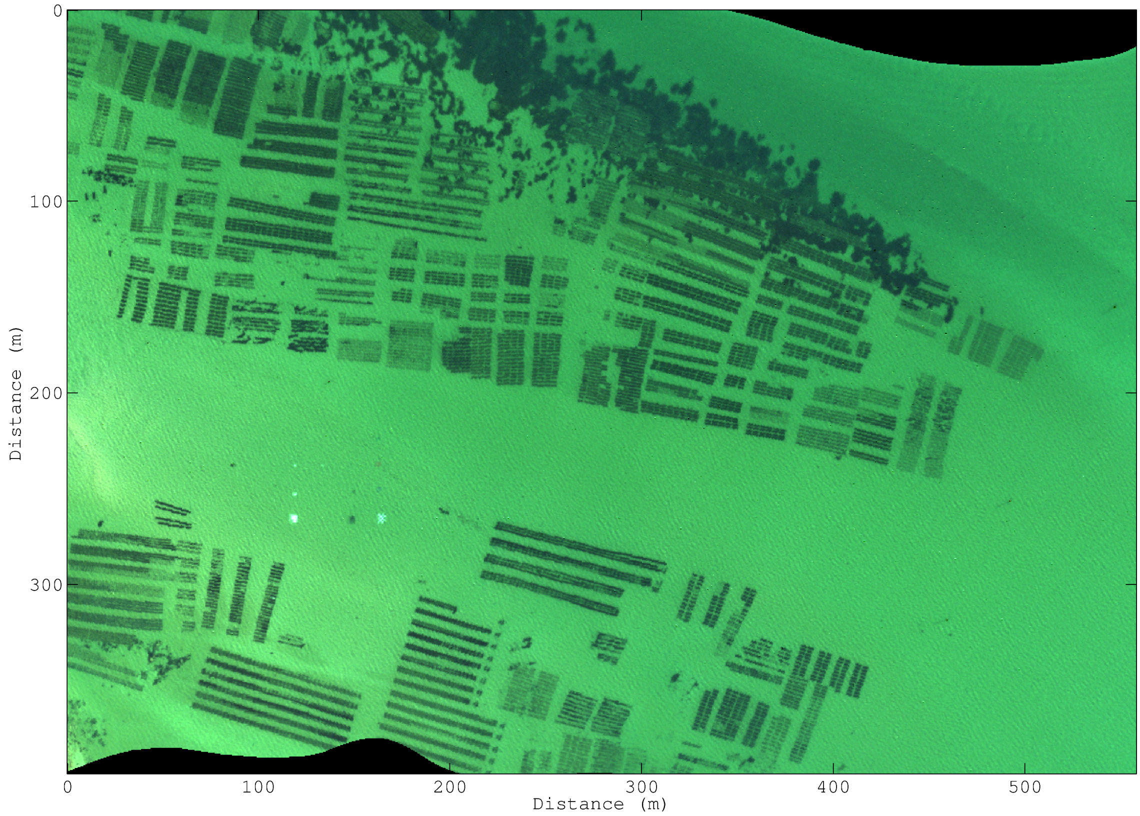

Figure 3: True color composite image derived from the deglinted subsurface remote-sensing reflectance image that was used to assess the bottom cover estimation (note that the dynamic range of the image was enhanced by multiplying every pixel by a factor of 10).

bottom types were identified in this area. Numerous oyster racks were present on a mostly sandy bottom. Some of these wooden structures were empty (e.g., in the upper left part of the image), but most of them were full of oyster bags at the time of acquisition. Depending on when these bags had been put on racks, they could partly or completely be covered with green algae and/or brown algae. Lastly, there was a large seagrass meadow in the upper right part of the image, as well as small patches of brown algae irregularly distributed within the image (e.g., between oyster racks in the lower left part). Note that the colored tarpaulins present on the left-hand side (in the middle of which depth was 2.83 m) were ignored in this study.

For each bottom class and based on expert knowledge, numerous endmember spectra were extracted from supplementary hyperspectral images acquired over the neighboring zones in Site S_2 during low tide (Fig. 4). It is worth mentioning that, due to intra-class variability and because these zones are a few hundred meters to a dozen kilometers from the zones used to assess the inversion methods (Fig. 1), the extracted endmember spectra may not perfectly

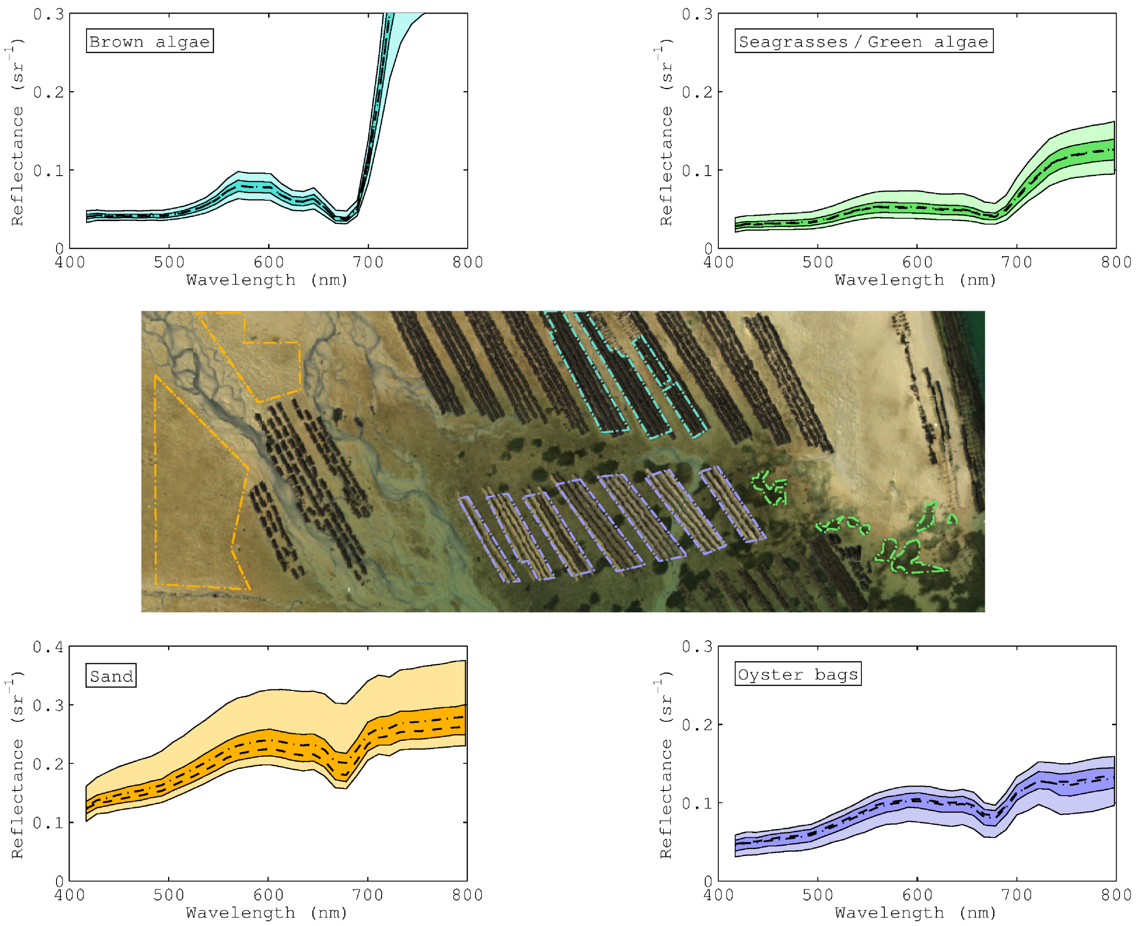

Figure 4: Reflectance distributions of sand, oyster bag, seagrass/green alga and brown alga classes as estimated from the areas emphasized in the airborne hyperspectral image shown in the middle. For each plot, the darkest and brightest shades correspond to the 25-75% and 5-95% quantiles resp., whereas the median and mean spectra are indicated by dashed and dash-dot lines resp..

match those encountered in the whole study area. Selecting reflectance spectra of emerged substrates directly from the remote-sensing images allowed us to avoid potential issues of intercalibration between airborne and ground-based sensors. However, note that, since empty wooden structures were too thin to fill entirely the 0.5×0.5 m^2 pixels of hyperspectral images, they were not included as a possible endmember. Further, green algae and seagrasses were grouped into a single class corresponding to green vegetation elements. Four bottom classes were thus used, namely sand, oyster bags, brown algae and seagrasses/green algae (note that these surfaces were assumed to be Lambertian). The corresponding reflectance distributions were estimated based on 150 to 3,000 image spectra, and all show some intra-class variability around the mean reflectance spectra (Fig. 4). Such variability may be due, e.g., to the bottom chemistry itself (e.g., variations in chlorophyll content in seagrasses/green algae) or to the bottom 3-D arrangement that may make the illumination conditions within the surface highly variable (Manolakis et al., 2003). Given the similar magnitudes of brown alga, seagrass/green alga and, to a lesser extent, oyster bag mean reflectance spectra, such variability potentially makes the identification of these three partially overlapping classes quite difficult.

3. Methodology

3.1. Forward modeling of subsurface remote-sensing reflectance

3.1.1. Bio-optical modeling

In this study, we use the semi-analytical model $\tilde{r}(\lambda)$ developed by Lee et al. (1998, 1999) to express the subsurface remote-sensing reflectance as measured from nadir as a function of depth H (in m), bottom albedo $\rho_b(\lambda)$ (unitless), total absorption and backscattering properties of the water column $a(\lambda)$ and $b_b(\lambda)$ resp. (in m^{-1}), and subsurface solar zenith angle θ_s (in ^{o}):

$$\tilde{r}(\lambda) = r_\infty(\lambda)\left(1 - \mathrm{e}^{-(k_d(\lambda)+k_u^c(\lambda))H}\right) + \frac{\rho_b(\lambda)}{\pi}\mathrm{e}^{-\left(k_d(\lambda)+k_u^b(\lambda)\right)H} \quad (1)$$

where the subsurface remote-sensing reflectance of optically-deep water $r_\infty(\lambda)$ (in sr^{-1}) and attenuation coefficients $k_d(\lambda)$, $k_u^c(\lambda)$ and $k_u^b(\lambda)$ (in m^{-1}) are related to $a(\lambda)$, $b_b(\lambda)$ and θ_s by:

$$\begin{aligned}
r_\infty(\lambda) &= \left(0.084 + 0.17\frac{b_b(\lambda)}{a(\lambda)+b_b(\lambda)}\right)\frac{b_b(\lambda)}{a(\lambda)+b_b(\lambda)} && (2)\\
k_d(\lambda) &= \frac{a(\lambda)+b_\mathrm{b}(\lambda)}{\cos\theta_\mathrm{s}} && (3)\\
k_u^b(\lambda) &= 1.04(a(\lambda)+b_b(\lambda))\left(1+5.4\frac{b_b(\lambda)}{a(\lambda)+b_b(\lambda)}\right)^{0.5} && (4)\\
k_u^c(\lambda) &= 1.03(a(\lambda)+b_b(\lambda))\left(1+2.4\frac{b_b(\lambda)}{a(\lambda)+b_b(\lambda)}\right)^{0.5}. && (5)
\end{aligned}$$

Eq. (1) to Eq. (5) have been used and validated in numerous studies dealing with shallow water remote sensing over a wide range of coastal waters (Lee et al., 1999, 2001; Klonowski et al., 2007; Goodman et al., 2008; Brando et al., 2009; Hedley et al., 2009; Dekker et al., 2011; Fearns et al., 2011; Garcia et al., 2014a; Jay & Guillaume, 2014; McKinna et al., 2015; Jay & Guillaume, 2016; Petit et al., 2017). In the absence of *in-situ* measurements of inherent optical properties to develop a site-specific bio-optical model, the total absorption and backscattering coefficients are given by the sum of the contributions of optically-active water constituents and parameterized according to the generic expressions of Lee et al. (1998) and Dekker et al. (2011):

$$a(\lambda) = a_w(\lambda) + [a_0(\lambda) + a_1(\lambda)\ln P]\,P + G\mathrm{e}^{-0.015(\lambda-440)} \tag{6}$$

$$b_b(\lambda) = b_{b,w}(\lambda) + X\left(\frac{550}{\lambda}\right)^{0.5} \tag{7}$$

where $a_w(\lambda)$ and $b_{b,w}(\lambda)$ (in m^{-1}) are the pure water absorption and backscattering coefficients (Buiteveld et al., 1994; Morel, 1974), $a_0(\lambda)$ and $a_1(\lambda)$ (unitless) are empirical spectra tabulated by Lee et al. (1998), P (in m^{-1}) is the absorption coefficient of phytoplankton at 440 nm, G (in m^{-1}) is the absorption coefficient of colored dissolved organic and detrital matter at 440 nm, and X (in m^{-1}) is the particle backscattering coefficient at 550 nm. The above parameterizations of absorption coefficients of phytoplankton and colored dissolved organic and detrital matter have been shown to be sufficiently accurate over a wide range of coastal waters (Lee et al., 1999, 2001; Lee & Carder, 2002; Goodman et al., 2008; Hedley et al., 2009; Dekker et al., 2011; Hedley et al., 2012a; Jay & Guillaume, 2014, 2016). Note also that the power law exponent used to model particle backscattering was set to -0.5, which is adequate for normal to more turbid coastal waters (Lee et al., 2001).

In order to accurately model the response of oyster racks (which are relatively thin compared to the 0.5×0.5 m^2 pixel size) while appropriately limiting the number of unknowns and, therefore, the estimation uncertainty, the bottom albedo is parameterized using a linear combination of two pure substrates similarly to Brando et al. (2009) and Hedley et al. (2009):

$$\rho_b(\lambda) = B_1\rho_{b,1}(\lambda) + B_2\rho_{b,2}(\lambda) \tag{8}$$

where $\rho_{b,1}(\lambda)$ and $\rho_{b,2}(\lambda)$ are two known substrate albedos (e.g., obtained from ground-based measurements or a generic spectral library). The scalars B_1 and B_2 (unitless) may represent the fractional covers of both substrates within the considered pixel, so in this case, only one bottom coefficient B is required, i.e., $B_1 = B$, $B_2 = 1 - B$ and $0 \leq B \leq 1$ (Klonowski et al., 2007; Goodman & Ustin, 2007; Brando et al., 2009; Hedley et al., 2009, 2012a). Alternatively, Fearns et al. (2011) and Garcia et al. (2014b) used a mixture of benthic reflectances normalized at 550 nm, and they estimated the relative brightness of each substrate without imposing any constraint on the mixture coefficients to be retrieved. In this case, a single multiplicative factor is used to model both the fractional cover and the brightness (or magnitude) of each substrate. Although the sum-to-one constraint applies for the fractional cover, the brightness of substrate $\rho_{b,1}$ is independent from that of substrate $\rho_{b,2}$. As a result, the mixture coefficients B_1 and B_2 are independent and do not necessarily sum to one. It is worth noting that, even though such a modeling enables the magnitudes of $\rho_{b,1}$ and $\rho_{b,2}$ to vary, it also adds an extra degree of freedom during the inversion process. This may increase the estimation noise and require post-processing steps in order to smooth estimated maps, e.g., using median filtering (Fearns et al., 2011). In the following, we test these two approaches in order to assess the impact of the sum-to-one constraint on estimation performance.

3.1.2. Probabilistic modeling

As widely accepted in the community (Jay & Guillaume, 2011; Hedley et al., 2012a; Jay et al., 2012; Gillis et al., 2013; Garcia et al., 2014b; Jay & Guillaume, 2014; Knudby et al., 2016), the measured subsurface remote-sensing reflectance, denoted in vector form $\boldsymbol{r} = [r(\lambda_1), ..., r(\lambda_L)]^t$ (where L is the number of wavebands), is assumed to follow a multivariate Gaussian distribution with mean $\boldsymbol{\mu} = \mathbb{E}[\boldsymbol{r}]$ and spectral covariance matrix $\boldsymbol{\Gamma} = \mathbb{E}[(\boldsymbol{r} - \mathbb{E}(\boldsymbol{r}))(\boldsymbol{r} - \mathbb{E}(\boldsymbol{r}))^t]$. The mean vector is parameterized using the bio-optical model presented in Section 3.1.1, which may be written in matrix notation as

$$\boldsymbol{\mu}(\boldsymbol{\Delta}) = (\mathbb{I} - \boldsymbol{K}_c)\boldsymbol{r}_\infty + \boldsymbol{K}_b \left(B_1 \frac{\boldsymbol{\rho}_{b,1}}{\pi} + B_2 \frac{\boldsymbol{\rho}_{b,2}}{\pi} \right) \tag{9}$$

where $\boldsymbol{\Delta} = [H, P, G, X, B_1, B_2]^t$, $\boldsymbol{r}_\infty = [r_\infty(\lambda_1), ..., r_\infty(\lambda_L)]^t$, $\mathbb{I}$ is the $L \times L$ identity matrix, $\boldsymbol{K}_c = \mathrm{diag}\left[\mathrm{e}^{-(k_d(\lambda_i)+k_u^c(\lambda_i))H}\right]_{i\in[\![1;L]\!]}$, $\boldsymbol{K}_b = \mathrm{diag}\left[\mathrm{e}^{-(k_d(\lambda_i)+k_u^b(\lambda_i))H}\right]_{i\in[\![1;L]\!]}$, and $\boldsymbol{\rho}_{b,i} = [\rho_{b,i}(\lambda_1), ..., \rho_{b,i}(\lambda_L)]^t$.

The different sources of deviations between the measured and simulated spectra can be modeled via an appropriate parameterization of $\boldsymbol{\Gamma}$. In the probabilistic modeling subsequently used within the proposed MILE (MaxImum Likelihood estimation including Environmental noise) inversion method (Section 3.2), we assume that the random variability around mean $\boldsymbol{\mu}(\boldsymbol{\Delta})$ can be described using the full spectral covariance matrix of the environmental noise, $\boldsymbol{\Gamma}_{surf}$, similarly to Hedley et al. (2012a), Garcia et al. (2014b) and Knudby et al. (2016). The subsurface remote-sensing reflectance is then modeled as

$$\boldsymbol{r} = \left[(\mathbb{I} - \boldsymbol{K}_c)\boldsymbol{r}_\infty + \boldsymbol{K}_b\left(B_1\frac{\boldsymbol{\rho}_{b,1}}{\pi} + B_2\frac{\boldsymbol{\rho}_{b,2}}{\pi}\right)\right] + \boldsymbol{n}_{surf} \tag{10}$$

where the random vector $\boldsymbol{n}_{surf}$ follows a multivariate Gaussian distribution with zero mean and covariance matrix $\boldsymbol{\Gamma}_{surf}$. Note that, in real scenarios, $\boldsymbol{\Gamma}_{surf}$ can be estimated over optically deep waters similarly to NEΔr_E.

However, Eq. (10) only allows the bottom remote-sensing reflectances $\left(\boldsymbol{\rho}_{b,1}/\pi\right)$ and $\left(\boldsymbol{\rho}_{b,2}/\pi\right)$ to vary according to the multiplicative factors B_1 and B_2. As an alternative to this usual bottom modeling, the proposed MILEBI (MaxImum Likelihood estimation including Environmental noise and Bottom Intra-class variability) probabilistic modeling uses a multivariate Gaussian distribution to describe the reflectance inherent variability of each benthic class. Due to the compromise offered between accuracy and mathematical tractability, the Gaussian modeling has been widely used to develop hyperspectral remote-sensing algorithms that must take into account the spread of each class of materials (and therefore potential overlaps between these classes) to obtain good performances, e.g., classification and target detection algorithms (Manolakis et al., 2003; Melgani & Bruzzone, 2004; Palmason et al., 2005). Preliminary tests (not shown here for the sake of brevity) demonstrated that, except for a small minority of samples corresponding to extreme data points, the bottom intra-class variabilities presented in Fig. 4 could indeed be reliably represented using multivariate Gaussian distributions. In this case, the subsurface remote-sensing reflectance can be modeled as

$$\boldsymbol{r} = \left\{(\mathbb{I} - \boldsymbol{K}_c)\boldsymbol{r}_\infty + \boldsymbol{K}_b\left[B_1(\boldsymbol{\mu}_{b,1} + \boldsymbol{n}_{b,1}) + B_2(\boldsymbol{\mu}_{b,2} + \boldsymbol{n}_{b,2})\right]\right\} + \boldsymbol{n}_{surf} \tag{11}$$

where $\boldsymbol{\mu}_{b,i}$ is the mean remote-sensing reflectance spectrum of bottom class i and $\boldsymbol{n}_{b,i}$ follows a multivariate Gaussian distribution with zero mean and covariance matrix $\boldsymbol{\Gamma}_{b,i}$. Separating deterministic terms from random terms in Eq. (11) leads to

$$\boldsymbol{r} = \left[(\mathbb{I} - \boldsymbol{K}_c)\boldsymbol{r}_\infty + \boldsymbol{K}_b\left(B_1\boldsymbol{\mu}_{b,1} + B_2\boldsymbol{\mu}_{b,2}\right)\right] + \left[\boldsymbol{n}_{surf} + \boldsymbol{K}_b(B_1\boldsymbol{n}_{b,1} + B_2\boldsymbol{n}_{b,2})\right]. \tag{12}$$

The corresponding total covariance matrix is obtained by applying $\boldsymbol{\Gamma} = \mathbb{E}\left[(\boldsymbol{r} - \mathbb{E}(\boldsymbol{r}))(\boldsymbol{r} - \mathbb{E}(\boldsymbol{r}))^t\right]$ to Eq. (12) and by assuming that $\boldsymbol{n}_{b,1}$, $\boldsymbol{n}_{b,2}$ and $\boldsymbol{n}_{surf}$ are independent:

$$\boldsymbol{\Gamma}(\boldsymbol{\Delta}) = \boldsymbol{K}_b\left[{B_1}^2\boldsymbol{\Gamma}_{b,1} + {B_2}^2\boldsymbol{\Gamma}_{b,2}\right]\boldsymbol{K}_b + \boldsymbol{\Gamma}_{surf}. \tag{13}$$

In Eq. (12), possible deviations between the observed subsurface remote-sensing reflectance $\boldsymbol{r}$ and the model (left-hand term of the sum) are not only due to the environmental noise, but also to the intrinsic spectral variability of each benthic class. As expected, for the i^{th} class, the influence of this variability is proportional to B_i, and becomes negligible when depth and/or turbidity increase(s) (because of progressive attenuation by $\boldsymbol{K}_b$). Also, if $\boldsymbol{\Gamma}_{b,1}$ and $\boldsymbol{\Gamma}_{b,2}$ perfectly describe the bottom intrinsic variabilities, the parameters B_1 and B_2 only represent fractional covers, so the sum-to-one constraint applies. In this case, the MILEBI probabilistic modeling disentangles the fractional cover (which is taken into account by a single multiplicative factor $B = B_1 = 1 - B_2$) from intra-class variabilities (which are taken into account through the bottom covariance matrices $\boldsymbol{\Gamma}_{b,1}$ and $\boldsymbol{\Gamma}_{b,2}$), which is not possible when using Eq. (10). Alternatively, relaxing the sum-to-one constraint may allow potential deviations from the assumed Gaussian modeling.

3.2. Inversion methods

In this study, various inversion methods are derived based on the above two probabilistic models of shallow water reflectance variability. All these inversion methods consist in maximizing the likelihood of observing $\boldsymbol{r}$ given the set $\boldsymbol{\Delta}$ of water column parameters to be estimated. Under the Gaussian assumption, the likelihood is defined as

$$\mathrm{P}(\boldsymbol{r}|\boldsymbol{\Delta}) = \left[(2\pi)^L|\boldsymbol{\Gamma}(\boldsymbol{\Delta})|\right]^{-1/2}\mathrm{e}^{-\frac{1}{2}(\boldsymbol{r}-\boldsymbol{\mu}(\boldsymbol{\Delta}))^t\boldsymbol{\Gamma}(\boldsymbol{\Delta})^{-1}(\boldsymbol{r}-\boldsymbol{\mu}(\boldsymbol{\Delta}))}. \tag{14}$$

The maximum likelihood estimate $\widehat{\mathbf{\Delta}}_{ML}(\boldsymbol{r})$ is the value of $\mathbf{\Delta}$ that maximizes the likelihood:

$$\widehat{\mathbf{\Delta}}_{ML}(\boldsymbol{r}) = \underset{\mathbf{\Delta}}{\operatorname{argmax}}\, \mathrm{P}(\boldsymbol{r}|\mathbf{\Delta}). \tag{15}$$

In Eq. (14), the mean vector $\boldsymbol{\mu}(\mathbf{\Delta})$ is given by Eq. (9) for every tested inversion method. The main difference between the methods actually lies in the parameterization of $\boldsymbol{\Gamma}(\mathbf{\Delta})$. In MILE, $\boldsymbol{\Gamma}(\mathbf{\Delta}) = \boldsymbol{\Gamma}_{surf}$ does not depend on $\mathbf{\Delta}$ since it only characterizes the above-water variability. Eq. (14) can thus be simplified, and the MILE estimate $\widehat{\mathbf{\Delta}}_{MILE}(\boldsymbol{r})$ is given by the minimum Mahalanobis distance between the measured and simulated spectra:

$$\widehat{\mathbf{\Delta}}_{MILE}(\boldsymbol{r}) = \underset{\mathbf{\Delta}}{\operatorname{argmin}}(\boldsymbol{r} - \boldsymbol{\mu}(\mathbf{\Delta}))^t \boldsymbol{\Gamma}_{surf}{}^{-1}(\boldsymbol{r} - \boldsymbol{\mu}(\mathbf{\Delta})). \tag{16}$$

In MILEBI, $\boldsymbol{\Gamma}(\mathbf{\Delta})$ depends on $\mathbf{\Delta}$, so Eq. (14) cannot be further simplified:

$$\widehat{\mathbf{\Delta}}_{MILEBI}(\boldsymbol{r}) = \underset{\mathbf{\Delta}}{\operatorname{argmax}} \left\{ \left[(2\pi)^L |\boldsymbol{\Gamma}(\mathbf{\Delta})|\right]^{-1/2} \mathrm{e}^{-\frac{1}{2}(\boldsymbol{r}-\boldsymbol{\mu}(\mathbf{\Delta}))^t \boldsymbol{\Gamma}(\mathbf{\Delta})^{-1}(\boldsymbol{r}-\boldsymbol{\mu}(\mathbf{\Delta}))} \right\} \tag{17}$$

where $\boldsymbol{\Gamma}(\mathbf{\Delta})$ is given by Eq. (13).

In this paper, MILE and MILEBI are compared to the widely used least-squares (LS) method. Note that the LS estimate can also be obtained by maximizing the likelihood in Eq. (14), taking $\boldsymbol{\Gamma} = \sigma^2 \mathbb{I}$ where σ is a positive real number and $\mathbb{I}$ is the $L \times L$ identity matrix (i.e., uncertainties of all spectral bands are assumed to be uncorrelated and of equal variances). The LS estimate $\widehat{\mathbf{\Delta}}_{LS}(\boldsymbol{r})$ is given by the minimum Euclidean distance between the measured and simulated spectra:

$$\widehat{\mathbf{\Delta}}_{LS}(\boldsymbol{r}) = \underset{\mathbf{\Delta}}{\operatorname{argmin}}(\boldsymbol{r} - \boldsymbol{\mu}(\mathbf{\Delta}))^t (\boldsymbol{r} - \boldsymbol{\mu}(\mathbf{\Delta})). \tag{18}$$

Comparing Eq. (16), Eq. (17) and Eq. (18) shows that, unlike LS, MILE and MILEBI utilize the information contained in the spectral covariance matrix to further constrain the inversion. In addition, both methods allow some deviations between the measured and simulated spectra by giving the less reliable wavebands little weights in the cost function. For MILE, these are located in the domains of strong environmental noise. For MILEBI, these wavebands not

Table 1: Methods compared in this study and derived from the likelihood function presented in Eq. (14). Subscript "S21" indicates the use of the sum-to-one constraint.

Method	$\boldsymbol{\Delta}$	$\boldsymbol{\mu}(\boldsymbol{\Delta})$	$\boldsymbol{\Gamma}(\boldsymbol{\Delta})$
$\mathrm{LS}_{\mathrm{S21}}$	$[H, P, G, X, B]$	$(\mathbb{I} - \boldsymbol{K}_c)\boldsymbol{r}_\infty + \boldsymbol{K}_b\left(B\frac{\boldsymbol{\rho}_{b,1}}{\pi} + (1-B)\frac{\boldsymbol{\rho}_{b,2}}{\pi}\right)$	$\sigma^2\mathbb{I}$
$\mathrm{MILE}_{\mathrm{S21}}$	$[H, P, G, X, B]$	$(\mathbb{I} - \boldsymbol{K}_c)\boldsymbol{r}_\infty + \boldsymbol{K}_b\left(B\frac{\boldsymbol{\rho}_{b,1}}{\pi} + (1-B)\frac{\boldsymbol{\rho}_{b,2}}{\pi}\right)$	$\boldsymbol{\Gamma}_{surf}$
$\mathrm{MILEBI}_{\mathrm{S21}}$	$[H, P, G, X, B]$	$(\mathbb{I} - \boldsymbol{K}_c)\boldsymbol{r}_\infty + \boldsymbol{K}_b\left(B\frac{\boldsymbol{\rho}_{b,1}}{\pi} + (1-B)\frac{\boldsymbol{\rho}_{b,2}}{\pi}\right)$	$\boldsymbol{K}_b\left[B^2\boldsymbol{\Gamma}_{b,1} + (1-B)^2\boldsymbol{\Gamma}_{b,2}\right]\boldsymbol{K}_b + \boldsymbol{\Gamma}_{surf}$
LS	$[H, P, G, X, B_1, B_2]$	$(\mathbb{I} - \boldsymbol{K}_c)\boldsymbol{r}_\infty + \boldsymbol{K}_b\left(B_1\frac{\boldsymbol{\rho}_{b,1}}{\pi} + B_2\frac{\boldsymbol{\rho}_{b,2}}{\pi}\right)$	$\sigma^2\mathbb{I}$
MILE	$[H, P, G, X, B_1, B_2]$	$(\mathbb{I} - \boldsymbol{K}_c)\boldsymbol{r}_\infty + \boldsymbol{K}_b\left(B_1\frac{\boldsymbol{\rho}_{b,1}}{\pi} + B_2\frac{\boldsymbol{\rho}_{b,2}}{\pi}\right)$	$\boldsymbol{\Gamma}_{surf}$
MILEBI	$[H, P, G, X, B_1, B_2]$	$(\mathbb{I} - \boldsymbol{K}_c)\boldsymbol{r}_\infty + \boldsymbol{K}_b\left(B_1\frac{\boldsymbol{\rho}_{b,1}}{\pi} + B_2\frac{\boldsymbol{\rho}_{b,2}}{\pi}\right)$	$\boldsymbol{K}_b\left[{B_1}^2\boldsymbol{\Gamma}_{b,1} + {B_2}^2\boldsymbol{\Gamma}_{b,2}\right]\boldsymbol{K}_b + \boldsymbol{\Gamma}_{surf}$

only correspond to the domains of strong environmental noise, but also to the domains of strong bottom intrinsic variabilities.

Implementing MILE, MILEBI and LS with or without the sum-to-one constraint on bottom mixture coefficients results in the six methods summarized in Table 1. Note that other cost functions, such as SAM or least-squares on spectral derivative (Brando et al., 2009; Botha et al., 2013; Petit et al., 2017), could also be tested, since, for example, SAM may provide more accurate bathymetry retrieval than LS (Petit et al., 2017). We, however, only compared MILE, MILEBI and LS (1) in order to focus primarily on the influence of $\boldsymbol{\Gamma}(\boldsymbol{\Delta})$ parameterization on the inversion, and (2) because LS generally offers a better tradeoff than SAM and least-squares on spectral derivative for accurately retrieving all the parameters at the same time (Petit et al., 2017).

3.3. Implementation of inversion methods

For the six methods presented in Table 1, the cost function was iteratively optimized using the trust-region reflective algorithm implemented in MATLAB® (version 8.0.0, The MathWorks Inc., Natick, MA, 2012) within the "lsqcurvefit" function. Lower and upper optimization bounds were similar to those found in the literature for turbid waters (Hedley et al., 2009; Garcia et al., 2014b, 2015), i.e., $0 \leq H \leq 30$ m, $0 \leq P \leq 0.5$ m^{-1}, $0 \leq G \leq 0.5$ m^{-1}, $0 \leq X \leq 0.08$ m^{-1}, $0 \leq B_1, B_2 \leq 1.5$ and $0 \leq B \leq 1$.

A special attention was given to the initialization step. While default parameter values (Lee et al., 2001; Klonowski et al., 2007; McKinna et al., 2015) or reflectance-derived values (Lee et al., 1999; Dekker et al., 2011; Jay & Guillaume, 2016) may be used as initial guesses, Garcia et al. (2014a,b) have shown that different initial guesses could lead to different local minima and therefore different parameter estimates. This step may be more critical in the case of maximum likelihood estimation because considering spectrally-correlated noise may introduce more local minima in the parameter solution space (Garcia et al., 2014b). In this paper, we thus implemented a Latin Hypercube Sampling scheme as proposed by Garcia et al. (2014b) to generate preliminary LUTs containing 100,000 initial guesses and corresponding simulated reflectance spectra. Normal distributions were used for H, P, G and X, and uniform distributions bounded by the above lower and upper bounds were used for B, B_1 and B_2. Empirical values were used for means and standard deviations of normal distributions: means were set to 0, while standard deviations were set such that the value of the probability density function at half maximum corresponded to one-third of the upper bound (e.g., we used a standard deviation of 8.5 m for H). Only positive sets of parameters were then kept to build the LUTs. The use of such normal distributions allowed us to sample more finely the regions of the parameter space where the reflectance strongly varies with depth and water clarity parameters, namely, shallow waters and high water clarity (Hedley et al., 2009; Jay & Guillaume, 2016). For each measured spectrum to be inverted, the 100 sets of parameters corresponding to the 100 closest spectra in the LUT were averaged to provide a single initial guess for the iterative optimization process. In vegetation remote sensing, averaging multiple best solutions instead of retaining only the best one is known to increase the estimation accuracy when the inversion problem is ill-posed and/or the reflectance model is not fully accurate (Darvishzadeh et al., 2011; Verrelst et al., 2015; Jay et al., 2017).

In this study, four substrates were identified as possible endmembers (Fig. 4). As only two of them could be used in the bottom reflectance model (Eq. (8)), we implemented the same type of approach as Brando et al. (2009), i.e., (1) each measured reflectance spectrum was inverted using each of the six possible pairs of substrates (note that this requires generating six preliminary LUTs for initialization), and (2) these six pairs were sorted according to their $\mathrm{P}(\boldsymbol{r}|\boldsymbol{\Delta})$ value. For similar reasons as for initialization and unlike Brando et al. (2009) who only retained the best pair (i.e., corresponding to the highest $\mathrm{P}(\boldsymbol{r}|\boldsymbol{\Delta})$ value, $\mathrm{P}_{\max}$), the solution was here obtained by averaging all pairs whose $\mathrm{P}(\boldsymbol{r}|\boldsymbol{\Delta})$ values were sufficiently close to $\mathrm{P}_{\max}$, i.e., within $n\%$ of $\mathrm{P}_{\max}$. In the following, the value of n was investigated based on simulated data (Section 4.2), testing $n = 0$ (i.e., only the best pair is retained), 1 and 2%. The optimum value was then used for processing the airborne data (Section 4.3).

The four bottom intra-class covariance matrices used in MILEBI and MILEBI$_{\mathrm{S21}}$ were estimated from hyperspectral images acquired at low tide, similarly to the mean reflectance spectra (see Section 2.4). It is worth noting that inverting the covariance matrices detailed in Table 1 requires (at least) as many samples (i.e., spectra) as spectral bands for $\boldsymbol{\Gamma}_{surf}$ and $\boldsymbol{\Gamma}_{b,i}$ estimations. The more samples we have, the more accurate the estimations. In this paper, a minimum of 150 spectra (for oyster bag class) were used, this number being substantially higher than the number of spectral bands (35).

3.4. Performance assessment

3.4.1. Simulated data

We conducted two series of simulations, each of which corresponded to a different model to generate the synthetic data set. For the first data set, we used the probabilistic modeling of Eq. (10), therefore assuming that the random variability is only described by $\boldsymbol{\Gamma}_{surf}$. The influence of water column properties was studied at four depths, i.e., 1, 5, 10 and 20 m, and intermediate water clarity as given by Garcia et al. (2015), i.e., $P = 0.1$ m^{-1}, $G = 0.1$ m^{-1}, and $X = 0.01$ m^{-1}. The bottom was given either as one of the four substrates shown in Fig. 4, or as a 50%/50% mixture of two substrates, thus resulting in ten tested bottom spectra. Note that intra-class variability was not simulated for this data set. We used the $\boldsymbol{\Gamma}_{surf}$ matrix that was estimated over optically deep waters from the airborne data set presented in Section 2, the diagonal of $\boldsymbol{\Gamma}_{surf}$ being given as the square of NEΔr_E shown in Fig. 2. The sun-sensor geometry was identical to that used for airborne acquisitions, i.e., nadir viewing and a solar zenith angle of 50^o.

The second synthetic data set was generated using the probabilistic modeling of Eq. (12). As compared with the first data set, the only difference related to the simulation of bottom reflectance, which was here not only modeled using multiplicative factors, but also using random vectors $\boldsymbol{n}_{b,1}$ and $\boldsymbol{n}_{b,2}$. These vectors were generated based on the intra-class covariance matrices estimated from airborne data (see Section 3.3).

For each data set, the "mvnrnd" MATLAB function allowed us to generate 100 noise-perturbed spectra for every depth and bottom reflectance, hence providing 4,000 simulated spectra in total. These spectra were then inverted using the six methods and according to the procedure described in Section 3.3. The estimation performances were evaluated in terms of mean absolute error (MAE), which has proven to be a more reliable measure of error than the classical root mean square error (Willmott & Matsuura, 2005).

3.4.2. Airborne data

The retrievals of bathymetry, absorption of phytoplankton at 440 nm and bottom cover were also assessed using the airborne data set (Section 2). For each 6×6 m^2 flat sandy-bottom area (thus containing 12×12 pixels), the semi-analytical model was inverted for each pixel using the six methods, and estimated values of H, P and bottom coefficients were compared to their actual values whenever possible. The six methods were also used to retrieve the bottom cover for the image presented in Fig. 3, the estimated benthic habitats being qualitatively evaluated by visual inspection.

4. Results and discussion

4.1. Influences of environmental noise and bottom intra-class variability on subsurface reflectance

A preliminary study was conducted to quantify the influences of environmental noise and bottom intra-class variability on the measured subsurface reflectance, based on the total covariance matrix presented in Eq. (13). Representing this matrix for the four depths investigated in the simulations (same water quality) and the four pure substrates presented in Fig. 4 allows us to see how these two sources of error make the observation deviate from the model (note that, if the bio-optical model in Eq. (9) would be perfect, the total covariance matrix would be the zero matrix).

In the absence of water, the four bottom intra-class covariance matrices show quite different patterns and magnitudes (Fig. 5). While, overall, sand and oyster bag variabilities steadily increase with wavelength, brown algae and, to a lesser extent, seagrasses/green algae, show lower variability in the blue and red domains due to the strong chlorophyll absorption leading

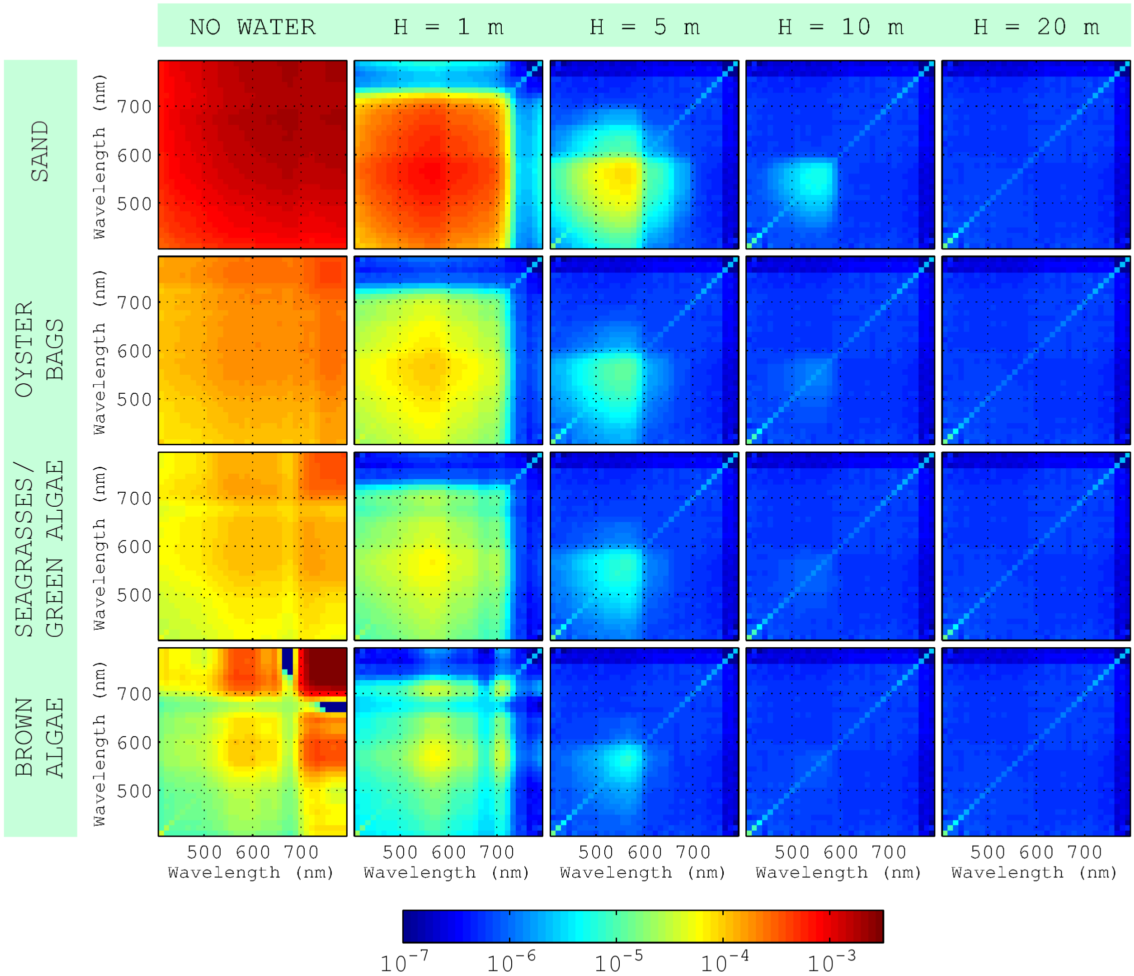

Figure 5: Total covariance matrix (as defined by Eq. (13)) as a function of depth for the four pure substrates investigated ($P = 0.1$ m^{-1}, $G = 0.1$ m^{-1} and $X = 0.01$ m^{-1}). The color scale is the same for every matrix.

to reflectance saturation. For the four substrates, the influence of bottom intra-class variability (resp., environmental noise) decreases (resp., increases) with increasing optical depth. At 1 m and, to a lesser extent, 5 m, the subsurface reflectance variability in the visible domain is primarily driven by the bottom intra-class variability, showing that the latter should not be neglected for such optically shallow waters as also observed by Hedley et al. (2012b). Note that, at 1 m and for most wavebands larger than 700 nm, the water attenuation is already such that the total covariance matrix is mainly dominated by the environmental noise for the four substrates. At 10 m, the influence of environmental noise tends to overshadow that of bottom intra-class variability; only the variability of the brightest benthic class, namely sand, affects the subsurface reflectance in the domain of lower absorption (i.e., in the green region for this water type). In optically deep waters (20 m), the bottom is not visible so only the environmental noise contributes to the total covariance matrix. Of course, note that the relative influences of environmental noise and bottom intra-class variability as functions of optical depth depend on their magnitude, meaning that they should be re-evaluated for every sensor, study area, etc.

To our knowledge, only a few authors (e.g., Hedley et al. (2012b)) have thoroughly analyzed the influence of bottom intra-class spectral variability on subsurface reflectance. Using the analytical expression in Eq. (13) appears as a simple but convenient way to undertake such an analysis and to investigate how accurate Eq. (8) is in modeling the total bottom reflectance.

4.2. Estimation results obtained with the simulated data

In Fig. 6 and Fig. 7, we show the inversion results obtained from the two synthetic data sets presented in Section 3.4.1. Importantly, as the bottom reflectance variability was simulated differently in these two data sets, we only present $\mathrm{LS}_{\mathrm{S21}}$, $\mathrm{MILE}_{\mathrm{S21}}$, LS and MILE (resp. $\mathrm{MILEBI}_{\mathrm{S21}}$ and MILEBI) bottom estimation results when using the first (resp. the second) data set. For both data sets, we, however, show the H, P, G and X estimation results for the six methods in order to study the influence of bottom mismodeling.

For each method, the H estimation error is similar for both data sets and increases with depth (Fig. 6). It could be shown that this increase is caused both by a progressive H underestimation and by an increasing estimation variance. Overall, $\mathrm{MILE}_{\mathrm{S21}}$ and $\mathrm{MILEBI}_{\mathrm{S21}}$ (resp., MILE and MILEBI) provide lower errors than $\mathrm{LS}_{\mathrm{S21}}$ (resp., LS). For example, at 10 m (first data set, $n = 0\%$), the MAEs are 1.52, 1.63 and 2.32 m for $\mathrm{MILEBI}_{\mathrm{S21}}$, $\mathrm{MILE}_{\mathrm{S21}}$ and $\mathrm{LS}_{\mathrm{S21}}$ resp.. Using the sum-to-one constraint generally improves the performances, especially for $H \geq 5$ m, MILEBI, MILE and LS respectively obtaining MAEs of 2.48, 2.46 and 3.14 m at 10 m.

On the one hand, the P and G errors tend to show a bowl-shaped pattern with respect to depth (the minimum being located at 5 m in most cases), especially when considering the second data set. On the other hand, the X error steadily declines with increasing depth (Fig. 6). Similarly to H, MILE- and MILEBI-based methods generally better estimate these water clarity parameters than LS-based methods. This is more visible for $H \geq 5$ m, for which similar errors are generally obtained with $\mathrm{MILE}_{\mathrm{S21}}$, $\mathrm{MILEBI}_{\mathrm{S21}}$, MILE and MILEBI.

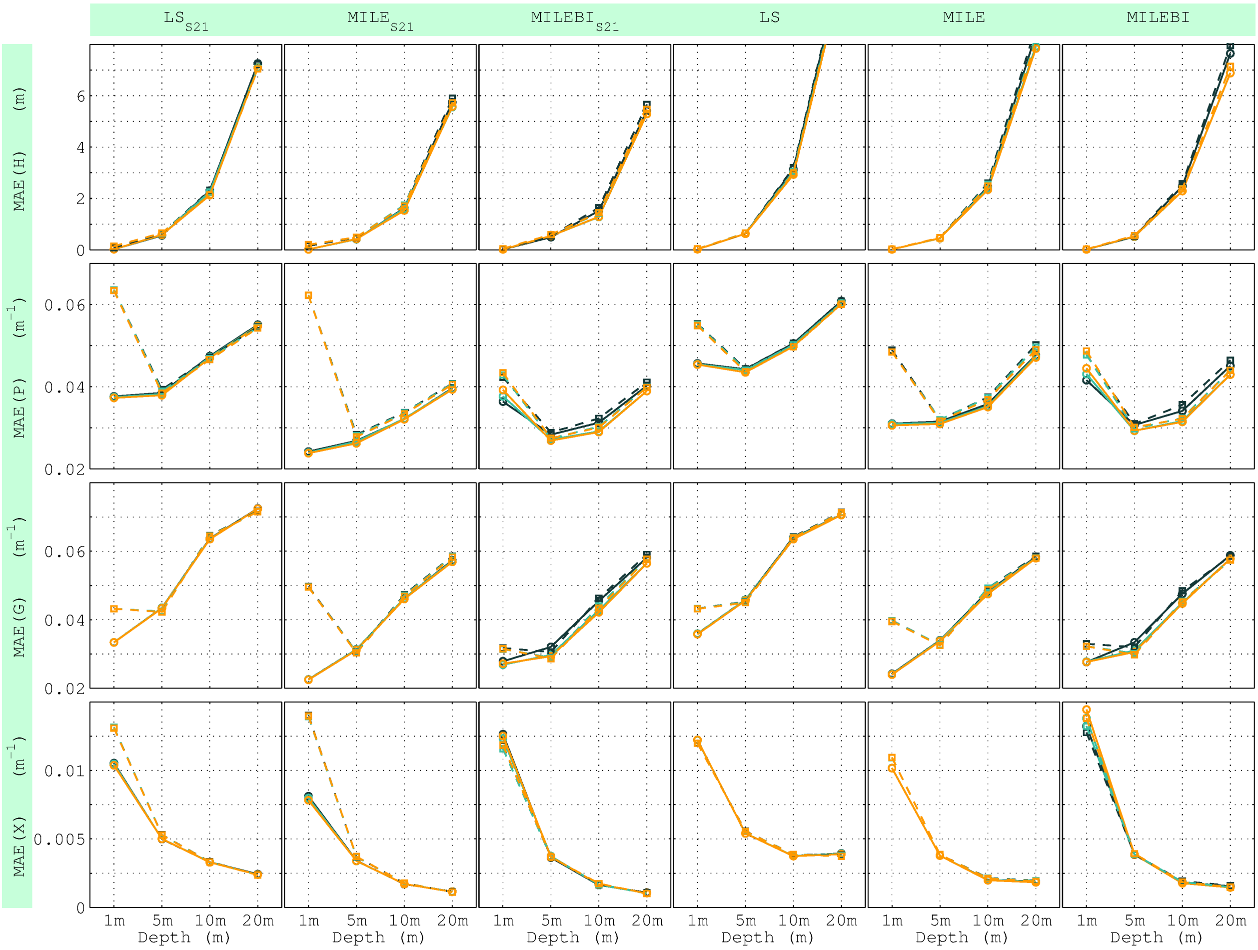

Figure 6: H, P, G and X (rows 1-4, resp.) estimation results obtained by applying the six methods (columns 1-6, resp.) presented in Table 1 to the synthetic data simulated using either Eq. (10) (solid lines) or Eq. (12) (dashed lines). Black, turquoise and orange lines respectively correspond to the use of $n = 0$, 1 and 2% for averaging the best bottom pairs.

For example, at 10 m (first data set, $n = 0\%$), the P (resp., X) retrieval error decreases by about 30% (resp., 48%) when using one of these four methods instead of LS$_{S21}$ or LS. While both data sets lead to similar results for $H \geq 5$ m, strong differences appear for $H = 1$ m. When using the first data set, MILE-based methods offer the best performances for P and G, followed by MILEBI- and LS-based methods. In the case of X, MILE$_{S21}$ and MILE still perform better, followed by LS- and MILEBI-based methods. However, the errors obtained with MILE- and LS-based methods increase when using the second data set. This increase is stronger (1) when the bottom mixture coefficients are constrained to sum to one (e.g., for P estimation, the MAEs obtained with LS$_{S21}$ and LS increase by 70 and 21% resp.), and (2) in the cases of MILE-based methods as compared to LS-based methods (e.g., for X estimation, the MAEs obtained with LS$_{S21}$ and MILE$_{S21}$ increase by 26 and 78% resp.).

On the other hand, MILEBI-based methods offer more similar results over both data sets, $\text{MILEBI}_{\text{S21}}$ generally performing better than the other methods for these three parameters when using the second data set.
Using $n = 0$, 1 or 2% for averaging the best bottom pairs does not significantly change the H, P, G and X inversion results for LS- and MILE-based methods. For $\text{MILEBI}_{\text{S21}}$ and MILEBI, increasing the value of n generally slightly degrades the estimation accuracy at 1 m (e.g., for P estimation, the MAE obtained with $\text{MILEBI}_{\text{S21}}$ increases by 7% when taking $n = 2\%$ as compared to $n = 0\%$). However, the performances generally improve for $H \geq 5$ m when taking either $n = 1$ or 2%. For example, at 10 m (first data set) and for both n values, the MAE obtained with $\text{MILEBI}_{\text{S21}}$ decreases by 15% for H and 7% for P.

The bottom estimation results show similar trends for every benthic class, method, data set and n value, i.e., the error increases with depth (Fig. 7). For $H \leq 5$ m, the easiest class to be retrieved is generally sand, followed by brown algae, seagrasses/green algae and oyster bags. For deeper waters, it is more difficult to note any clear trend among methods and benthic classes. Similarly to depth and water clarity parameters, MILE-based methods provide equal or better performances than LS-based methods for $H \leq 5$ m (e.g., for the sand coefficient, the MAEs obtained with LS_{S21} and MILE_{S21} at 5 m are 0.13 and 0.09 resp.). It is worth noting that, despite the additional bottom intra-class variability present in the second data set, the performances of MILEBI-based methods generally remain comparable to those of MILE-based methods. Also, it can be seen that applying the sum-to-one constraint significantly improves the retrieval for every method, especially for $H \geq 5$ m. For example, for the oyster bag coefficient, the MAE obtained with $\text{MILEBI}_{\text{S21}}$ at 5 m ($n = 0\%$) increases by 38% when relaxing the sum-to-one constraint.
Averaging over several bottom pairs instead of retaining only the best one generally has a positive effect for every method and $H \geq 10$ m (or even for $H \geq 5$ m in the cases of LS and MILE). For such optically deep waters, taking $n = 2\%$ and, to a lesser extent, $n = 1\%$, provides equal or better performances than taking $n = 0\%$ in most cases. For example, for the sand coefficient, the MAE obtained with LS_{S21} at 10 m decreases by 13% when taking $n = 2\%$ as compared to $n = 0\%$. For shallower waters, this averaging does not significantly

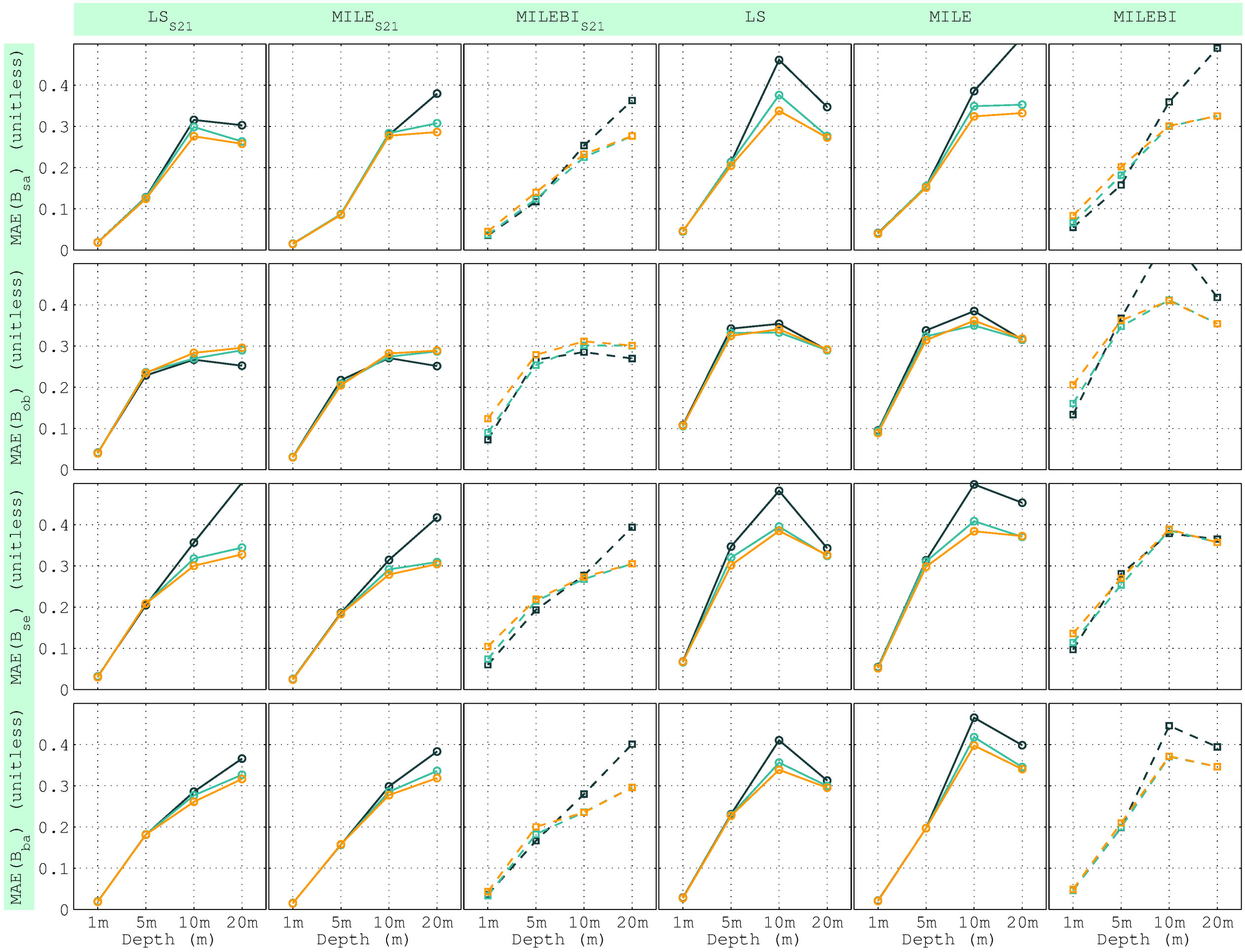

Figure 7: Bottom estimation results obtained by applying the six methods (columns 1-6, resp.) presented in Table 1 to the synthetic data simulated using either Eq. (10) (solid lines) or Eq. (12) (dashed lines). Black, turquoise and orange lines respectively correspond to the use of $n = 0$, 1 and 2% for averaging the best bottom pairs. B_{sa}, B_{ob}, B_{se} and B_{ba} (rows 1-4, resp.) refer to the coefficients of sand, oyster bag, seagrass/green alga and brown alga spectra, resp..

change the retrieval accuracy for LS- and MILE-based methods. However, taking $n = 2\%$, and, to a lesser extent, $n = 1\%$, slightly degrades the MILEBI$_{S21}$ and MILEBI bottom estimation results. In the following results, n is therefore set to 1% as this value offers a good compromise between optically shallow and deep waters for the six methods.

4.3. Estimation results obtained with the airborne data

Similarly to simulations, for every method, the H estimation error increases with depth as a result of a progressive H underestimation and an increasing estimation variance (Fig. 8). This underestimation occurs for shallower waters in the cases of LS-based methods as compared to MILE- and MILEBI-based methods. Unlike for simulations, the sum-to-one constraint leads to poorer performances for every method. MILEBI provides the highest overall

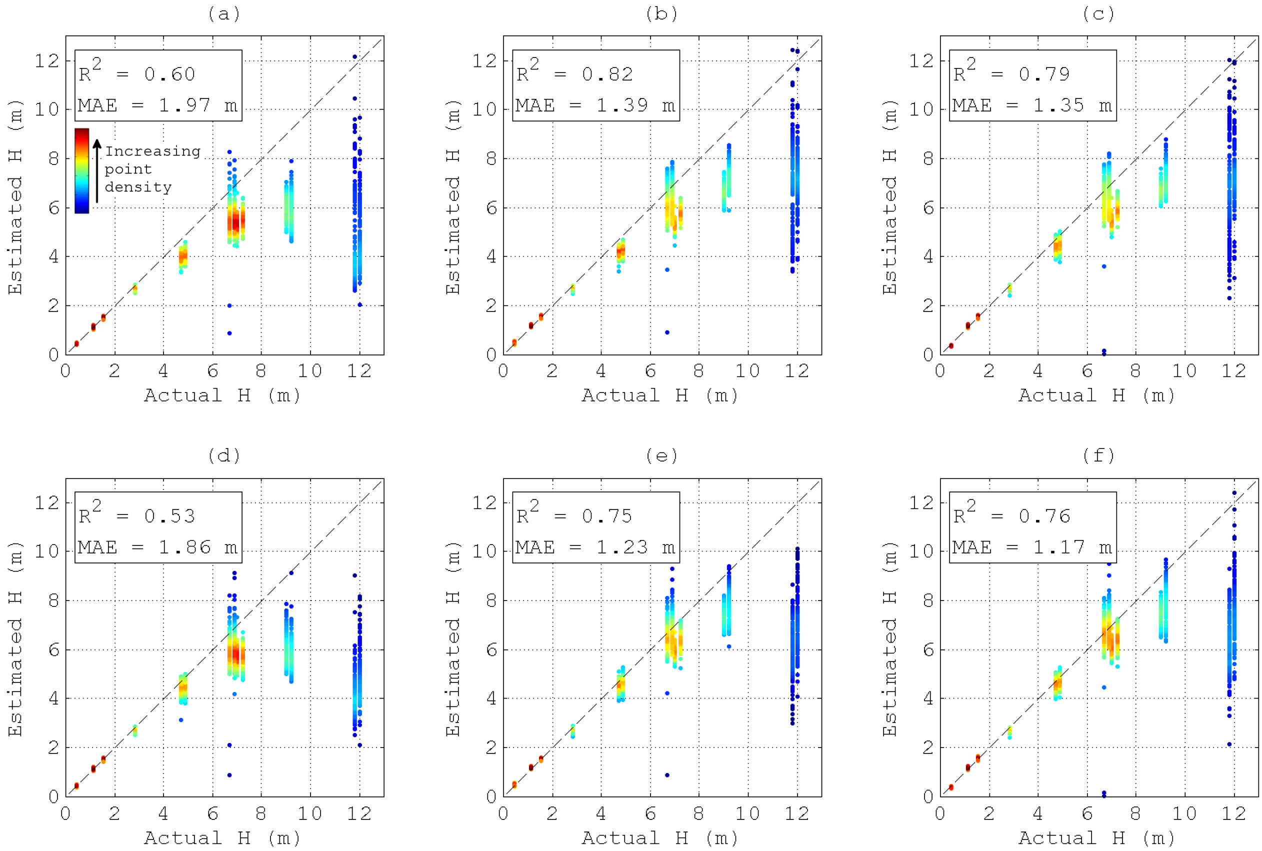

Figure 8: Depth estimation results obtained from airborne data ($n = 1\%$): (a) $\mathrm{LS_{S21}}$, (b) $\mathrm{MILE_{S21}}$, (c) $\mathrm{MILEBI_{S21}}$, (d) LS, (e) MILE and (f) MILEBI.

accuracy (MAE = 1.17 m), followed by MILE (MAE = 1.23 m), $\mathrm{MILEBI_{S21}}$ (MAE = 1.35 m) and $\mathrm{MILE_{S21}}$ (MAE = 1.39 m). On the other hand, $\mathrm{LS_{S21}}$ and LS obtain significantly higher errors, with MAEs of 1.97 and 1.86 m respectively. Note that Fig. 8 suggests that the water tends to be quasi-optically deep for $H > 10$ m, thus potentially making the comparison misleading. When removing the samples corresponding to $H > 10$ m, the MAEs become 0.49, 0.53, 0.75 and 0.81 m for MILEBI, MILE, $\mathrm{MILEBI_{S21}}$ and $\mathrm{MILE_{S21}}$, respectively. On the other hand, LS and $\mathrm{LS_{S21}}$ still obtain poorer performances, with MAEs of 0.92 and 1.12 m, respectively. These MAEs are about twice as high as those obtained using MILEBI and MILE.

Similar observations are made from the P inversion results (Fig. 9), i.e., (1) MILE- and MILEBI-based methods perform better than LS-methods, and (2) relaxing the sum-to-one constraint improves the estimation accuracy. MILEBI and MILE still provide the best performances with MAE ≈ 0.016 m^{-1}, while $\mathrm{LS_{S21}}$ and LS lead to MAE ≈ 0.027 m^{-1}.

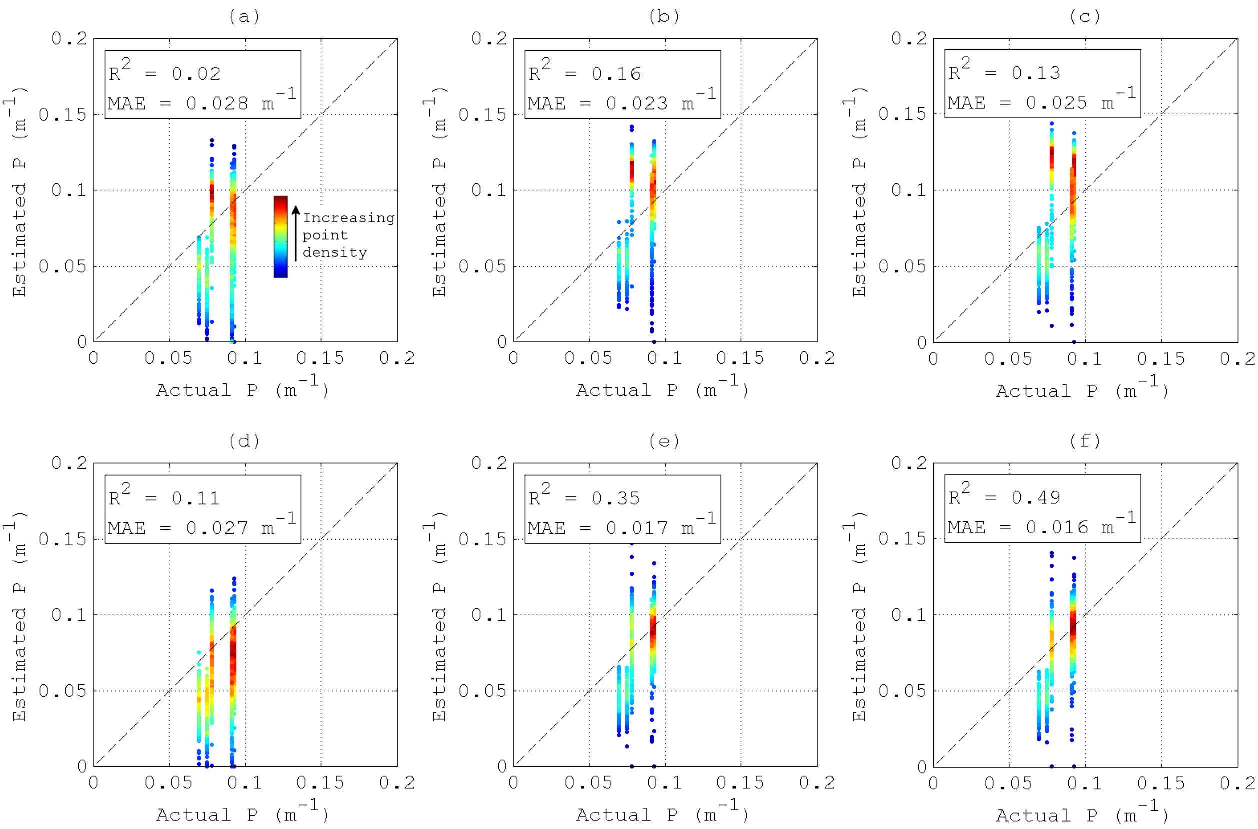

Figure 9: P estimation results obtained from airborne data ($n = 1\%$): (a) $\mathrm{LS}_{\mathrm{S21}}$, (b) $\mathrm{MILE}_{\mathrm{S21}}$, (c) $\mathrm{MILEBI}_{\mathrm{S21}}$, (d) LS, (e) MILE and (f) MILEBI.

The bottom estimation results obtained from the 14 areas of known depth (Fig. 10) show the same pattern for every method, i.e., (1) the sandy-bottom cover is accurately retrieved in shallow waters, and (2) the estimated sand coefficient decreases as depth increases, which is compensated for by increasing coefficients of darker substrates. This decrease occurs for shallower waters (i.e., for $H \geq 4.70$ m) for the three methods that do not constrain the sum to one, i.e., LS, MILE and MILEBI. For example, for these methods and $H \geq 4.70$ m, the estimated sand coefficient generally does not exceed 0.5, while the estimated brown alga coefficient is mostly close to 1.5. On the other hand, $\mathrm{LS}_{\mathrm{S21}}$, $\mathrm{MILE}_{\mathrm{S21}}$ and $\mathrm{MILEBI}_{\mathrm{S21}}$ generally lead to reasonable estimates of bottom cover until around 9.00 m, the best performances being obtained using $\mathrm{MILE}_{\mathrm{S21}}$ with a minimum estimated sand coefficient of 0.6.

In Fig. 11, the same concise and qualitative RGB representation as Petit et al. (2017) is adopted to show the estimated spatial distributions of the four investigated substrates based on the image presented in Fig. 3. Beforehand, for each pixel, the four estimated bottom

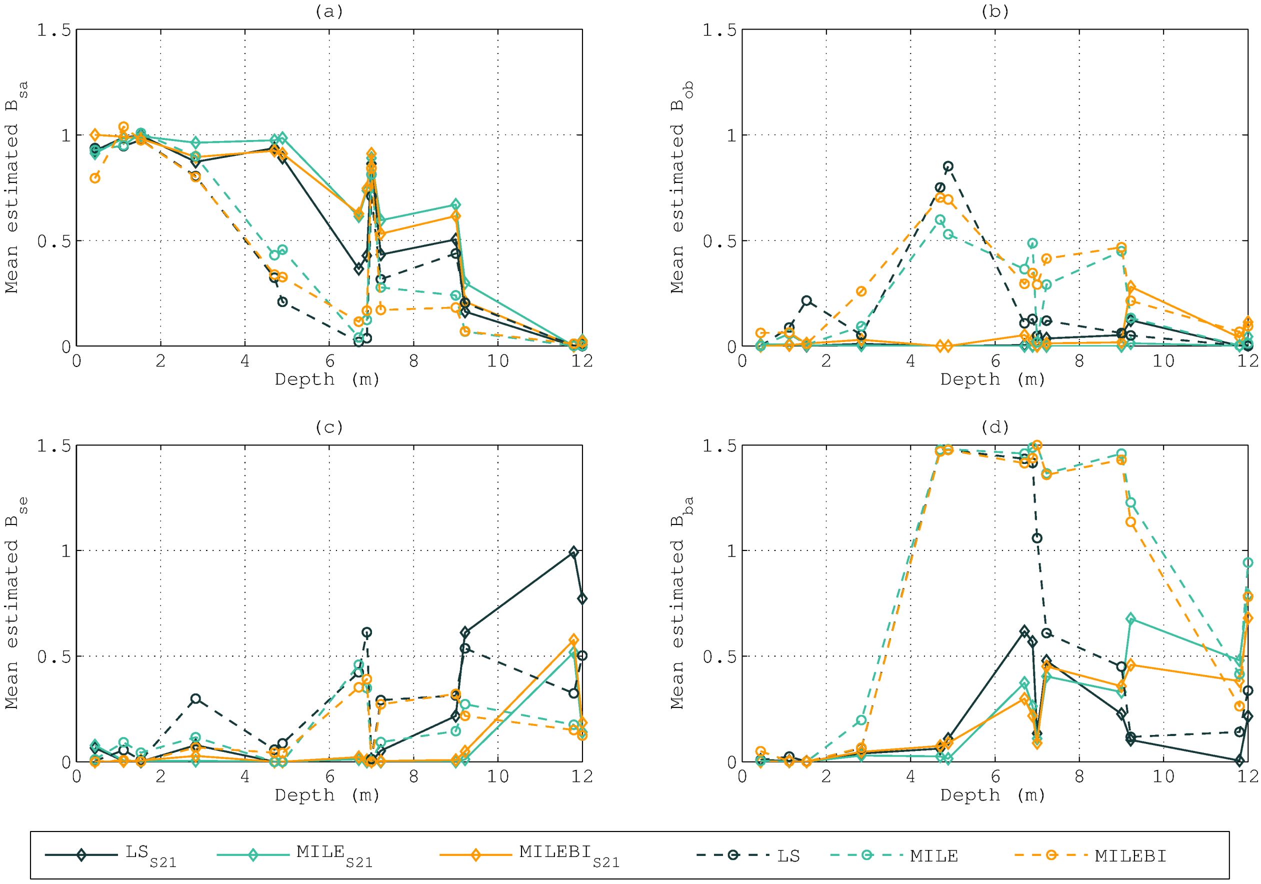

Figure 10: Mean estimated coefficients for (a) sand (B_{sa}), (b) oyster bags (B_{ob}), (c) seagrasses/green algae (B_{se}) and (d) brown algae (B_{ba}) for the 14 sandy-bottom areas ($n = 1\%$).

coefficients were normalized by their sum (that obviously equals one for LS$_{\text{S21}}$, MILE$_{\text{S21}}$ and MILEBI$_{\text{S21}}$) so that the obtained normalized coefficients were closer to the actual fractional covers (if we assume that the effect of intra-class variability is lower than that of fractional cover), which facilitates the comparison of the six methods. This allows representing (1) the distributions of oyster bags, seagrasses/green algae and brown algae through the blue, green and red channels of the color composite image, resp., and (2) the distribution of sand through the absence of blue, green and red, i.e., through the pixel darkness.

The large sandy-bottom area is accurately retrieved by LS$_{\text{S21}}$, MILE$_{\text{S21}}$ and MILEBI$_{\text{S21}}$, the LS$_{\text{S21}}$ map being slightly noisier than the other two, e.g., in the deeper (upper right) part of the image. Except in the shallower (left-hand) part of the image for MILEBI, relaxing the sum-to-one constraint leads to poorer results in the main sandy area. Indeed, even if LS, MILE and MILEBI retrieve some sand, they greatly overestimate the presence of seagrasses/green algae, brown algae and oyster bags respectively.

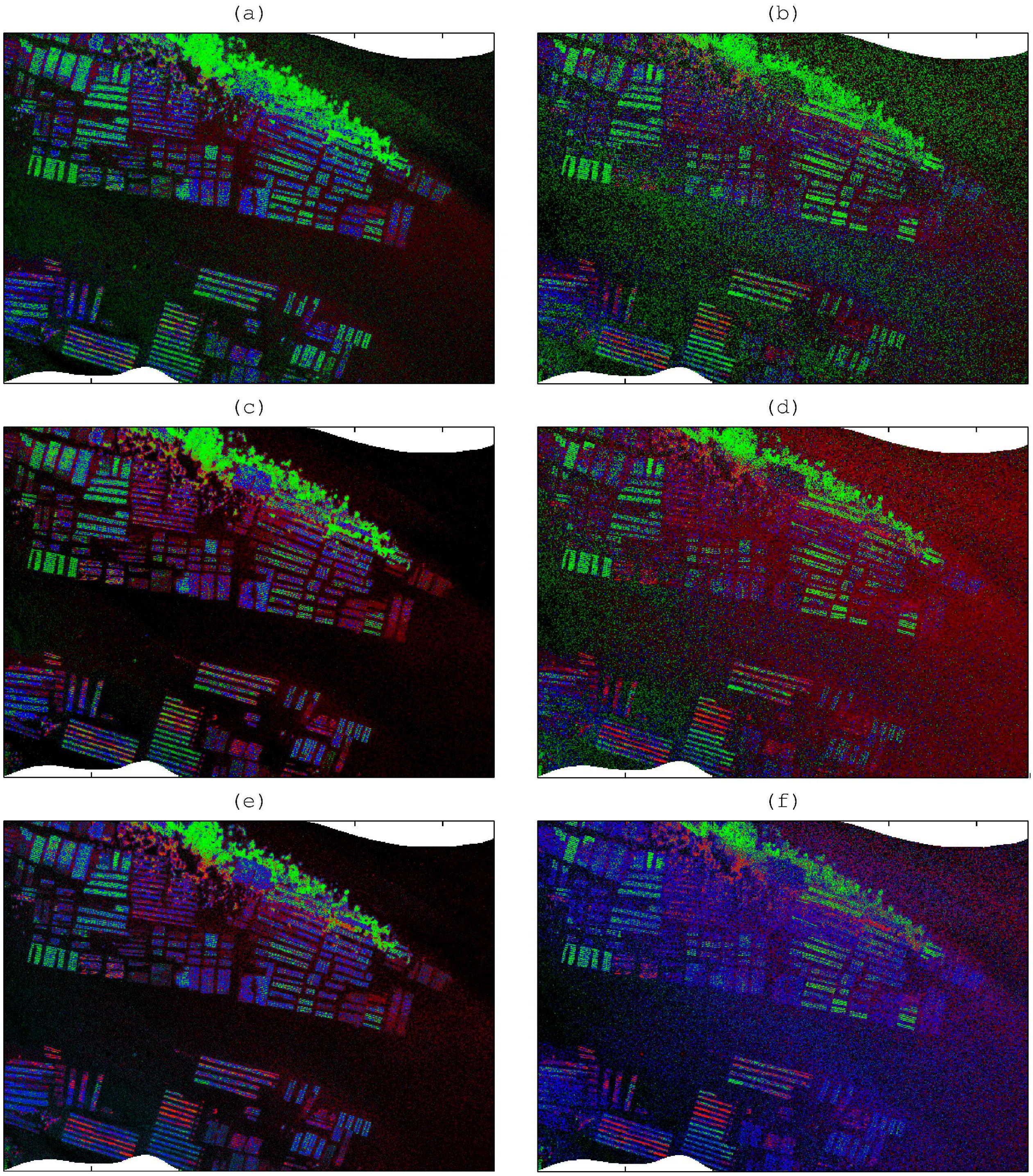

Figure 11: Color composite images showing the estimated spatial distributions of the four investigated substrates based on the image presented in Fig. 3: (a) $\mathrm{LS_{S21}}$, (b) LS, (c) $\mathrm{MILE_{S21}}$, (d) MILE, (e) $\mathrm{MILEBI_{S21}}$, and (f) MILEBI ($n = 1\%$). The normalized estimated coefficients of oyster bags, seagrasses/green algae and brown algae are respectively coded by the blue, green and red channels. The normalized estimated sand coefficient is coded by the pixel darkness (i.e., the absence of red, green and blue).

Overall, the six methods accurately retrieve the seagrass meadow. Some confusions with brown algae however occur in the lower and shallower part of the meadow when using MILE, MILEBI, and to a lesser extent, $\text{MILEBI}_{\text{S21}}$, MILE_{S21} and LS_{S21}.

Similarly to what is observed with simulations, the retrieval of oyster bag distribution is generally less accurate. The results are seemingly more consistent with $\text{MILEBI}_{\text{S21}}$ and MILEBI, as both methods obtain higher and more homogeneously-distributed oyster bag coefficients over oyster racks as compared to the other methods. Note that only the MILE_{S21} and MILEBI-based methods can reliably detect the deepest oyster racks located within the seagrass meadow. On the other hand, LS_{S21} obtains a spatially-inconsistent mixture of oyster bags and seagrasses, while LS and MILE retrieve a sand-dominated bottom.

It is worth noting that the brown algae retrieved by $\text{MILEBI}_{\text{S21}}$ over some oyster racks in the lower left part of the image are more sparsely detected by MILE_{S21} and almost not detected by LS_{S21}. These brown algae are, however, consistently retrieved by the three methods with relaxed sum-to-one constraint.

4.4. Discussion of estimation performances

4.4.1. General considerations

By definition, a bio-optical model is only a model, which means that various sources of error may make it deviate from the observation. Given the number of potential sources (e.g., environmental noise or bottom intra-class variability), the difficulty to properly take them into account (e.g., skyglint) and the low water-leaving radiance, it seems quite challenging to include them explicitly within the modeling and to estimate the corresponding additional parameters during the inversion process. Yet, the results presented in Fig. 5 show that such variability may make the shallow water reflectance strongly differ from the bio-optical model. As a result, it may significantly decrease the estimation accuracy as obtained using the classical LS method, since the latter tries to perfectly match the model with the observation. Alternatively, we propose to include these deviations within a probabilistic forward model of shallow water reflectance variability, thus assuming that they all can be described through an additive zero-mean multivariate Gaussian noise that is fully determined by its spectral covariance matrix. The MILE- and MILEBI-based inversion methods are derived from such probabilistic modeling, and the results derived from simulated and airborne data show that they all succeed in decreasing the detrimental influence of environmental noise as compared to LS-based methods, especially in optically deep waters. In addition, MILEBI-based methods decrease the influence of bottom intra-class variability, especially in very optically shallow waters.

4.4.2. Common trends in method performances

Overall, the results obtained with simulated and airborne data show similar trends and are consistent with expectations for every method. For example, depth and benthic cover estimations become less accurate as depth increases due to the decreasing bottom influence on subsurface reflectance (Fig. 6, Fig. 7, Fig. 8 and Fig. 10). The retrievals of water clarity parameters differ between absorbing (P and G) and scattering (X) components that respectively decrease and increase the subsurface reflectance (Fig. 6). For P and G, the depth of minimum error is the one that offers the best compromise between (1) maximizing the subsurface reflectance so that there is more contrast between absorbing and non-absorbing regions (which facilitates the retrieval), and (2) minimizing the influence of bottom variability on subsurface reflectance. For X, the error is minimum in optically deep waters, where the bottom does not affect the subsurface reflectance.

4.4.3. Influence of averaging the best bottom pairs

Due to the ill-posedness of the inversion problem (resulting in compensations between model parameters) or to potential deviations between the measured reflectance and the model, the actual bottom pair may not be the one that leads to the lowest cost function value. In simulations, the inversion is particularly ill-posed for quasi-optically deep waters, where (1) H and coefficients of dark bottoms often tend to compensate, and (2) all the dark benthic classes (e.g., seagrasses/green algae and brown algae) nearly have the same effect on subsurface reflectance (Fig. 6 and Fig. 7). In this case, selecting a particular dark substrate in the bottom spectral library instead of another dark substrate is not strongly justified, given the different sources of error between the observed and modeled reflectances that actually make both substrates equally likely. The results (Fig. 6 and Fig. 7) demonstrate that, alternatively, taking the average of multiple best bottom pairs (if sufficiently close to the best pair) can decrease the ill-posedness influence and increase the overall retrieval accuracy, acting as a regularization step. Testing the effect of the n value (that directly controls the number of best pairs to be averaged), we show that a high n value (even greater than 2%) can be chosen for optically deep waters, where a reasonable aim is only to discriminate among bright and dark substrates. In very shallow waters, a too large n value may, however, increase the confusion between classes, therefore making the value of 1% a good compromise for our data. Of course, this value should be reassessed for each data set, as it is expected to depend on, e.g., the environmental noise and/or the benthic classes encountered on the study site.

4.4.4. Influence of sum-to-one constraint

The results show that the sum-to-one constraint always leads to better inversion results if the shallow water reflectance model is perfect (e.g., when applying LS- and MILE-based methods to the first data set, or MILEBI-based methods to the second data set), because reducing the number of parameters to be retrieved reduces the estimation uncertainty. In practice, the observation may, however, deviate from the model. These deviations may be caused either by the observation, e.g., in the case of imperfect preprocessing of at-sensor radiance (e.g., atmospheric and sea surface corrections), or by the model, e.g., in the case of imperfect bio-optical modeling. In this study, such deviations are present when considering airborne remote-sensing data or when applying LS- and MILE-based (resp., MILEBI-based) methods to the second (resp., first) synthetic data set. In these cases, relaxing the sum-to-one constraint adds a degree of freedom, which enables unmodeled (or mismodeled) variability to be compensated for by misestimation of bottom cover rather than by misestimation of depth and/or water clarity parameters. This is demonstrated by the results obtained with airborne data, since (1) Fig. 8 and Fig. 9 show that LS, MILE and MILEBI better retrieve H and P as compared to LS_{S21}, MILE_{S21} and MILEBI_{S21} resp. (note that this is consistent with the results of Petit et al. (2017) in the case of LS), and (2) Fig. 10 shows that LS_{S21}, MILE_{S21} and MILEBI_{S21} provide better bottom retrievals than LS, MILE and MILEBI resp.. However, relaxing the sum-to-one constraint does not always degrade the bottom retrieval: indeed, if the bottom intra-class variabilities affect the subsurface reflectance (i.e., mostly for low optical depths, see Fig. 5), allowing both benthic reflectances in Eq. (8) to vary in a multiplicative way enables LS and MILE to better capture this intra-class variability and improve the overall performances.

MILEBI_{S21} thus appears as an interesting alternative to LS- and MILE-based methods, because (1) it takes into account potentially complex (i.e., not only multiplicative) bottom intra-class variabilities through their associated covariance matrix, and (2) it limits the problem ill-posedness as it does not require any additional parameter to be estimated. The benthic covers derived from airborne data (Fig. 11) illustrate this dual improvement, as MILEBI_{S21} not only provides accurate performances in the deepest sandy-bottom areas similarly to LS_{S21} and MILE_{S21}, but also retrieves the presence of brown algae over oyster racks in shallower waters, similarly to LS, MILE and MILEBI.

4.4.5. Robustness of inversion methods

All LS-, MILE- and MILEBI-based methods require some prior knowledge on the considered scene, this knowledge concerning either the mean endmember reflectances or the covariance matrices. However, obtaining an accurate prior knowledge may be difficult, which requires investigating how such errors can affect the method performances.

It should first be noted that obtaining an accurate estimate of the environmental noise matrix (as necessary for MILE- and MILEBI-based methods) is usually not problematic, since it only necessitates finding a homogeneous area in the image. This may easily be done using the methodology proposed by Wettle et al. (2004), and areas of optically deep water are ideal to perform this estimation. Using this matrix for inversion allows MILE-based methods to greatly improve the retrieval of depth and water clarity parameters in sufficiently deep waters as compared to LS-based methods (Fig. 6, Fig. 8 and Fig. 9). It also improves the remote sensing of shallow waters if Eq. (8) accurately models the actual bottom reflectance. However, if the latter cannot accurately be modeled by Eq. (8) (e.g., due to complex intra-class variabilities or poorly-known mean endmember reflectances) while having a strong effect on subsurface reflectance (i.e., in very optically shallow waters), the performances of MILE-based methods may decrease more strongly than those of LS-based methods (Fig. 6). In such cases, MILE is shown to better estimate depth and water clarity parameters than LS, LS_{S21} and MILE_{S21} (Fig. 6, Fig. 8 and Fig. 9), especially because relaxing the sum-to-one constraint reduces the detrimental influence of bottom intra-class variability. Alternatively, MILEBI and MILEBI_{S21} allow the modeled endmember spectra to vary around their mean through the use of bottom intra-class covariance matrices. Both methods are thus less affected by an imperfect knowledge of endmember reflectances. This aspect is one of the primary advantages of these methods as compared to LS- and MILE-based methods, and may be of tremendous importance when mapping poorly-known shallow water environments, for which the use of a single mean reflectance spectrum for each benthic class may seem unrealistic.

However, obtaining accurate estimates of bottom covariance matrices may sometimes be difficult since, similarly to the mean endmember reflectances used by the six tested methods, and as emphasized in Section 2.4, these matrices are estimated from a limited number of spectra that may not be fully representative of the variability encountered in the whole study area. That said, the results obtained with simulated data (Fig. 6) suggest that accurate knowledge of these matrices may only be necessary for very optically shallow waters, as MILE- and MILEBI-based obtain nearly the same results over both data sets beyond 5 m. As the optical depth increases, the water attenuation and environmental noise smooths the spectral details present in bottom covariance matrices (Fig. 5), so rough estimates become sufficient to take this variability into account. For very optically shallow waters, unlike LS- and MILE-based methods, MILEBI-based methods show similar performances for both synthetic data sets (Fig. 6), although the first data set is generated using zero covariance matrices that strongly differ from those used in MILEBI_{S21} and MILEBI. This important result demonstrates the robustness of these two methods against imperfect knowledge of bottom covariance matrices, which may have important implications for their implementation at larger scales (e.g., global scale).

5. Conclusions and perspectives

In this study, we propose a realistic probabilistic model of shallow water reflectance variability as well as two associated inversion methods, denoted MILE and MILEBI. As compared to classical least-squares fitting, these methods improve the remote sensing of shallow waters by utilizing specific parameterizations of the spectral covariance matrix. MILE and MILEBI not only constrain model inversion based on the off-diagonal terms of covariance matrices, but also allow the measured data to differ from the model by giving the less reliable wavebands lower weights in the cost function. For MILE, these wavebands correspond to the domains where the environmental noise is the strongest. For MILEBI, the less reliable wavebands not only correspond to the domains of strong environmental noise, but also to the domains where the bottom intra-class variability is the highest. To our knowledge, MILEBI is one of the first shallow water remote-sensing methods that explicitly take into account the inherent variability of each benthic class without adding any multiplicative parameter to be estimated during the inversion process (the bottom covariance matrices, however, need to be estimated beforehand, similarly to the mean endmember reflectances).

Based on simulated and airborne data, we show that these specific covariance parameterizations enable MILE and MILEBI to generally perform better than LS. Further, studying the influence of constraining bottom mixture coefficients to sum to one shows that this constraint provides better inversion results if the reflectance model reliably describes the observation. In the presence of unmodeled (or mismodeled) variability in the remote-sensing data (e.g., due to bottom intra-class variability, imperfect atmospheric correction or bio-optical modeling, etc), relaxing this constraint may decrease the detrimental influences of these deviations, however at the cost of an increasingly noisy bottom retrieval as the optical depth increases. In practice, as there are always some slight deviations between measured and simulated data, these results thus suggest that most inversion methods cannot accurately retrieve all the targeted parameters at the same time, and that applying different constraints during the inversion will lead these deviations to affect the estimation of other unconstrained parameters. That said, the sum-to-one constrained version of MILEBI combines the advantage of limiting the number of parameters to be estimated (thus reducing the problem ill-posedness) with that of allowing the observation to differ from the model. This dual aspect makes this method promising to remotely sense complex shallow water environments.

Future studies would certainly benefit from the probabilistic forward model of shallow water reflectance variability presented in Eq. (12) so as to generate more realistic data sets than those usually generated using Eq. (10). This model could also be combined with other bottom reflectance models (e.g., a single substrate model, linear models including more than two substrates or even non-linear mixing models) in order to further refine the modeling of bottom reflectance and improve the inversion performances. This may be important for more accurately simulating the response of very shallow waters, for which an increase in bottom modeling complexity significantly affects the measured subsurface reflectance.

As far as the inversion is concerned, perspectives include refining the initialization part, that may be critical for MILE methods in very shallow waters (results not shown). Optimizing the construction of the LUT used for initialization (size, parameter distributions, etc) is likely to speed up the inversion while keeping similar estimation performances. Alternatively, the Mahalanobis distance used in MILE could easily be used as a metric within a LUT-based inversion approach such as ALLUT (Hedley et al., 2009) in order to further speed up the inversion process or to avoid local minima. Note that the approach recently proposed by Jay & Guillaume (2016) could also be implemented to regularize the inversion by introducing prior knowledge on targeted parameters.

Ultimately, an important perspective is the assessment of MILE and MILEBI performances for shallow water remote sensing at the global scale, e.g., in the context of the forthcoming "Environmental Mapping and Analysis Program" mission (Guanter et al., 2015). For this purpose, besides properly estimating the environmental noise on the image itself, a generic library of bottom mean reflectance spectra will be necessary to parameterize the total benthic reflectance. This library may be built from a comprehensive spectral database gathering all the expected bottom classes in the considered study site. For example, the 12-class database presented by Hochberg et al. (2003) could be of great help for coral reef remote sensing. This database could also be used to build an associated generic library of intra-class covariance matrices to implement MILEBI. As shown by Hochberg et al. (2003) in Fig. 3, the intra-class variability at the global scale is such that using a single mean reflectance spectrum for each bottom class to map this class across different areas worldwide seems to be highly unrealistic. MILEBI thus offers an interesting alternative to LS and MILE to take such variability into account in a more accurate manner. In particular, given the high intra-class variabilities presented by Hochberg et al. (2003) and the significant overlaps between these classes, MILEBI may greatly improve the remote sensing of coral reefs.

Acknowledgments

This work was supported by the French Defense Procurement Agency (DGA) with the reference ANR-15-ASTR-0019 (HypFoM). We are also grateful to Actimar, that carried out the field measurement campaign (Smet et al., 2010) within the exploratory research and innovation project “HypLitt”, funded by the French Defence Agency (DGA). ActiMar was a company based in Brest, France, and specialized in operational oceanography and high-resolution remote sensing (www.actimar.fr). These activities are now carried out by the Hytech Imaging company (www.hytech-imaging.fr). Many thanks to Marc Lennon, John D. Hedley and the anonymous reviewers for their valuable comments. Note, finally, that the codes of the proposed methods are available on demand.

References

AFNOR (December 1999). *Qualité de l’eau, Dosage de la chlorophylle a et d’un indice phopigments - Méthode par spectrométrie d’absorption moléculaire.*

Albert, A., & Gege, P. (2006). Inversion of irradiance and remote sensing reflectance in shallow water between 400 and 800 nm for calculations of water and bottom properties. *Applied Optics*, *45*, 2331–2343.

Albert, A., & Mobley, C. (2003). An analytical model for subsurface irradiance and remote sensing reflectance in deep and shallow case-2 waters. *Optics Express*, *11*, 2873–2890.

Andréfouët, S., Berkelmans, R., Odriozola, L., Done, T., Oliver, J., & Müller-Karger, F. (2002). Choosing the appropriate spatial resolution for monitoring coral bleaching events using remote sensing. *Coral Reefs*, *21*, 147–154.

Aschbacher, J., & Milagro-Pérez, M. P. (2012). The European Earth monitoring (GMES) programme: Status and perspectives. *Remote Sensing of Environment*, *120*, 3–8.

Botha, E. J., Brando, V. E., Anstee, J. M., Dekker, A. G., & Sagar, S. (2013). Increased spectral resolution enhances coral detection under varying water conditions. *Remote Sensing of Environment*, *131*, 247–261.

Brando, V., Anstee, J., Wettle, M., Dekker, A., Phinn, S., & Roelfsema, C. (2009). A physics based retrieval and quality assessment of bathymetry from suboptimal hyperspectral data. *Remote Sensing of Environment*, *113*, 755–770.

Brando, V. E., & Dekker, A. G. (2003). Satellite hyperspectral remote sensing for estimating estuarine and coastal water quality. *Geoscience and Remote Sensing, IEEE Transactions on*, *41*, 1378–1387.

Buiteveld, H., Hakvoort, J., & Donze, M. (1994). Optical properties of pure water. In *Ocean Optics XII* (pp. 174–183). International Society for Optics and Photonics.

Darvishzadeh, R., Atzberger, C., Skidmore, A., & Schlerf, M. (2011). Mapping grassland leaf area index with airborne hyperspectral imagery: A comparison study of statistical approaches and inversion of radiative transfer models. *ISPRS Journal of Photogrammetry and Remote Sensing*, *66*, 894–906.

Dekker, A. G., Phinn, S. R., Anstee, J., Bissett, P., Brando, V. E., Casey, B., Fearns, P., Hedley, J., Klonowski, W., Lee, Z. P. et al. (2011). Intercomparison of shallow water bathymetry, hydro-optics, and benthos mapping techniques in Australian and Caribbean coastal environments. *Limnology and Oceanography: Methods*, *9*, 396–425.

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.

Fearns, P., Klonowski, W., Babcock, R., England, P., & Phillips, J. (2011). Shallow water substrate mapping using hyperspectral remote sensing. *Continental Shelf Research*, *31*, 1249–1259.

Garcia, R. A., Fearns, P. R., & McKinna, L. I. (2014a). Detecting trend and seasonal changes in bathymetry derived from HICO imagery: A case study of shark bay, western australia. *Remote Sensing of Environment*, *147*, 186–205.

Garcia, R. A., Hedley, J. D., Tin, H. C., & Fearns, P. R. (2015). A method to analyze the potential of optical remote sensing for benthic habitat mapping. *Remote Sensing*, *7*, 13157–13189.

Garcia, R. A., McKinna, L. I., Hedley, J. D., & Fearns, P. R. (2014b). Improving the optimization solution for a semi-analytical shallow water inversion model in the presence of spectrally correlated noise. *Limnol. Oceanogr. Methods*, *12*, 651–669.

Giardino, C., Candiani, G., Bresciani, M., Lee, Z., Gagliano, S., & Pepe, M. (2012). BOMBER: A tool for estimating water quality and bottom properties from remote sensing images. *Computers & Geosciences*, *45*, 313–318.

Gillis, D. B., Bowles, J. H., & Moses, W. J. (2013). Improving the retrieval of water inherent optical properties in noisy hyperspectral data through statistical modeling. *Optics express*, *21*, 21306–21316.

Goodman, J., & Ustin, S. L. (2007). Classification of benthic composition in a coral reef environment using spectral unmixing. *Journal of Applied Remote Sensing*, *1*, 011501–011501.

Goodman, J. A., Lee, Z., & Ustin, S. L. (2008). Influence of atmospheric and sea-surface corrections on retrieval of bottom depth and reflectance using a semi-analytical model: a case study in Kaneohe Bay, Hawaii. *Applied Optics*, *47*, F1–F11.

Guanter, L., Kaufmann, H., Segl, K., Foerster, S., Rogass, C., Chabrillat, S., Kuester, T., Hollstein, A., Rossner, G., Chlebek, C. et al. (2015). The EnMAP spaceborne imaging spectroscopy mission for earth observation. *Remote Sensing*, *7*, 8830–8857.

Hedley, J., Harborne, A., & Mumby, P. (2005). Technical note: Simple and robust removal of sun glint for mapping shallow-water benthos. *International Journal of Remote Sensing*, *26*, 2107–2112.

Hedley, J., Roelfsema, C., Koetz, B., & Phinn, S. (2012a). Capability of the Sentinel-2 mission for tropical coral reef mapping and coral bleaching detection. *Remote Sensing of Environment*, *120*, 145–155.

Hedley, J., Roelfsema, C., & Phinn, S. R. (2009). Efficient radiative transfer model inversion for remote sensing applications. *Remote Sensing of Environment*, *113*, 2527–2532.

Hedley, J. D., Roelfsema, C. M., Phinn, S. R., & Mumby, P. J. (2012b). Environmental and sensor limitations in optical remote sensing of coral reefs: Implications for monitoring and sensor design. *Remote Sensing*, *4*, 271–302.

Hochberg, E. J., & Atkinson, M. J. (2003). Capabilities of remote sensors to classify coral, algae, and sand as pure and mixed spectra. *Remote Sensing of Environment*, *85*, 174–189.

Hochberg, E. J., Atkinson, M. J., & Andréfouët, S. (2003). Spectral reflectance of coral reef bottom-types worldwide and implications for coral reef remote sensing. *Remote Sensing of Environment*, *85*, 159–173.

Jay, S., & Guillaume, M. (2011). Estimationofwater column parameters with a maximum likelihood approach. In *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2011 3rd Workshop on* (pp. 1–4). IEEE.

Jay, S., & Guillaume, M. (2014). A novel maximum likelihood based method for mapping depth and water quality from hyperspectral remote-sensing data. *Remote Sensing of Environment*, *147*, 121–132.

Jay, S., & Guillaume, M. (2016). Regularized estimation of bathymetry and water quality using hyperspectral remote sensing. *International Journal of Remote Sensing*, *37*, 263–289.

Jay, S., Guillaume, M., & Blanc-Talon, J. (2012). Underwater target detection with hyperspectral data: Solutions for both known and unknown water quality. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, *5*, 1213–1221.

Jay, S., Maupas, F., Bendoula, R., & Gorretta, N. (2017). Retrieving LAI, chlorophyll and nitrogen contents in sugar beet crops from multi-angular optical remote sensing: comparison of vegetation indices and PROSAIL inversion for field phenotyping. *Field Crops Research*, *210*, 33–46.

Joshi, I., & D'Sa, E. J. (2015). Seasonal variation of colored dissolved organic matter in Barataria Bay, Louisiana, using combined Landsat and field data. *Remote Sensing*, *7*, 12478–12502.

Klonowski, W. M., Fearns, P. R., & Lynch, M. J. (2007). Retrieving key benthic cover types and bathymetry from hyperspectral imagery. *Journal of Applied Remote Sensing*, *1*, 011505–011505.

Knudby, A., Ahmad, S. K., & Ilori, C. (2016). The potential for Landsat-based bathymetry in Canada. *Canadian Journal of Remote Sensing*, *42*, 367–378.

Kutser, T., Dekker, A. G., & Skirving, W. (2003). Modeling spectral discrimination of great barrier reef benthic communities by remote sensing instruments. *Limnology and Oceanography*, *48*, 497–510.

Lee, Z., & Carder, K. L. (2002). Effect of spectral band numbers on the retrieval of water column and bottom properties from ocean color data. *Applied Optics*, *41*, 2191–2201.

Lee, Z., Carder, K. L., Chen, R. F., & Peacock, T. G. (2001). Properties of the water column and bottom derived from Airborne Visible Infrared Imaging Spectrometer (AVIRIS) data. *Journal of Geophysical Research: Oceans*, *106*, 11639–11651.

Lee, Z., Carder, K. L., Mobley, C. D., Steward, R. G., & Patch, J. S. (1998). Hyperspectral remote sensing for shallow waters. i. a semianalytical model. *Applied Optics*, *37*, 6329–6338.

Lee, Z., Carder, K. L., Mobley, C. D., Steward, R. G., & Patch, J. S. (1999). Hyperspectral remote sensing for shallow waters: 2. deriving bottom depths and water properties by optimization. *Applied Optics*, *38*, 3831–3843.

Lee, Z., Weidemann, A., & Arnone, R. (2013). Combined effect of reduced band number and increased bandwidth on shallow water remote sensing: The case of Worldview 2. *IEEE Transactions on Geoscience and Remote Sensing*, *51*, 2577–2586.

Manolakis, D., Marden, D., & Shaw, G. A. (2003). Hyperspectral image processing for automatic target detection applications. *Lincoln laboratory journal*, *14*, 79–116.

Maritorena, S., Morel, A., & Gentili, B. (1994). Diffuse reflectance of oceanic shallow waters: Influence of water depth and bottom albedo. *Limnology and Oceanography*, *39*, 1689–1703.

McKinna, L. I., Fearns, P. R., Weeks, S. J., Werdell, P. J., Reichstetter, M., Franz, B. A., Shea, D. M., & Feldman, G. C. (2015). A semianalytical ocean color inversion algorithm with explicit water column depth and substrate reflectance parameterization. *Journal of Geophysical Research: Oceans*, *120*, 1741–1770.

Melgani, F., & Bruzzone, L. (2004). Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on geoscience and remote sensing*, *42*, 1778–1790.

Mishra, D., Narumalani, S., Rundquist, D., & Lawson, M. (2006). Benthic habitat mapping in tropical marine environments using QuickBird multispectral data. *Photogrammetric Engineering & Remote Sensing*, *72*, 1037–1048.

Mobley, C. D. (1994). *Light and water: radiative transfer in natural waters*. Academic press.

Mobley, C. D., Sundman, L. K., Davis, C. O., Bowles, J. H., Downes, T. V., Leathers, R. A., Montes, M. J., Bissett, W. P., Kohler, D. D., Reid, R. P. et al. (2005). Interpretation of hyperspectral remote-sensing imagery by spectrum matching and look-up tables. *Applied Optics*, *44*, 3576–3592.

Morel, A. (1974). Optical properties of pure water and pure sea water. *Optical aspects of oceanography*, *1*, 1–24.

Mouroulis, P., Van Gorp, B., Green, R. O., Dierssen, H., Wilson, D. W., Eastwood, M., Boardman, J., Gao, B.-C., Cohen, D., Franklin, B. et al. (2014). Portable Remote Imaging Spectrometer coastal ocean sensor: Design, characteristics, and first flight results. *Applied optics*, *53*, 1363–1380.

Palmason, J. A., Benediktsson, J. A., Sveinsson, J. R., & Chanussot, J. (2005). Classification of hyperspectral data from urban areas using morphological preprocessing and independent component analysis. In *Geoscience and Remote Sensing Symposium, 2005. IGARSS'05. Proceedings. 2005 IEEE International* (pp. 4–pp). IEEE volume 1.

Petit, T., Bajjouk, T., Mouquet, P., Rochette, S., Vozel, B., & Delacourt, C. (2017). Hyperspectral remote sensing of coral reefs by semi-analytical model inversion–comparison of different inversion setups. *Remote Sensing of Environment*, *190*, 348–365.

Richter, R. (2012). *Atmospheric / topographic correction for airborne imagery: ATCOR-4 User Guide*. Wessling, Germany: DLR IB 565-02/11.

Smet, S., Sicot, G., & Lennon, M. (2010). *Evaluation des capacités de le télédétection hyperspectrale et développement de méthodes innovantes de traitement d'images pour des applications Défense en zone littorale (HypLitt)*. Technical Report , contrat de recherche DGA 2010 34 0014.

Verrelst, J., Camps-Valls, G., Muñoz-Marí, J., Rivera, J. P., Veroustraete, F., Clevers, J. G., & Moreno, J. (2015). Optical remote sensing and the retrieval of terrestrial vegetation biogeophysical properties–a review. *ISPRS Journal of Photogrammetry and Remote Sensing*, *108*, 273–290.

Wettle, M., Brando, V. E., & Dekker, A. G. (2004). A methodology for retrieval of environmental noise equivalent spectra applied to four hyperion scenes of the same tropical coral reef. *Remote Sensing of Environment*, *93*, 188–197.

Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate research*, *30*, 79–82.