



# MEETC2: Ocean color atmospheric corrections in coastal complex waters using a Bayesian latent class model and potential for the incoming sentinel 3 - OLCI mission

Bertrand Saulquin, Ronan Fablet, Ludovic Bourg, Grégoire Mercier, Odile Fanton d'Andon

## ► To cite this version:

Bertrand Saulquin, Ronan Fablet, Ludovic Bourg, Grégoire Mercier, Odile Fanton d'Andon. MEETC2: Ocean color atmospheric corrections in coastal complex waters using a Bayesian latent class model and potential for the incoming sentinel 3 - OLCI mission. Remote Sensing of Environment, 2016, 172, pp.39 - 49. 10.1016/j.rse.2015.10.035 . hal-01611644

**HAL Id: hal-01611644**

**<https://hal.science/hal-01611644>**

Submitted on 6 Oct 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# MEETC2: Ocean Color Atmospheric corrections in coastal complex waters using a Bayesian latent class model and potential for the incoming Sentinel 3 - OLCI mission.

Bertrand Saulquin (1, 2), Ronan Fablet (2, 3), Ludovic Bourg (1), Grégoire Mercier (2, 3),  
Odile Fanton d'Andon (1).

(1) ACRI-ST, Sophia-Antipolis, 260 route du Pin Montard, BP 234  
06904 Sophia-Antipolis, France

(2) Institut Mines-Telecom, Télécom Bretagne; UMR CNRS 3192 Lab-STICC, Technopôle  
Brest Iroise CS 83818, 29238 Brest, France

(3) Université Européenne de Bretagne, 35000 Rennes, France

Corresponding author: [bertrand.saulquin@acri-st.fr](mailto:bertrand.saulquin@acri-st.fr)

Submission date: 10/06/2015

## ***Abstract***

From top-of-atmosphere (TOA) observations, atmospheric correction for ocean color inversion aims at distinguishing atmosphere and water contributions. From a methodological point of view, our approach relies on a Bayesian inference using Gaussian Mixture Model prior distributions on reference spectra of aerosol and water reflectance. A reference spectrum for the aerosol characterizes the specific signature of the aerosols on the observed aerosol reflectance. A reference

spectrum for the water characterizes the specific signature of chlorophyll-a, suspended particulate matters and colored dissolved organic matters on the observed sea surface reflectance. In our Bayesian inversion scheme, prior distributions of the marine and aerosol variables are corrected using the observed values of covariates, typically acquisition geometry acquisition conditions and pre-estimates of the aerosol and water reflectance in the near-infrared part of the spectrum, to optimize the 25 random initializations for our MEETC2 algorithm.

We evaluate our estimates of the sea surface reflectance from the MERIS TOA observations. Using the MERMAID radiometric in-situ dataset, we obtain significant improvements in the estimation of the sea surface reflectance, especially for the 412, 442, 490 and 510 nm bands, compared with the standard ESA MEGS algorithm and the a state-of-the-art neural network approach (C2R). The mean gain value on the relative error for the 13 bands between 412 and 885 nm is of 57% compared with MEGS algorithm and 10% compared with the C2R. The water leaving reflectances are used in Ocean Color for the estimation of the chl-a concentration, the colored dissolved organic matters absorption and the suspended particulate matters concentration underlying the potential of such approach to improve the standard level 2 products in coastal areas. We further discuss the potential of MEETC2 for the incoming OLCI / Sentinel 3 mission that should be launched in 2015.

Index term: 1) Atmospheric corrections in complex coastal waters. 2) Ocean color. 3) Bayesian inversion. 4) Gaussian Mixture Model.

## **1 Introduction**

The inversion of Ocean Color signal in coastal areas from top-of-atmosphere (TOA) measurements remains a scientific challenge. This is a crucial point for the ocean color community as many governmental policies such as the European Water Framework directive (WFD) rely on

45 estimation of coastal water quality, itself possibly derived from space-based ocean-color  
46 measurements ([http://ec.europa.eu/environment/water/water-framework/index\\_en.html](http://ec.europa.eu/environment/water/water-framework/index_en.html)). Hence,  
47 ocean color inversion is certainly among highest priority research topics for ocean-color  
48 community. Different aspects may explain the difficulties encountered in this inversion process.  
49 Firstly, the contribution of suspended matters to the reflectance in the near infrared (700-900 nm) is  
50 an issue as many algorithms expect these reflectances to be null. This assumption is called the black  
51 pixel hypothesis and relies on the strong natural absorption of the water in this domain (Antoine et  
52 al. 2006, Gordon and Wang, 1994). Secondly, bio-optical modelling, i.e. the estimation of the  
53 water-leaving reflectance from the Inherent Optical Properties (IOPs, namely the absorption and  
54 backscattering of the sea water constituents) in complex coastal waters is also challenging. Despite  
55 accurate physical models exist for open clear waters that cover 85% of the oceans, their derivation  
56 for coastal waters is more complex (Maritorena et al., 2002; IOCCG Report 3&5, 2000). Lastly,  
57 aerosol and water reflectance spectra may show important correlation in the near infrared, a spectral  
58 domain typically used by the standard algorithms to distinguish the two contributions.

59 As a consequence, available operational standard level-2 reflectance products may perform  
60 poorly in coastal areas, and consequently these products are often flagged as anomalous values for  
61 such areas (MERIS DPM, 2005). The result for end users is typically that very few observations are  
62 available in coastal areas if the standard quality standards are applied. For available pixels in coastal  
63 turbid waters, reflectances in the blue and green bands are often underestimated and may involve  
64 physically-meaningless negative values (Jamet et al., 2011; Goyens et al., 2013). Obviously Park et  
65 al.(2004) show this strongly affects relevance of level-2 products for the end users, which typically  
66 use water-reflectance spectra as inputs to estimate the chlorophyll-a and the suspended particulate  
67 matter concentrations (SPM, Doxaran et al. 2002), or the vertical light attenuation (Saulquin et al.  
68 2013; Morel et al. 2006, Wang et al., 2009; Jamet et al., 2012).

69 Over the last fifteen years, many coastal water algorithms have been developed to address user's  
70 needs for reliable water-reflectance data in coastal areas. Among them, the Schiller & Doerffer

(1999) MERIS Case2-Regional (C2R) based on a non-linear machine learning model, namely a Neural Networks (NN, Krasnopolsky & Schiller, 2003; Schiller & Doerffer, 1999), estimates water reflectance over turbid areas. The learning paradigm relies on the calibration of a non-linear model to relate the available satellite-derived observations to the geophysical quantity of interest from a training dataset. This training dataset typically consists of a collection of in-situ measurements along with the satellite-derived measurements. This learning-based strategy may suffer from two major drawbacks: weak geophysical/biological interpretability of this ‘black-box’ model and the assumption on the representativity of the training dataset. They may restrict the applicability of the model to a specific region and questions its validity with respect to the generally unknown variability of the atmospheric and water conditions.

Here, we develop a Bayesian latent class approach to address these limitations. To our knowledge, Bayesian model mixtures have been seldom explored for ocean color inversion (Frouin & Pelletier, 2014). The key feature of our model is the inversion of water and atmospheric signals from TOA observations using a multi-hypothesis setting. Rather than considering a single model, linear or not, we develop a Bayesian framework where the priors stated as mixture of models. Mixture models are trained both for water and aerosol contributions and lead to the identification of the reference spectrum families characterized by their mean spectrum and the associated covariance matrices. This training phase exploits in-situ data or radiative transfer simulations in the atmosphere and the water (Barker et al., 2008; Berk, 1999, Deuzé et al. 1989). Contrary to the machine learning approaches (NN, or Support Vector Regressions, SVR; Burges, 1998), the identified a priori distributions of the water and aerosol variables are directly linked to interpretable reference water or atmospheric spectra.

Our inversion scheme, referred to hereafter as MEETC2, is applied here to the estimation of water reflectances in complex waters from the MEdium Resolution Imaging Spectrometer (MERIS) TOA observations. Nevertheless the methodology is generic and may be directly applied to other sensors such as the incoming OLCI sensor embedded onto the sentinel 3 platform. Model

97 calibration and validation involve the MERIS MAtchup In-situ Database (MERMAID) radiometric  
98 in-situ dataset (Barker et al. 2008). Quantitative comparisons with the standard MEGS v8 (Antoine  
99 et al. 2006) and the MERIS C2R Neural Network outputs clearly demonstrate the relevance of our  
100 approach.

## 101 **2 Review of the standard Ocean Color inversion method**

### 102 **2.1 Atmospheric correction principles**

103

104 Ocean-color sensor measures at TOA the upwelling radiance ( $L_u$ ) in  $\text{mW.m}^{-2}.\text{sr}^{-1}$  backscattered  
105 by the ocean-atmosphere system. This radiance originates from photons scattered by air molecules  
106 and/or aerosols, which may also have been reflected directly at the sea surface (glint effect, Cox &  
107 Munk, 1954a&b), and may potentially have penetrated into the ocean. The measured TOA  
108 reflectance ( $\rho_{TOA}$ ) is the ratio between the upwelling radiance  $L_u$  and the downwelling irradiance  
109  $E_d$ , i.e.  $L_u$  integrated over the solid angle  $[0;2\pi]$ . The water reflectance contribution measured at  
110 TOA, i.e. transmitted through the atmosphere, is generally lower than 20% of the signal (Robinson  
111 et al. 2004). Due to this low signal/noise ratio, the unmixing of the atmospheric contribution from  
112 the water one reveals particularly complex. This inversion proves even more complex in coastal  
113 areas where the spectral water contribution may be closer to the aerosol contribution.

114 The traditional signal decomposition expresses measured TOA reflectance for each wavelength  
115  $\lambda$  as a sum of elementary contributions:

$$\rho_{gc}(\lambda) = \rho_{Ray}(\lambda) + \rho_{aer}(\lambda) + t_d(\lambda) \cdot \rho_w(\lambda) + \rho_{coupl}(\lambda) + \varepsilon(\lambda) \quad (1)$$

116

117 where  $\rho_{GC}$  is the observed TOA reflectance  $\rho_{TOA}$  corrected from the glint and gaseous  
118 absorption,  $\rho_{Ray}$  (known) the reflectance of a purely molecular atmosphere (no aerosol),

119  $\rho_{aer}$  (unknown) the reflectance of the aerosols,  $\rho_{coupl}$  (unknown) the coupling between air and  
 120 aerosol molecules,  $t_d$  (unknown) the diffuse transmittance of the atmosphere,  $\rho_w$  (unknown) the  
 121 water reflectance which is here the main quantity of interest to be estimated.  $\varepsilon$  is considered a white  
 122 noise process. Here we consider a classical multivariate normal distribution (MVN) noise with null  
 123 mean and spectral covariance matrix  $\Sigma_\varepsilon$ .

124 We consider here the Rayleigh-corrected reflectance variable  $\rho_{RC}(\lambda)$  (Antoine et al., 2005,  
 125 Santer et al. 1999, Gordon and Wang, Gordon 1997):

$$\rho_{RC}(\lambda) = \rho_{gc}(\lambda) - \rho_{Ray}(\lambda) = \rho_{aer}(\lambda) + t_d(\lambda) \cdot \rho_w(\lambda) + \rho_{coupl}(\lambda) + \varepsilon(\lambda) \quad (2)$$

126 The diffuse transmittance  $t_d$  is the product of both air molecules and aerosol particles scattering:

$$t_d(\lambda) = e^{-[0.5 \cdot \tau_{ray}(\lambda) + (1 - w_a(\lambda) \cdot F_a(\lambda)) \cdot \tau_a(\lambda)] \cdot M} \quad (3)$$

127 where  $\tau_{ray}(\lambda)$  is the Rayleigh optical thickness,  $\tau_a(\lambda)$  is the aerosol optical thickness,  $M$  the air  
 128 mass factor,  $w_a$  the aerosol single scattering albedo,  $F_a$  the forward probability scattering.  $\tau_a$  is  
 129 linked with the estimated aerosol reflectance for primary scattering (MERIS DPM, 2005):

$$\rho_{aer}(\lambda) = \frac{P(\lambda) \cdot w_a(\lambda)}{4(\cos(\Theta_s) + \cos(\Theta_v))} (1 - e^{-\tau_a(\lambda) \cdot M}) \quad (4)$$

130 where  $P(\lambda) \cdot w_a(\lambda)$  is the aerosol phase function times the single scattering albedo for the  
 131 current scattering angle,  $\Theta_s$  and  $\Theta_v$  are respectively the sun and the view zenith angles. For a fixed  
 132 geometry, aerosols contributions in the NIR are often assumed to follow an exponential decay  
 133 (Gordon & Wang, 1994):

$$\rho_{aer}(\lambda) = \rho_{aer}(\lambda_0) e^{c(\lambda - \lambda_0)} \quad (5)$$

134 where  $\lambda_0 = 865$  nm and  $c$  is the exponential decay of the aerosol spectrum, i.e. representative  
 135 of the aerosol type. Though relevant in the NIR domain, the assumption of an exponential decay

136 appears too restrictive in the 400-700 nm range where multiple scattering between aerosol and air  
137 molecules may become significant. Following Steinmetz et al. (2011), a polynomial model is  
138 considered to provide a more general model of aerosol contributions. Using our training dataset (cf  
139 § 4) a polynomial of order 3 was found as relevant to estimate the aerosol contributions:

$$\rho_{aer}(\lambda) = \rho_{aer}(\lambda_0) + a_1(\lambda - \lambda_0) + a_2(\lambda - \lambda_0)^2 + a_3(\lambda - \lambda_0)^3 \quad (6)$$

### 140 **2.1.1 The MERIS standard processing atmospheric correction scheme**

141

142 In the standard Level 2 processing of MERIS, the following four-step scheme is applied to  
143 estimate the water-leaving reflectances (Antoine & Morel, 2005):

144 1/ The signal is corrected from absorbing gaseous such as ozone, oxygen, water vapor and  
145 nitrogen dioxide.

146 2/ The estimated contribution of suspended matter particles in the NIR is removed from TOA  
147 observations after single scattering transmittance through the atmosphere. This step is known as the  
148 Bright Pixel Atmospheric Correction (BPAC) and detailed in the next section.

149 3/ A mixture of two aerosol models among 34 (for MERIS) is estimated from the values of ratio  
150  $\rho_{path} = \rho_{gc} / \rho_{ray}$  (Eq.2) at 779 and 865 nm, leading to the estimation of both the aerosol  
151 reflectance  $\rho_{aer}(\lambda)$  and the multiple scattering transmittance  $t_d(\lambda)$  (Eq.3).

152 4/ Water reflectance contribution is estimated by subtracting the estimated aerosol contribution  
153 from  $\rho_{aer}(\lambda)$  using Eq. 2.

154

### 155 **The Bright Pixel Atmospheric Correction (BPAC)**



Whereas, in open ocean waters, one can exploit null contribution of water reflectance in the near infrared (NIR) range to infer aerosol contributions, no such simple inversion scheme applies in coastal waters, which are characterized by a non-null contribution in this domain (Ruddick et al., 2005&2006) and possible correlations between atmospheric and water spectra. This is a major issue to be dealt with in the atmospheric corrections in coastal waters. BPAC is an iterative algorithm to correct the TOA signal from the estimated contribution of turbidity. It aims at removing the water contribution, caused by suspended matters, of the TOA observed reflectance (Moore et al., 1999). This step is essential in the standard MERIS level 2 processing as the estimation of the aerosols is performed using the NIR bands under the assumption  $\rho_w(NIR) = 0$ . Moore proposed for MERIS a two steps algorithm which iterates: the estimation of  $\rho_{aer}(709, 865)$ ,  $c$  and  $\rho_{aer}(779)$  using  $\rho_{path}(779, 865)$ , then, using the estimated residuals  $\hat{\rho}_w$  in the NIR from Eq.2 and a parametric model, the estimation of the SPM concentration and related  $\hat{\rho}_w$  at TOA. This converging algorithm suffers from important drawbacks for very turbid waters. In such areas, the considered water model does not allow retrieving high concentrations of SPM (Goyens et al. 2013). It typically leads to an over-correction of the blue water-reflectance, i.e. an underestimation of  $\rho_w$  at 412 and 442 nm with the standard Level 2 processing, and may resort to geophysically-meaningless negative reflectance values.

### 3 Method

#### 3.1 Spectral reference signatures of the sea water using Non-Negative Matrix Factorization

Given the spectral overlap of water and aerosol contributions especially in coastal areas, inversion of (Eq.1) requires some prior knowledge on water contributions. We propose here to determine from the training dataset a parametric spectral representation of water contributions. We

180 use here a Non-Negative Matrix Factorization (NNMF) with projected gradients (Lin, 2007).  
 181 Similarly to PCA, it relies on an additive decomposition on a basis learnt from the data. In contrast  
 182 to PCA, it does not involve orthogonality constraints but imposes non-negativity for both the basis  
 183 function and the projection coefficients. NNMF is among the most popular approaches in  
 184 multispectral and hyperspectral remote sensing as a mean to unmix contributions issued from  
 185 various sources in a sensed environment (Jia & Qian, 2009). Formally, NNMF leads to the  
 186 following parametric representation of a given water spectrum  $\rho_w(\lambda)$ :

$$\rho_w(\lambda) = W(\lambda, n) * h(n) \quad (7)$$

187 where  $W(\lambda, n) > 0$  spectral reference signature basis identified by NNMF using the training data,  
 188 and  $h(n) > 0$  refer to the coordinates of the spectrum  $\rho_w(\lambda)$  in the decomposition space. It may be  
 189 noticed that NNMF decomposition could also be replaced here by a bio-optical model.  
 190 Nevertheless, to our knowledge none of this model is today performant enough to estimate, in  
 191 coastal areas, the water leaving reflectance spectrum from the water's constituents. The NNMF  
 192 decomposition, by imposing non-negativity of both the coordinates and the reference spectral  
 193 signatures appropriately constrains our inversion to converge toward physically-realistic solutions  
 194 (cf § 4.3.1) conversely to the standard Level-2 processing (ESA and NASA). In our case, four  
 195 spectral reference signatures were needed to address the training in-situ spectrum variability (cf  
 196 4.2).

### 197 **3.2 Bayesian setting**

198

199 From Eq.1, the variables to be estimated are  $X_w = \{h_i\}$ , i.e. the coordinates of  $\rho_w$  in the basis  $W$   
 200 (Eq. 7), and  $X_a = \{a_i\}$  i.e. the polynomial coefficients of the aerosol models (Eq.6). Conversely to a  
 201 standard least square estimation framework, we develop a Bayesian setting. It relies not only on the  
 202 likelihood of the residuals  $\delta\rho_{RC} = \hat{\rho}_{RC} - \rho_{RC}$  but also on the prior distributions of  $X_a$  and  $X_w$ .

203 Formally, we consider the Maximum A Posteriori estimation (MAP, Harold & Sorenson, 1980)  
 204 which aims at maximizing the posterior probability  $P(X_a, X_w | \rho_{RC}, \varphi)$ :

$$P(X_a, X_w | \rho_{RC}, \varphi) \propto P(\rho_{RC} | X_a, X_w, \varphi) \cdot P(X_a, X_w | \varphi)$$

We suppose here that  $x_a$  and  $x_w$  are independent i.e.: (8)

$$P(X_a, X_w | \rho_{RC}, \varphi) \propto P(\rho_{RC} | X_a, X_w, \varphi) \cdot P(X_a | \varphi) \cdot P(X_w | \varphi)$$

205 The first term  $P(\rho_{RC} | x_a, x_w, \varphi)$  is the likelihood of the observation model (Eq. 1) with respect  
 206 to variables  $X_a$ ,  $X_w$  and  $\varphi$ .  $\varphi$  is here a vector of covariates composed of the observation geometry  
 207 and pre-estimates of the water and aerosol contributions in the NIR performed in the bright pixel  
 208 estimation step (BPE, § 4.3).  $P(X_a | \varphi)$  and  $P(X_w | \varphi)$  refer to the priors on  $X_a$  and  $X_w$  variables  
 209 given the covariates.

210 In the proposed Bayesian framework,  $P(\rho_{RC} | X_a, X_w, \varphi)$  is modeled with a multivariate normal  
 211 distribution with a null mean vector and full covariance matrix  $\Sigma_\epsilon$ . As detailed in the next sections,  
 212  $P(X_a | \varphi)$  and  $P(X_w | \varphi)$  a priori distributions are modeled using a mixture of MVN distributions,  
 213 namely a Gaussian Mixture Models (GMM, Reynolds, 1995). The MAP criterion cost function is  
 214 finally expressed using the log likelihood:

$$C = -\log(P(X_a, X_w | \rho_{RC}, \varphi)) \quad (9)$$

## 215 ***Covariates and non-homogeneous prior distributions***

216

217 Covariates are here geophysical parameters significantly correlated with the variables of  
 218 interest. From a physical point of view, the observed shape of aerosol reflectance spectrum  $\rho_{aer}(\lambda)$ ,  
 219 i.e.  $a_i$  coefficients of Eq.6, is correlated with the variables which describe geometry of acquisition  
 220 conditions ( $\Theta_s$ , the sun zenith angle,  $\Theta_v$ , the view zenith angle, and  $\delta\psi$ , the delta azimuth), and the

221 variables which describe the aerosol type and quantity,  $\rho_{aer}(865)$  and  $c$  (Eq. 5), estimated using the  
 222 NIR part of the spectrum (cf § 4.3). To characterize the correlation between variables and possible  
 223 covariates, we use a linear discriminant analysis (McLachlan et al., 2004) and the training dataset.

224 The selection of the significant contributors led to consider  $\varphi_a = \{\rho_{aer}(865), c, \theta_v, \theta_s\}$  for the  
 225 aerosol variable  $X_a$  and  $\varphi_w = \{\rho_w(780), c, \theta_v, \theta_s, \delta\psi\}$  for the water variable  $X_w$ . To derive the  
 226 priors of  $X_a$  and  $X_w$ , we model, using the training dataset and the Expectation Maximization  
 227 algorithm (Dempster, 1977), the joint distributions  $P(X_w, \varphi_w)$  and  $P(X_a, \varphi_a)$  as GMMs:

$$P(X_w, \varphi_w) = \sum_{Z_w=i} \Lambda_i g_{\Sigma_{\{X_w, \varphi_w\}_i}}(\{X_w, \varphi_w\} - \mu_{\{X_w, \varphi_w\}_i}) \quad (10)$$

$$P(X_a, \varphi_a) = \sum_{Z_a=j} \Lambda_j g_{\Sigma_{\{X_a, \varphi_a\}_j}}(\{X_a, \varphi_a\} - \mu_{\{X_a, \varphi_a\}_j})$$

228 We use subscript  $i$  (resp.  $j$ ) for the water-specific (resp. aerosol-specific) GMM.  $\Lambda_i$  is the prior  
 229 probability of mode  $i$  in the GMMs. It refers to the probability of the hidden state variable  $Z_w$  to be  
 230 in mode  $i$ ,  $P(Z_w=i)$  (resp.  $P(Z_a=j)$ ).  $g_{\Sigma_{\{X_w, \varphi_w\}_i}}$  (resp.  $g_{\Sigma_{\{X_a, \varphi_a\}_j}}$ ) is a zero-mean MVN distributions  
 231 with covariance matrix  $\Sigma_{\{X_w, \varphi_w\}_i}$  (resp.  $\Sigma_{\{X_a, \varphi_a\}_j}$ ) and mean vector  $\mu_{\{X_w, \varphi_w\}_i}$  (resp.  $\mu_{\{X_a, \varphi_a\}_j}$ ) for  
 232 the joint variables  $\{X_w, \varphi_w\}$  (resp.  $\{X_a, \varphi_a\}$ ) for mode  $i$  (resp.  $j$ ).

233 We update, given the covariate's values, the a priori distributions of Eq. (10) to obtain the a  
 234 priori conditional distributions of Eq.8:

$$P(X_w|\varphi_w) = \sum_{Z_w=i} \Lambda_{i|\varphi_w} g_{\Sigma_{X_w|\varphi_w,i}}(X_w - \mu_{X_w|\varphi_w,i}) \quad (11)$$

$$P(X_a|\varphi_a) = \sum_{Z_a=j} \Lambda_{j|\varphi_a} g_{\Sigma_{X_a|\varphi_a,j}}(X_a - \mu_{X_a|\varphi_a,j})$$

235 These non-homogeneous priors on  $X_a$  and  $X_w$  involve conditional means, covariances and priors  
 236 given the covariates and the initial GMM model estimated onto the joint variables. For instance, for  
 237 mode  $i$  of the aerosol prior, the conditional parameters are given by (Petersen, 2008):

$$\begin{aligned}\mu_{X_a|\varphi_a,j} &= E(X_a|\varphi_a, Z_a = j) = \mu_{X_{a,j}} + \Sigma_{X_a,\varphi_a,j} \cdot \Sigma_{\varphi_a,j}^{-1} \cdot (\varphi_a - \mu_{\varphi_{a,j}}) \\ \Sigma_{X_a|\varphi_a,j} &= \Sigma_{X_{a,j}} - \Sigma_{X_a,\varphi_a,j} \cdot \Sigma_{\varphi_a,j}^{-1} \cdot \Sigma_{\varphi_a,X_{a,j}}\end{aligned}\tag{12}$$

$$\Lambda_{X_a|\varphi_a,j} = \Lambda_j * P(X_a|\varphi_a, Z_a = j) / \sum_l \Lambda_j * P(X_a|\varphi_a, Z_a = l)$$

### 238 **3.3 Functional scheme**

239

240 Figure 1 summarizes the functional scheme for the proposed Bayesian inversion given the  
 241 calibrated model parameters, i.e. means and covariance matrices for the GMM models and the  
 242 MVN distribution of the residuals,  $\hat{\rho}_{RC}(\lambda) - \rho_{RC}(\lambda)$ , of Eq. (1). In the first step, the Bright Pixel  
 243 Estimation (BPE) is based on the water similarity spectrum (Ruddick et al., 2005) to estimate  
 244  $\rho_w(780)$ ,  $\rho_{aer}(865)$  and  $c$  the slope of the aerosol. An iterative convergent algorithm is used.  
 245 Given the estimated covariates, we update in step 2 the GMM for  $X_a$  and  $X_w$  conditionally to the  
 246 covariates (Eq.(12). Step 1 & 3 involve gradient descent based inversions and a Taylor series of Eq.  
 247 (4). As the MAP criterion may not be a concave criterion, the initialization of the gradient descent,  
 248 step 3, is a key issue as gradient-based maximization may converge toward local minima. We  
 249 proceed as follows: 25 aerosol parameters are randomly generated using the updated distributions.  
 250  $X_w$  initialization is performed using the estimated  $\hat{\rho}_w(780)$ ,  $X_a$  initialization and Eq. 2. Overall,  
 251 among the 25 computations, we select in step 4 the solution corresponding to the highest value of  
 252 the MAP criterion (Eq.9).

253

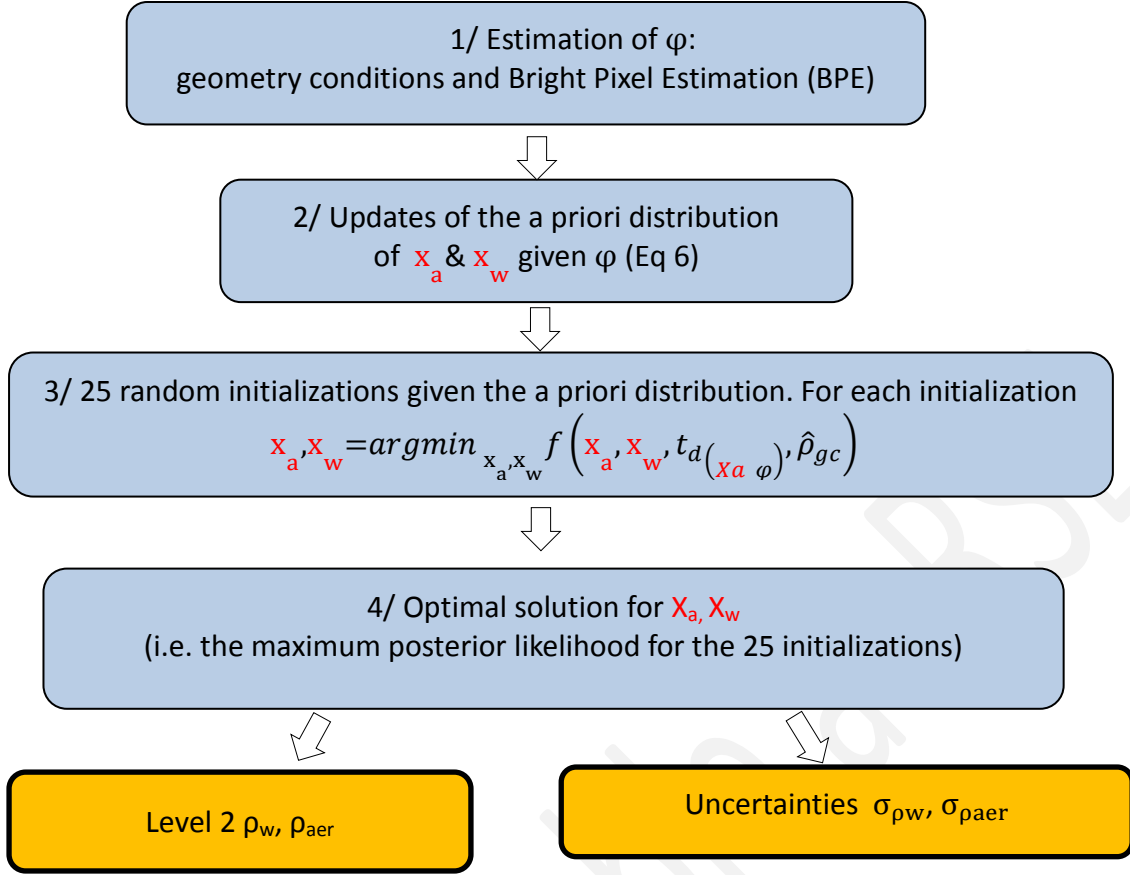


Figure 1: operational scheme for the atmospheric correction MEETC2 bayesian inversion

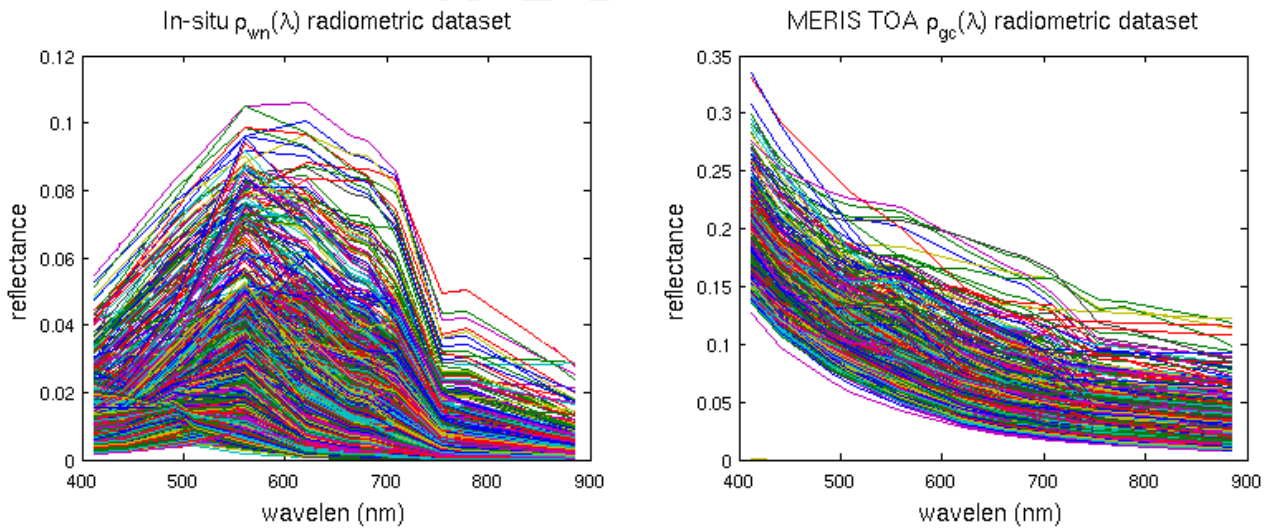
## 4 NUMERICAL EXPERIMENTS

To validate the proposed methodology, the 5976 radiometric in-situ profiles have been randomly splitted into two sets of equal size: a training dataset and a validation dataset. Model parameters are estimated using the training dataset. The optimal number of clusters,  $k$ , used in the GMM to estimate  $X_a$  and  $X_w$  a priori Probability Density Functions (PDF), is determined using the Bayes Information Criterion (BIC) (Bhat & Kumar, 2010) and the explained variance criterion (Saulquin et al., 2014). Validation is performed with the validation dataset, using scatter plots between estimated and in-situ  $\rho_w(\lambda)$ , histograms, and related regression statistics. We evaluate type-II regression statistics, i.e. a regression model that accounts for uncertainties for both  $y$  and  $x$  as the in-situ measurements also involves uncertainties (Laws, 1997).

#### 275 4.1 The in-situ MERMAID dataset

276

277 The MERMAID (<http://hermes.acri.fr/mermaid/home/home.php>) in-situ matchup database is a  
 278 comprehensive dataset that gathers in-situ measurements of water leaving radiances, IOPs, and  
 279 MERIS TOA reflectances measured at the same location. Many sites are available and among them,  
 280 the most known are the NASA bio-Optical Marine Algorithm Dataset (NOMAD, Werdell &  
 281 Bailey, 2005), the “BOUée pour l’acquiSition d’une Série Optique à Long terme” (BOUSSOLE,  
 282 Antoine et al., 2006) mooring program, the Aerosol Robotic Network (AERONET; Zibordi et al.  
 283 2009) stations, the Helgoland transect (Petersen et al. 2008) that provides a full dataset of  
 284 radiometric in-situ measurements in the Baltic Sea complex waters, and the MUMM Trios dataset  
 285 (Ruddick et al., 2006). Our initial dataset gathers 5976 matchups (without glint) measured at the  
 286 MERIS wavelengths: 412.5, 442.5, 490, 510, 560, 630, 665, 681, 708, 753.75, 778.75, 865 and 885  
 287 nm. For each in-situ measurement of Figure 2 (left), we use the corresponding 3 by 3 MERIS  
 288 pixels (Bailey & Werdell, 2006).

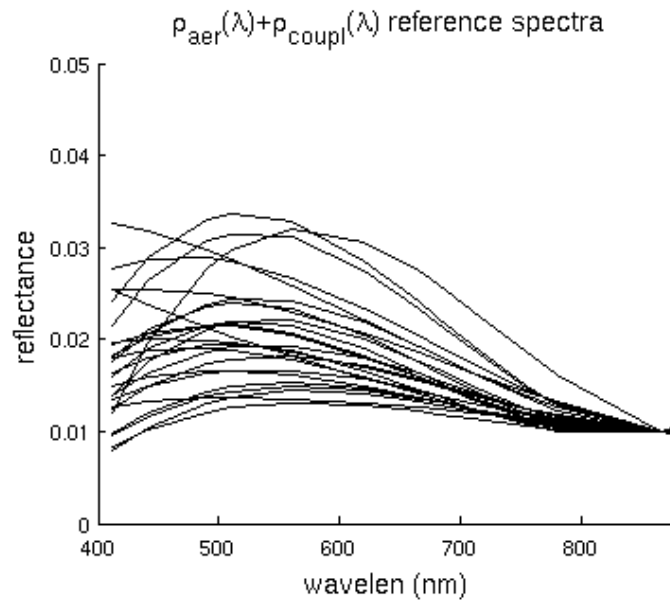


289 Figure 2: Left, the 5976 in-situ water reflectance spectra in complex waters. Right, the  
 290 corresponding (matchups)  $\rho_{gc}$  (TOA) observed from the MERIS sensor.

291

## 292 4.2 Calibrated priors

294 A 25-mode mixture model (cf § 3.3) was selected to model the joint distribution of  $\{X_a, \varphi_a\}$ .  
 295 Figure 3 shows the 25 aerosol modes reconstructed from the GMM centers for  $\rho_{aer}(865) =$   
 296 0.01. We remind that the PDF of  $\{X_a, \varphi_a\}$  involves a full covariance matrix  $\Sigma_{0_{Xai}}$  for each mode  
 297 that is accounted for in the maximization of Eq. 8.



298  
 299 Figure 3: The 35 aerosol modes reconstructed from the GMM and Eq 6.

300 From the NNMF applied to the in-situ water spectra, a four reference spectral signatures,  $W(\lambda)$   
 301 was needed to reconstruct 99% of the variance of the in-situ spectra training dataset (Eq.7, Figure  
 302 4a). The NNMF reference spectral signatures characterize the influence of the optically active  
 303 constituents of the water column onto the observed water leaving reflectance spectra:

- 304 ■ Reference n°1 (dark blue) is a typical spectral signature observed in presence of SPM  
 305 (Doxaran et al., 2002; Bricaud et al. 1998).
- 306 ■ Reference n°2 (green) highlights the spectral signature of CDOM absorption with its typical  
 307 decrease toward the blue.



- Reference n°3 (red) is the typical spectral signature of the chl-a, i.e. absorption in the blue and the resulting observed in the green (560 nm).
- Reference n°4 (light blue) is the spectral signature of the pure water (Pope & Fry, 1997).

A 35-mode GMM is optimal to fit the prior distribution of  $\{X_w, \varphi_w\}$ . Figure 4 shows the 35 water reflectance reference spectra. Similarly to the prior distribution of aerosol contributions, a full covariance matrix  $\Sigma_{0_{xwi}}$  is estimated for each mode  $i$ .

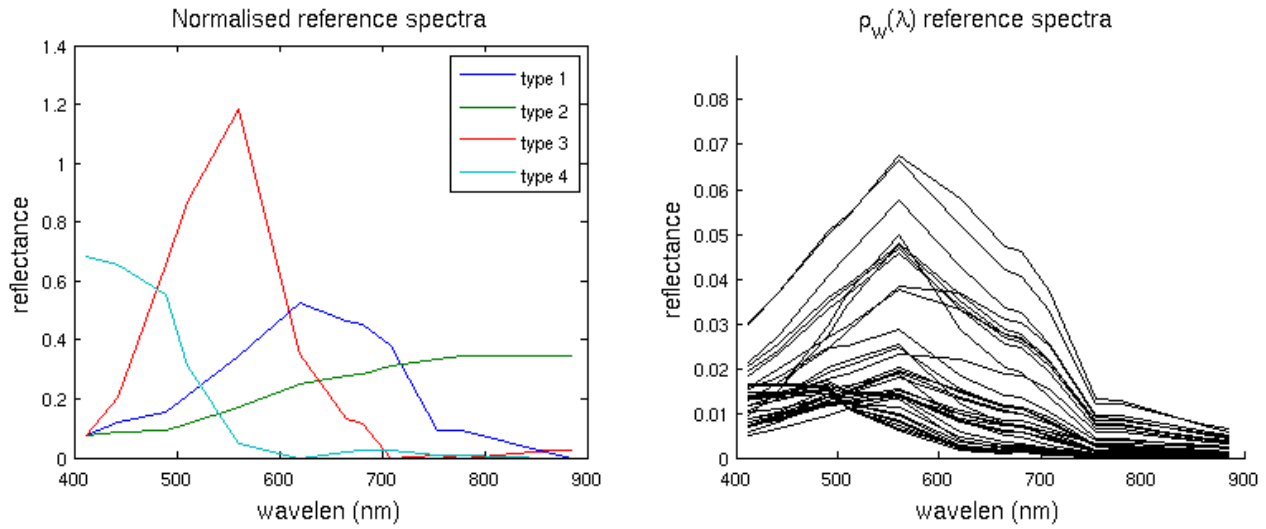


Figure 4: Left, the reference spectral signature basis,  $W(\lambda)$ , estimated using NNMF with projected gradients. Right, the 35 reference water models reconstructed using the GMM centers and Eq 7.

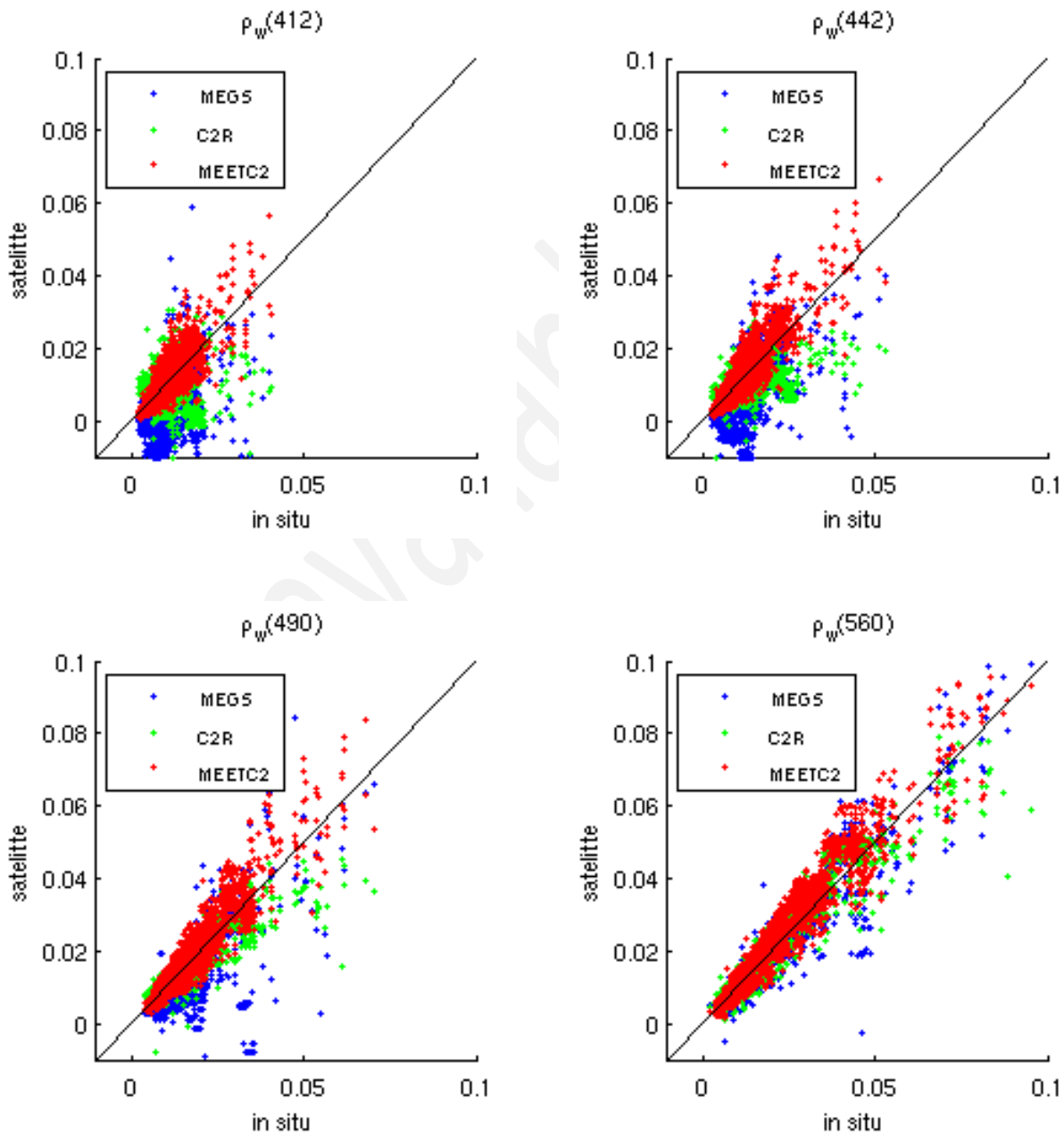
### 4.3 Ocean color inversion results

#### 4.3.1 Inversion performance for the Mermaid dataset

We perform a quantitative evaluation of the performance of the proposed Bayesian inversion model, MEETC2, for the Mermaid dataset and coastal waters. For the validation dataset, i.e. the half of the 5976 spectra, we analyze for each wavelength the estimated water reflectances  $\hat{\rho}_w$  against in-situ measurements (Figure 5, red). In addition to the proposed Bayesian inversion, we

324 also report on Figure 5 the inversion performed with MEGS v8 (blue), and C2R (green). Table 1  
325 summarizes the corresponding statistical results for the 13 wavelengths.

326 On this validation dataset, MEETC2 clearly outperforms MEGS and C2R at bands 412, 442,  
327 490 and 510 nm in term of mean-bias, mean absolute error, slope,  $R^2$  coefficient and  $\sigma$ . From 620  
328 to 885 nm MEETC2 slightly outperforms the two other models. Overall, the gain on the relative  
329 absolute error over the 12 bands is of 57% compared with MEGS and 10% compared with C2R.



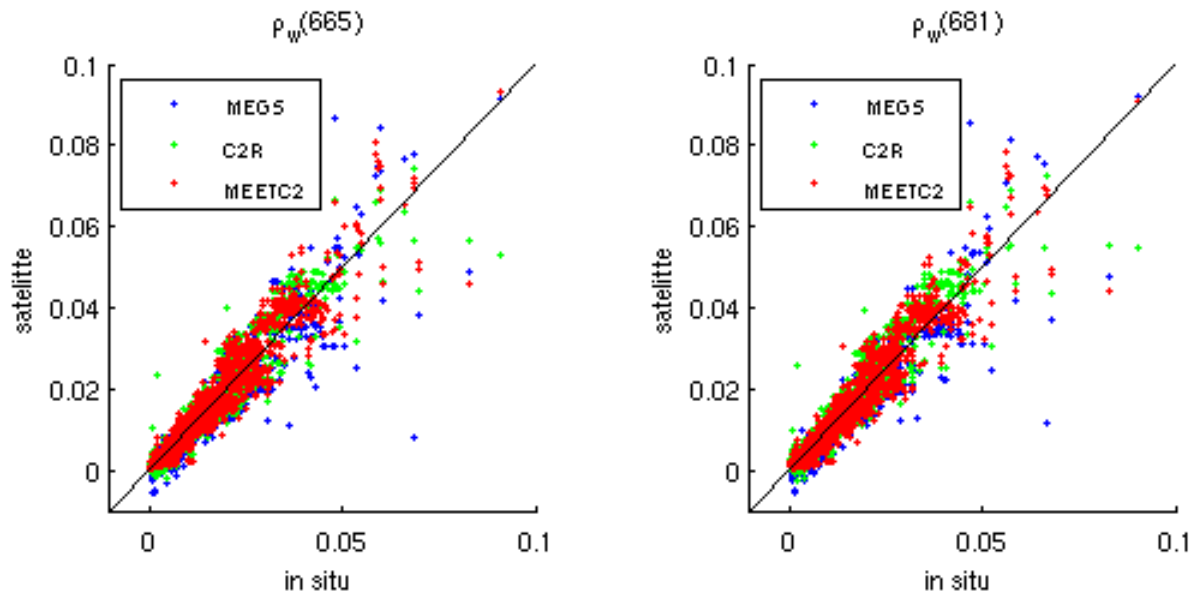


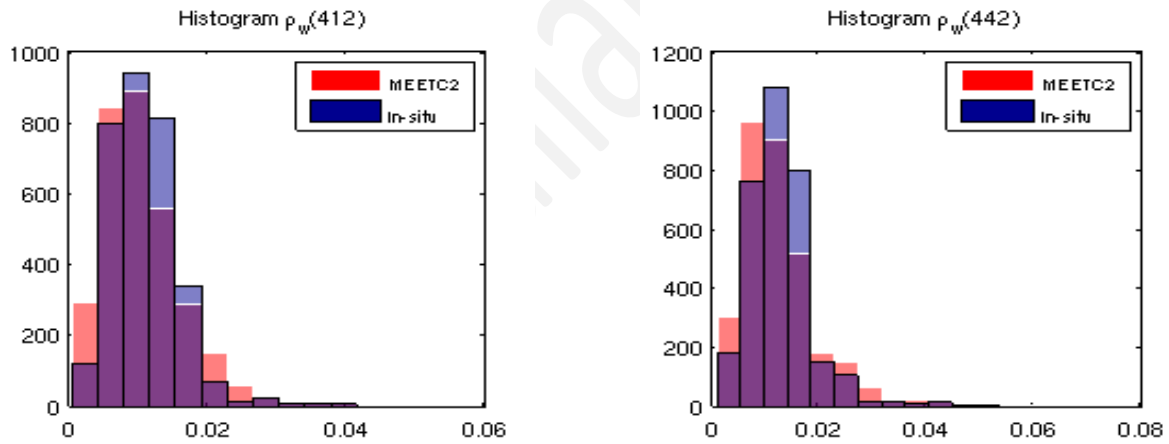
Figure 5: comparisons between the estimated  $\hat{\rho}_w$  at 412, 442, 490, 560, 665 and 681 nm using MEETC2 vs in-situ (red), MEGS 8 vs in-situ (blue) and C2R (NN) vs in-situ (green).

Table 1: Statistical analyses of the estimated water reflectances vs. in-situ data for the proposed Bayesian model (MEETC2), the standard MEGS processor and the neural-net-based algorithm C2R. For each wavelength, we report the mean error (bias), the slope of the regression with the in situ data, the associated  $R^2$  score and standard deviation,  $\sigma$ . We report in bold the algorithms which provided the best performance.

$\lambda(\text{nm})$		Mean error	Slope	$R^2$ (Pearson)	$\sigma$
412.5	MEETC2	<b>-0.0004</b>	<b>1.14</b>	<b>0.70</b>	<b>0.0039</b>
	MEGS	-0.0083	0.44	0.16	0.0149
	C2R	-0.0023	0.42	0.15	0.0075
442.5	MEETC2	<b>-0.0002</b>	<b>1.13</b>	<b>0.75</b>	<b>0.0041</b>
	MEGS	-0.0060	0.39	0.38	0.0128
	C2R	-0.0031	0.31	0.60	0.0060
490	MEETC2	<b>0.0004</b>	<b>0.97</b>	<b>0.77</b>	<b>0.0049</b>
	MEGS	-0.0033	0.92	0.76	0.0098
	C2R	-0.0022	0.77	0.57	0.0046
510	MEETC2	<b>0.0007</b>	<b>0.96</b>	<b>0.85</b>	<b>0.0046</b>
	MEGS	-0.0020	0.91	0.65	0.0085
	C2R	-0.0013	0.72	0.78	0.0040
560	MEETC2	<b>0.0007</b>	<b>1.04</b>	0.88	<b>0.0049</b>
	MEGS	-0.0007	0.95	0.81	0.0056
	C2R	-0.0007	0.88	<b>0.90</b>	0.0050
620	MEETC2	<b>0.0006</b>	<b>1.00</b>	<b>0.93</b>	<b>0.0038</b>
	MEGS	-0.0014	1.05	0.85	0.0050
	C2R	-0.0012	0.97	0.90	0.0039
665	MEETC2	<b>1.1809e-03</b>	<b>0.97</b>	<b>0.88</b>	<b>0.0033</b>

	MEGS	-0.8881-03	1.07	0.85	0.0043
	C2R	0.2650-03	1.02	<b>0.88</b>	<b>0.0033</b>
681	MEETC2	<b>0.0657e-03</b>	<b>0.99</b>	<b>0.92</b>	<b>0.0033</b>
	MEGS	-0.6257e-03	1.06	0.85	0.0041
	C2R	0.5037e-03	1.02	0.89	<b>0.0033</b>
708	MEETC2	<b>-0.0136e-03</b>	<b>0.94</b>	<b>0.87</b>	<b>0.0030</b>
	MEGS	-0.6400e-03	1.13	0.83	0.0039
	C2R	0.7668e-03	1.10	<b>0.87</b>	0.0037
753	MEETC2	<b>0.2656 e-03</b>	<b>0.90</b>	<b>0.90</b>	<b>0.0014</b>
	MEGS	-0.4222 e-03	1.35	0.78	0.0027
778	MEETC2	<b>0.2418 e-04</b>	<b>0.90</b>	<b>0.89</b>	<b>0.0014</b>
	MEGS	-0.4084e-04	1.21	0.74	0.0029
865	MEETC2	<b>0.3793 e-04</b>	0.94	<b>0.88</b>	<b>0.0009</b>
	MEGS	-0.2259e-04	<b>1.02</b>	0.75	0.0017
885	MEETC2	<b>0.2794 e-04</b>	0.95	<b>0.88</b>	<b>0.0007</b>
	MEGS	-0.2269e-04	<b>1.02</b>	<b>0.88</b>	0.0014

337 We further analyze the extent to which we recover realistic water reflectances from the  
338 proposed Bayesian inversion MEETC2. To this end, we compared for each wavelength, the  
339 distribution of in-situ measurements to the MEETC2 estimates. Figure 6 shows a global agreement  
340 between the distributions of  $\hat{\rho}_w(\lambda)$ , compared to the reference in-situ distributions, for wavelengths  
341 412, 442, 510 and 560 nm.



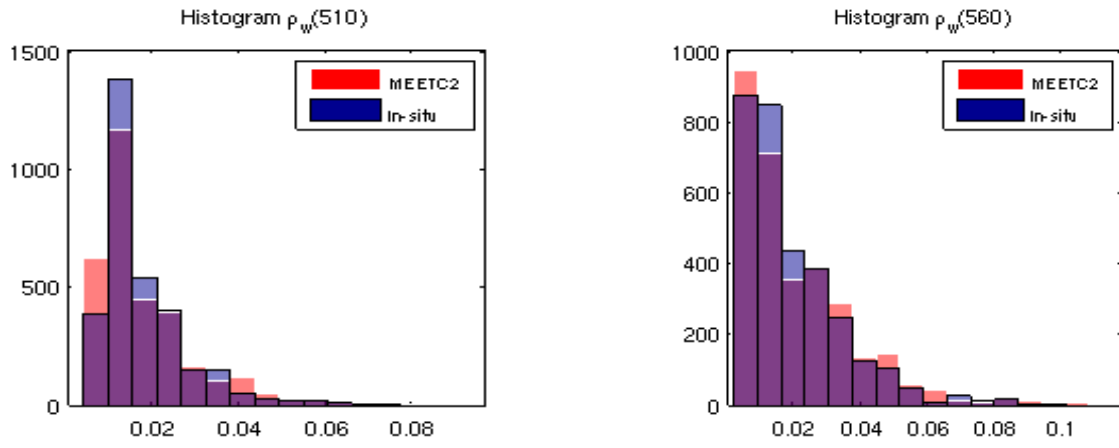
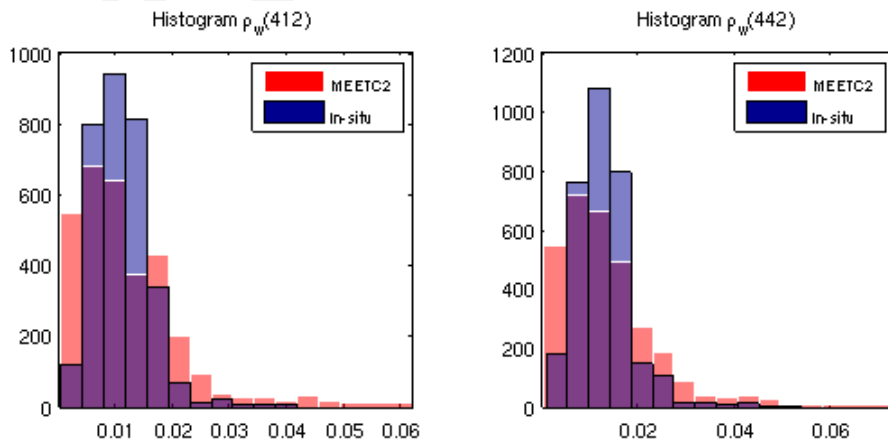


Figure 6: Comparison of the distributions of the estimated water reflectances  $\hat{\rho}_w$  at 412, 442, 510, 560 nm for in-situ measurements (blue) and the proposed inversion (MEETC2 model, red).

To illustrate the added value of the introduction of priors on both water and aerosol spectra, we implement model (Eq. 8) without priors on  $X_a$  and  $X_w$ , i.e. the cost function of Eq. 9. In this case, the cost function in the inversion is directly comparable with the one of a Generalized Least Square Model (GLS) with error covariance matrix  $\Sigma_\epsilon$ . Figure 7 shows the corresponding results obtained, using the same validation dataset. We clearly see in Figure 7a smoothing effects for bands 412, 442, 510 and 560 nm on the estimated distributions of  $\hat{\rho}_w$ . The resulting bias with the in-situ is lower using the MAP estimator and a priori knowledge (Figure 6).



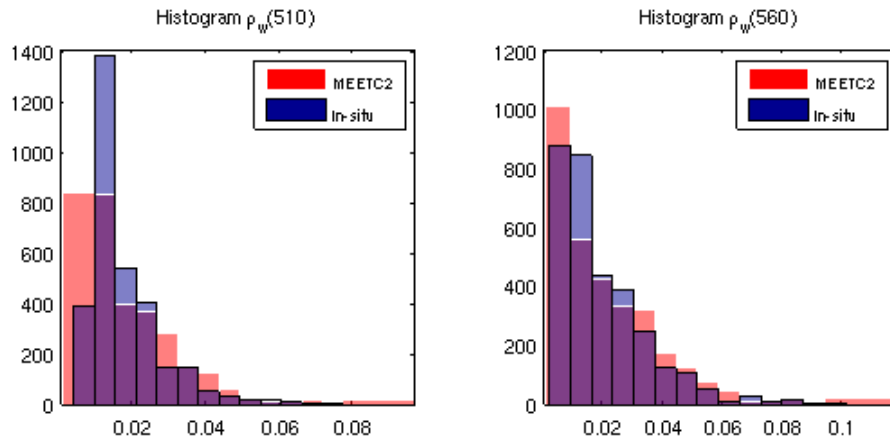


Figure 7: Comparison of the distributions of the estimated water reflectances  $\hat{\rho}_w$  at 412, 442, 510, 560 using a cost function without a priori for the inversion (Eq. 9) vs in-situ. In that case, the MAP criterion reduces to the Maximum Likelihood criterion.

#### 4.3.2 Example of estimated water reflectance on a very turbid area

Figure 8 shows the estimated  $\hat{\rho}_w$ , using the 20090322 MERIS Full Resolution (FR) level 1 observations over the French La Gironde's estuary, using the three algorithms. At springtime in this area, a bloom occurs leading to high chl-a concentrations (typically of magnitude from 5 to 15  $\text{mg.m}^{-3}$ ). At the same time, the seasonal river outflow involves high, SPM concentrations and CDOM absorption (Doxaran et al. 2009). In the same manner, Figure 9 shows the estimated  $\hat{\rho}_w$ , using the 20040208 MERIS Full Resolution (FR) level 1 observations over the French La Seine's estuary.

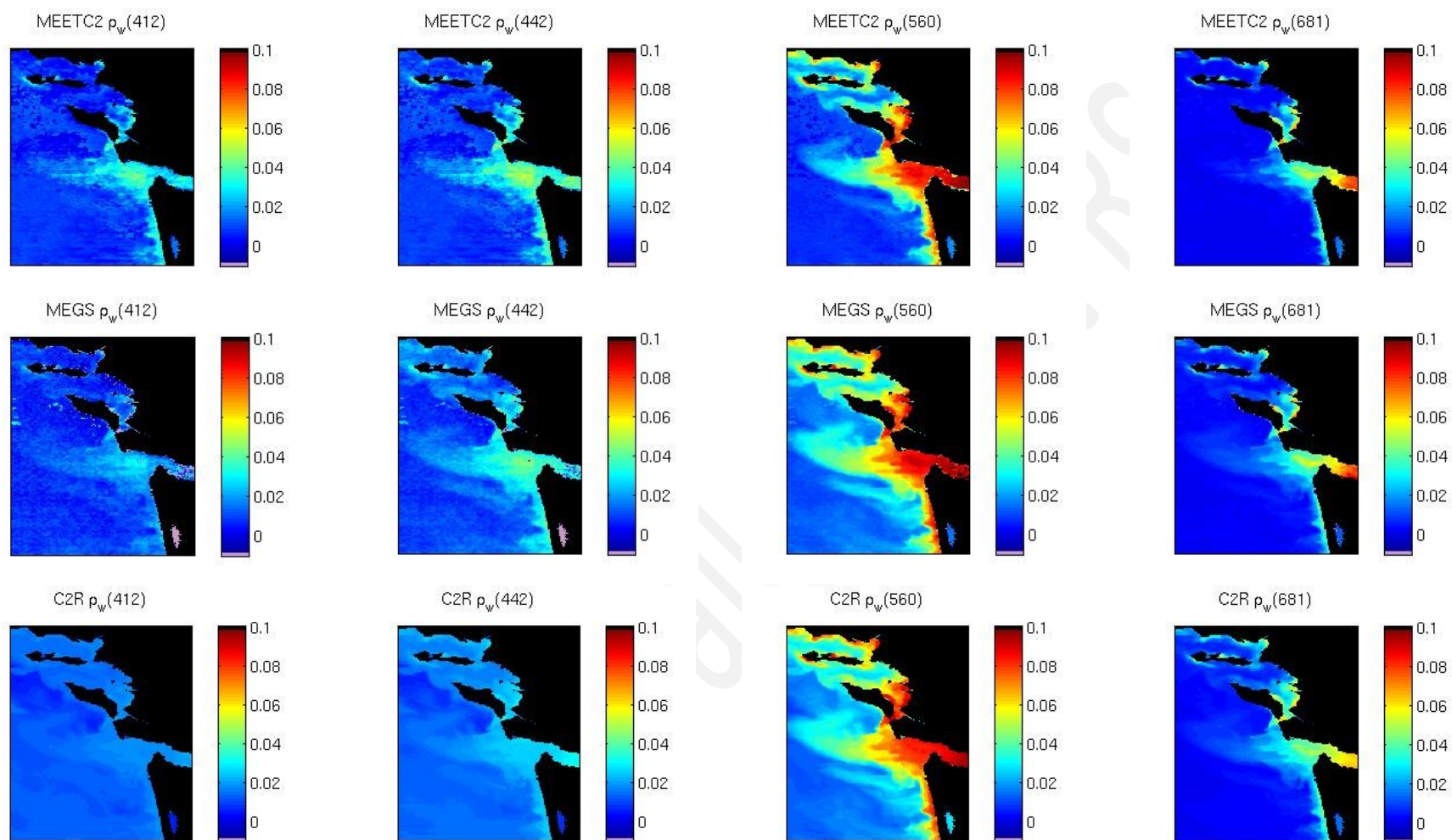


Figure 8: Estimated  $\rho_w(412, 442, 560, 680)$  (left to right) from the MERIS FR Level 1 image of the 20090322 over the French river La Gironde's estuary. Top, MEETC2 retrievals, middle, MEGS v8 and bottom C2R retrievals.



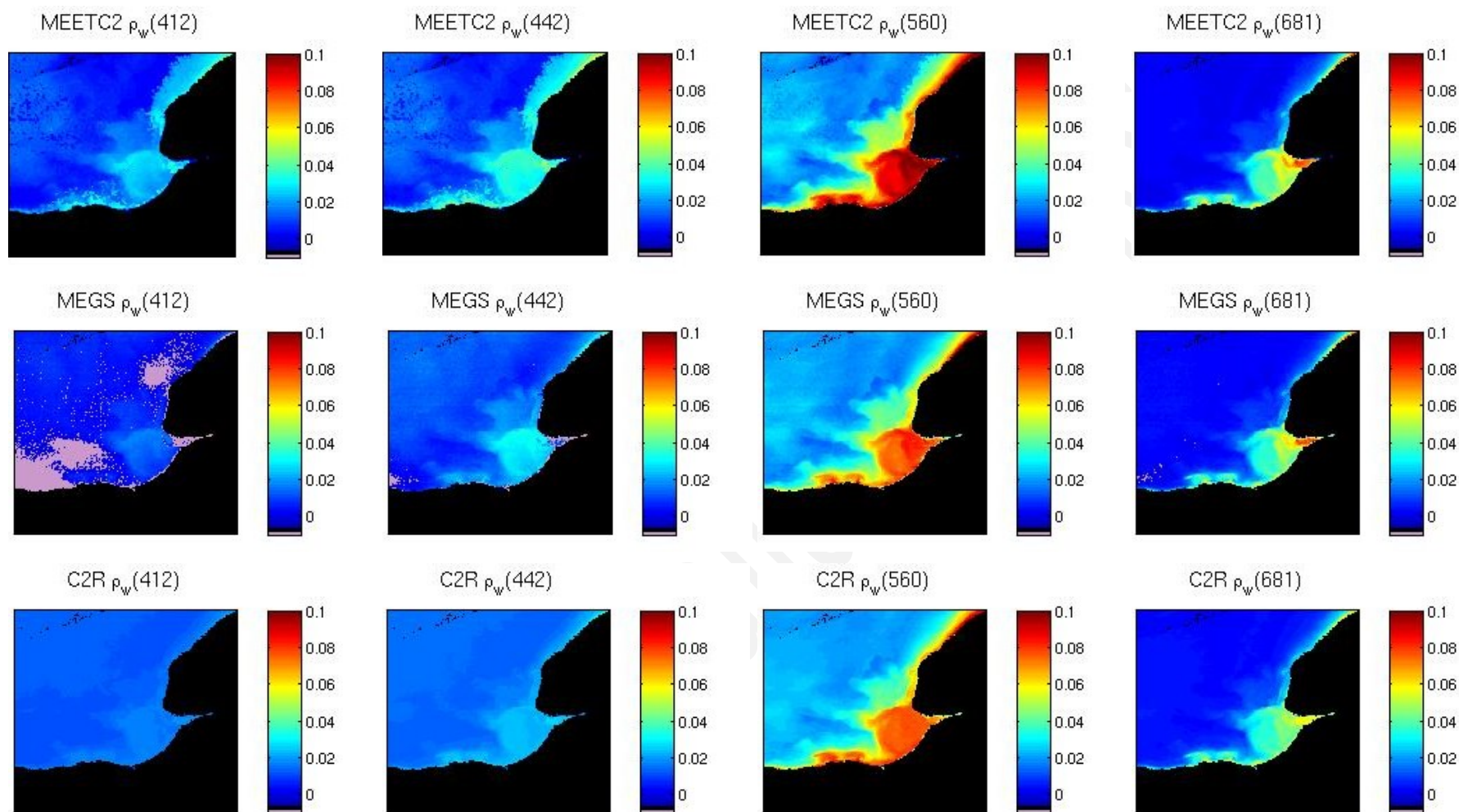


Figure 9: Estimated  $\rho_w(412, 442, 560, 680)$  (left to right) from the MERIS FR Level 1 image of the 20040208 over the French river La Seine's estuary. Top, MEETC2 retrievals, middle, MEGS v8 and bottom C2R retrievals. In pink are highlighted negative reflectances.

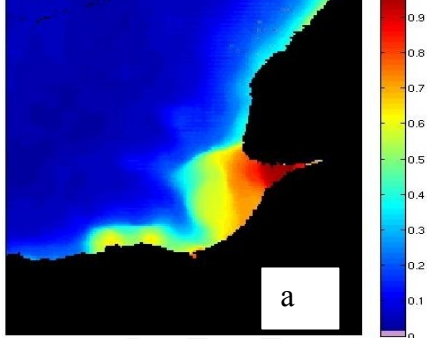


### 363 4.3.3 Estimated water types associated with the MEETC2 inversion

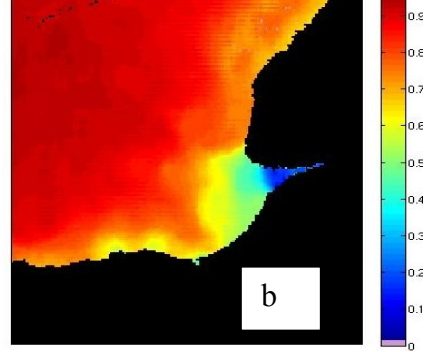
364

365 The NNMF reference spectral signatures in Figure 4 (left) characterize the influence of the  
 366 optically active constituents of the water onto the observed water leaving reflectance spectra. This  
 367 reference basis may be used to validate indirectly, using oceanographic knowledge, the spatial  
 368 coherency of the estimated  $\hat{\rho}_w$ . Figure 10 depicts the projection coefficients of the estimated  
 369 MEETC2  $\hat{\rho}_w$  onto the reference spectral signatures. Figure 10c depicts the presence of chl-a over  
 370 the all area as expected for this spring period and region. We observe the typical clear contrasted  
 371 situation in an estuary between waters whose spectral shape is mainly constrained by SPM (Figure  
 372 10a) and clearer waters (Figure 10d) in the oceanic part of the estuary. This spatial consistency of  
 373 the distribution of the water types from the estimated  $\hat{\rho}_w$ , relatively to our knowledge of the  
 374 seasonal behavior in this area, contributes to validate the shapes of our estimated spectra.

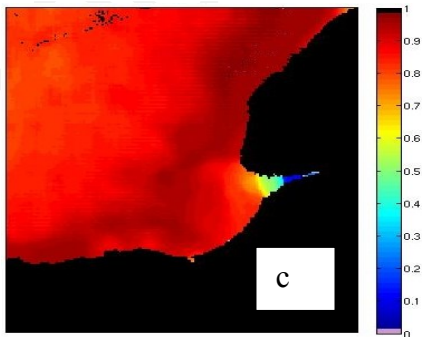
Type 1: SPM dominated  $\rho_w$



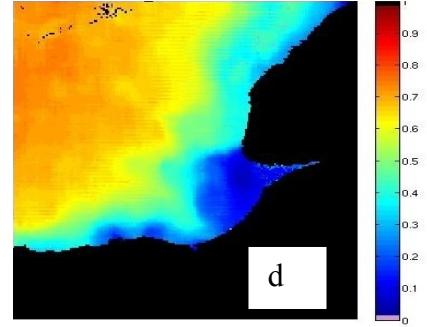
Type 2: Cdom dominated  $\rho_w$



Type 3: Chl-a dominated  $\rho_w$



Type 4: pure water dominated  $\rho_w$



375 Figure 10: projection coordinates of the MEETC2 estimated  $\hat{\rho}_w$  in the four NNMF reference  
 376 spectral signature basis.

## 377 **5 Discussion**

378

### 379 **A significant improvement of ocean color inversion in coastal waters.**

380 Retrieving reliable Ocean Color reflectances from space in coastal areas remains a major  
381 challenge for a number of operational and scientific issues, including for instance the delivery of  
382 reliable satellite-derived products in coastal areas for the space agencies, bio-optical and biological  
383 modeling, as well as environmental monitoring policies such as the WFD. Using the MERMAID  
384 satellite/in-situ collocated observation database, a Bayesian latent class model was shown to  
385 significantly enhance the inversion of water reflectances for complex waters compared to the  
386 standard MEGS inversion scheme and the C2R, a Neural Network trained using similar in-situ data  
387 (Schiller & Doerffer, 1999).

388 The improvements were especially noticeable for the 412, 442, 490 and 510 nm bands, which  
389 are used in Ocean Color for the estimation of the chl-a concentration, CDOM absorption and light  
390 attenuation underlying the potential of such approach to improve the standard level 2 products in  
391 coastal areas. An additional important feature of the proposed inversion, based onto the Non  
392 Negative Matrix Factorization water model, is strictly positive estimates of the water leaving  
393 reflectances in coastal areas. Meaningless negative estimates, as observed in the standard MEGS  
394 products are not anymore possible.

395 The complexity of the inversion is particularly stressed by the number of needed hidden models,  
396 respectively 25 for coastal aerosol reflectances and 35 for water reflectances, to address the spectral  
397 variability of both water and atmospheric contributions in such areas, and to unmix the possibly  
398 correlated aerosol and water spectra.

399

400

## **A physically-interpretable modelling framework**

Conversely to Neural Network, the modes retrieved by the Gaussian Mixture Models correspond to identifiable aerosols, such as identified in the MERIS and the OLCI reference aerosol database, and water types. The fact that we explicitly distinguish parametric representations of aerosol and water spectra makes also easier the independent calibration of the models and our Bayesian model may benefit in a much simpler manner for newly collected and/or simulated dataset to improve each prior distribution independently. This is regarded as a key property for future operational applications with respect to ongoing advances in radiative transfer modelling, in-situ monitoring and future satellite missions.

## **Operational potential in the framework of the ocean sensor of upcoming Sentinel 3 platform**

The incoming OLCI Ocean Color sensor, embedded on the Sentinel 3 platform, should succeed the MERIS sensor in 2015. The available spectral bands will be close to the MERIS ones. Beyond genericity of our Bayesian framework, we thus expect the considered parameterization, especially the NNMF-based representation, the GMM-based priors and the covariance models, to be directly transferable to the future OLCI observations. Our ongoing work addresses the development of an operational product based on the proposed Bayesian mode that will be freely distributed in the Odesa software (<http://www.odesa-info.eu/info/>). The dependency of both aerosol and water prior distributions to the observation geometry conditions will be addressed soon using radiative transfer simulations such as the Successive Order Scattering radiative transfer code (Deuzé, 1989) and Hydrolight (Mobley, 1998) to cover the full possible range of observation conditions.

From a modeling perspective, additional developments appear of interest, especially new covariates, e.g. humidity and wind conditions to further constrain the prior distributions of the water and aerosol variables. Parallelized implementation is also under investigation, as, conversely to existing MEGS and C2R processors, our optimization is computationally more demanding than

these as it relies on quasi-randomized initializations for the atmospheric initial model, i.e. multiple initializations given the observed geometry conditions and per-estimates in the near infrared. Optimal and noiseless results will be obtained with increased number of random initializations to converge towards the ‘true’ solution. This random initialization issue and the associated computing cost, is classic for genetic algorithms (Davis, 1991) and the new generation of satellite products such as the Soil Moisture Ocean Salinity (SMOS) product (Font et al., 2010).

## 6 References

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

Antoine, D., Chami, M., Claustre, H., d'Ortenzio, F., Morel, A., Becu, G., Gentilli, B., Louis, F., Ras, J., Roussier, E., Scott, A., Tailliez, D., Hooker, S. B., Guevel, P., Deste, J.-F., Dempsey, C. & Adams, D. (Eds.) 2006. BOUSSOLE: A Joint CNRS-INSU, ESA, CNES, and NASA Ocean Color Calibration and Validation Activity NASA Technical Memorandum No. 2006-214147. NASA/GSFC, Greenbelt, MD.

Antoine, D., & Morel, A. (2005). MERIS ATBD 2.7: Atmospheric correction of the MERIS observations over ocean case 1 waters.

Antoine, D., & Morel, A. (1999). A multiple scattering algorithm for atmospheric correction of remotely sensed ocean colour (MERIS instrument): principle and implementation for atmospheres carrying various aerosols including absorbing ones. *International Journal of Remote Sensing*, 20(9), 1875-1916.

447 Barker, K., Mazeran, C., Lerebourg, C., Bouvet, M., Antoine, D., Ondrusek, M., Zibordi,,  
 448 Lavender, S. (2008), MERMAID : The MERis MATCHup In-situ Database, proceedings of the 2nd  
 449 MERIS/(A)ATSR User Workshop, ESA/ESRIN, Italy, September 2008.

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

452 Berk, A., Anderson, G. P., Bernstein, L. S., Acharya, P. K., Dothe, H., Matthew, M. W., ... &  
 453 Hoke, M. L. (1999, October). MODTRAN4 radiative transfer modeling for atmospheric correction.  
 454 In *SPIE's International Symposium on Optical Science, Engineering, and Instrumentation* (pp. 348-  
 455 353). International Society for Optics and Photonics. Bhat, H. S., & Kumar, N. (2010). On the  
 456 derivation of the Bayesian Information Criterion. School of Natural Sciences, University of  
 457 California.

458 Bricaud, A., Morel, A., Babin, M., Allali, K., & Claustre, H. (1998). Variations of light  
 459 absorption by suspended particles with chlorophyll a concentration in oceanic (case 1) waters:  
 460 Analysis and implications for bio-optical models. *Journal of Geophysical Research: Oceans* (1978–  
 461 2012), 103(C13), 31033-31044.

462 Burges, C. J. (1998). A tutorial on support vector machines for pattern recognition. *Data mining*  
 463 *and knowledge discovery*, 2(2), 121-157.

464 Cox, C. and W. H. Munk (1954a): Statistics of the sea surface derived from sun glitter.  
 465 J. Mar. Res., 13, 198–227

466 Cox, C., & Munk, W. (1954). Measurement of the roughness of the sea surface from  
 467 photographs of the sun's glitter. *JOSA*, 44(11), 838-850.

468 Davis, L. (Ed.). (1991). *Handbook of genetic algorithms* (Vol. 115). New York: Van Nostrand  
 469 Reinhold.

470 Dempster, A.P.; Laird, N.M.; Rubin, D.B. (1977). "Maximum Likelihood from Incomplete Data  
471 via the EM Algorithm". *Journal of the Royal Statistical Society, Series B* 39 (1): 1–38. JSTOR  
472 2984875. MR 0501537.

473 Deuzé J.L, M. Herman, and R. Santer, « Fourier series expansion of the transfer  
474 equation in the atmosphere-ocean system », *J. Quant. Spectrosc. Radiat. Transfer*, vol. 41, no. 6,  
475 pp. 483-494, 1989.

476 Doxaran D., J.M. Froidefond, P. Castaing and M. Babin, (2009). Dynamics of the turbidity  
477 maximum zone in a macrotidal estuary (the Gironde, France): Observations from field and MODIS  
478 satellite data. *Estuarine, Coastal and Shelf Science* 81, 321–332. ([Get PDF Reprint](#))

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

482 Font, J., Camps, A., Borges, A., Martín-Neira, M., Boutin, J., Reul, N., Mecklenburg, S. (2010).  
483 SMOS: The challenging sea surface salinity measurement from space. *Proceedings of the IEEE*,  
484 98(5), 649-665.

485 Frouin, R., & Pelletier, B. (2014). Bayesian Methodology for ocean color remote sensing.

486 Gordon, H. R. and M. Wang (1994). "Influence of oceanic whitecaps on atmospheric correction  
487 of ocean-color sensors." *Applied Optics* 33(33): 7754-7763.

488 Gordon, H. R., Du, T., & Zhang, T. (1997). Atmospheric correction of ocean color sensors:  
489 analysis of the effects of residual instrument polarization sensitivity. *Applied optics*, 36(27), 6938-  
490 6948.

491 Goyens, C., Jamet, C., & Schroeder, T. (2013). Evaluation of four atmospheric correction  
 492 algorithms for MODIS-Aqua images over contrasted coastal waters. *Remote Sensing of*  
 493 *Environment*, 131, 63-75.

494 Harold W. Sorenson, (1980) "Parameter Estimation: Principles and Problems", Marcel Dekker.

495 Holben, B. N., Eck, T. F., Slutsker, I., Tanre, D., Buis, J. P., Setzer, A. & Smirnov, A. (1998).  
 496 AERONET-A federated instrument network and data archive for aerosol characterization. *Remote*  
 497 *sensing of environment*, 66(1), 1-16.

498 IOCCG (2000): Remote Sensing of Ocean Colour in Coastal, and Other Optically-Complex,  
 499 Waters. Edited by Shubha Sathyendranath, pp. 140.

500 Krasnopolsky, V. M., & Schiller, H. (2003). Some neural network applications in environmental  
 501 sciences. Part I: forward and inverse problems in geophysical remote measurements. *Neural*  
 502 *Networks*, 16(3), 321-334.

503 Jamet, C., Loisel, H., & Dessailly, D. (2012). Retrieval of the spectral diffuse attenuation  
 504 coefficient  $K_d(\lambda)$  in open and coastal ocean waters using a neural network inversion. *Journal of*  
 505 *Geophysical Research: Oceans* (1978–2012), 117(C10).

506 Jamet, C., Loisel, H., Kuchinke, C. P., Ruddick, K., Zibordi, G., & Feng, H. (2011).  
 507 Comparison of three SeaWiFS atmospheric correction algorithms for turbid waters using  
 508 AERONET-OC measurements. *Remote Sensing of Environment*, 115(8), 1955-1965.

509 Jia, S., & Qian, Y. (2009). Constrained nonnegative matrix factorization for hyperspectral  
 510 unmixing. *Geoscience and Remote Sensing, IEEE Transactions on*, 47(1), 161-173

511 Laws, E. A. (1997). *Mathematical methods for oceanographers: An introduction*. John Wiley &  
 512 Sons.

513        Levenberg, K. (1944). A Method for the Solution of Certain Non-Linear Problems in Least  
514        Squares. *Quarterly of Applied Mathematics* 2: 164–168.

515        Lin, C. J. (2007). Projected gradient methods for nonnegative matrix factorization. *Neural*  
516        *computation*, 19(10), 2756-2779.

517        Maritorena, S., Siegel, D. A., & Peterson, A. R. (2002). Optimization of a semianalytical ocean  
518        color model for global-scale applications. *Applied Optics*, 41(15), 2705-2714.

519        McLachlan, G. J. (2004). *Discriminant Analysis and Statistical Pattern Recognition*. Wiley  
520        Interscience.

521        MERIS Level 2 Detailed Processing Model, Doc. no PO-TN-MEL-GS-0006, issue 7, revision 2,  
522        June 2005.

523        Mobley, C. D. (1998). *Hydrolight 4.0 Users Guide*. SEQUOIA SCIENTIFIC INC MERCER  
524        ISLAND WA.

525        Moore, G. F., Aiken, J., Lavender, S. (1999), The atmospheric correction scheme of water  
526        colour and the quantitative retrieval of suspended particulate matter in Case II waters : application  
527        to MERIS, *International Journal of Remote Sensing*, 20, 1713–1733.

528        Morel, A., Huot, Y., Gentili, B., Werdell, P. J., Hooker, S. B., & Franz, B. A. (2007).  
529        Examining the consistency of products derived from various ocean color sensors in open ocean  
530        (Case 1) waters in the perspective of a multi-sensor approach. *Remote Sensing of Environment*,  
531        111(1), 69-88.

532        Morel, A., Claustre, H., Antoine, D., & Gentili, B. (2007). Natural variability of bio-optical  
533        properties in Case 1 waters: attenuation and reflectance within the visible and near-UV spectral  
534        domains, as observed in South Pacific and Mediterranean waters. *Biogeosciences Discussions*, 4(4),  
535        2147-2178.



536 Morel, A., Gentili, B., Chami, M., & Ras, J. (2006). Bio-optical properties of high chlorophyll  
537 Case 1 waters, and of yellow substance-dominated Case 2 waters. *Deep-Sea Research*, 53, 1439 –  
538 1459.

539 Morel, A., and B. Gentili (1996), Diffuse reflectance of oceanic waters 3, Implication of  
540 bidirectionality for the remote-sensing problem, *Applied Optics*, 35, 4850–4862.

541 Park, Y., De Cauwer, V., Nechad, B., & Ruddick, K. (2004). Validation of MERIS water  
542 products for Belgian coastal waters: 2002-2003.

543 Petersen, W., Wehde, H., Krasemann, H., Colijn, F., & Schroeder, F. (2008). FerryBox and  
544 MERIS–Assessment of coastal and shelf sea ecosystems by combining in situ and remotely sensed  
545 data. *Estuarine, Coastal and Shelf Science*, 77(2), 296-307.

546 Petersen, K. B., & Pedersen, M. S. (2008). The matrix cookbook. *Technical University of*  
547 *Denmark*, 7-15.

548 Pope, R. M., & Fry, E. S. (1997). Absorption spectrum (380–700 nm) of pure water. II.  
549 Integrating cavity measurements. *Applied optics*, 36(33), 810-8723.

550 Reynolds, D. A. (1995). Speaker identification and verification using Gaussian mixture speaker  
551 models. *Speech communication*, 17(1), 91-108.

552 Robinson, I. S. (2004). Measuring the oceans from space: the principles and methods of satellite  
553 oceanography. Springer Science & Business Media.

554 Ruddick K., De Cauwer V., Park Y. & Moore G. (2006). Seaborne measurements of near  
555 infrarewater-leaving reflectance: The similarity spectrum for turbid waters. *Limnology and*  
556 *Oceanography*, Vol. 51(2), pp. 1157–1179

557       Ruddick, K. G., De Cauwer, V., Van Mol, B. (2005) Use of the near infrared similarity  
558       reflectance spectrum for the quality control of remote sensing data, in Remote Sensing of the  
559       Coastal Oceanic Environment, edited by Frouin, Robert J., Babin, Marcel, Sathyendranath, Shubha.  
560       Proceedings of the SPIE, Volume 5885, pp. 1–12.

561       Santer, R., Carrere, V., Dubuisson, P., & Roger, J. C. (1999). Atmospheric correction over land  
562       for MERIS. *International Journal of Remote Sensing*, 20(9), 1819-1840.

563       Saulquin, B., Hamdi, A., Gohin, F., Populus, J., Mangin, A., & d'Andon, O. F. (2013).  
564       Estimation of the diffuse attenuation coefficient KdPAR using MERIS and application to seabed  
565       habitat mapping. *Remote Sensing of Environment*, 128, 224-233.

566       Schiller, H., & Doerffer, R. (1999). Neural network for emulation of an inverse model  
567       operational derivation of Case II water properties from MERIS data. *International Journal of*  
568       *Remote Sensing*, 20(9), 1735-1746.

569       Steinmetz, F., Deschamps, P. Y., & Ramon, D. (2011). Atmospheric correction in presence of  
570       sun glint: application to MERIS. *Optics express*, 19(10), 9783-9800.

571       Saulquin, B., Fablet, R., Ailliot, P., Mercier, G., Doxaran, D., Mangin, A., ... & Hembise, O.  
572       (2015). Characterization of Time-Varying Regimes in Remote Sensing Time Series: Application to  
573       the Forecasting of Satellite-Derived Suspended Matter Concentrations. *Selected Topics in Applied*  
574       *Earth Observations and Remote Sensing, IEEE Journal of*, 8(1), 406-417.

575       Wang, M., Son, S., & Harding, L. W. (2009). Retrieval of diffuse attenuation coefficient in the  
576       Chesapeake Bay and turbid ocean regions for satellite ocean color applications. *Journal of*  
577       *Geophysical Research: Oceans (1978–2012)*, 114(C10).

578 Werdell, P. J., & Bailey, S. W. (2005). An improved in-situ bio-optical data set for ocean color  
579 algorithm development and satellite data product validation. *Remote Sensing of Environment*, 98(1),  
580 122-140.

581 Zibordi, G., Mélin, F., Berthon, J. F., Holben, B., Slutsker, I., Giles, D., ... & Seppälä, J. (2009).  
582 AERONET-OC: a network for the validation of ocean color primary products. *Journal of*  
583 *Atmospheric and Oceanic Technology*, 26(8), 1634-1651.

584  
585

## 586 **7 List of caption**

587

588 Figure 1: no caption

589 Fig2 left: In-situ  $\rho_w(\lambda)$  radiometric dataset

590 Fig2 right: MERIS TOA  $\rho_{gc}(\lambda)$  radiometric dataset

591 Fig 3:  $\rho_{aer}(\lambda) + \rho_{coupl}(\lambda)$  coastal reference spectra

592 Fig4 left: Reference spectral signatures

593 Fig4 right:  $\rho_w(\lambda)$  coastal reference spectra

594 Fig5:  $\rho_w(412), \rho_w(442),$

$\rho_w(490), \rho_w(560),$

$\rho_w(665), \rho_w(681)$

595 Fig6: Histogram  $\rho_w(412)$ , Histogram  $\rho_w(442)$ ,

Histogram  $\rho_w(510)$ , Histogram  $\rho_w(560)$

596 Fig7: Histogram  $\rho_w(412)$ , Histogram  $\rho_w(442)$ ,

Histogram  $\rho_w(510)$ , Histogram  $\rho_w(560)$

597 Fig 8: MEETC2  $\rho_w(412)$ , MEETC2  $\rho_w(442)$ , MEETC2  $\rho_w(560)$ , MEETC2  $\rho_w(681)$

MEGS  $\rho_w(412)$ , MEGS  $\rho_w(442)$ , MEGS  $\rho_w(560)$ , MEGS  $\rho_w(681)$

C2R  $\rho_w(412)$ , C2R  $\rho_w(442)$ , C2R  $\rho_w(560)$ , C2R  $\rho_w(681)$

598

599 Fig 9: MEETC2  $\rho_w(412)$ , MEETC2  $\rho_w(442)$ , MEETC2  $\rho_w(560)$ , MEETC2  $\rho_w(681)$

MEGS  $\rho_w(412)$ , MEGS  $\rho_w(442)$ , MEGS  $\rho_w(560)$ , MEGS  $\rho_w(681)$

C2R  $\rho_w(412)$ , C2R  $\rho_w(442)$ , C2R  $\rho_w(560)$ , C2R  $\rho_w(681)$

600

601 Fig 10a: Type 1: SPM dominated  $\rho_w$

602 Fig 10b: Type 2: Cdom dominated  $\rho_w$

603 Fig 10c: Type 3: Chl-a dominated  $\rho_w$

604 Fig 10d: Type 4: pure water dominated  $\rho_w$

605