

Fields of non-linear regression models for atmospheric correction of satellite ocean-color imagery

Robert Frouin and Bruno Pelletier

*Scripps Institution of Oceanography, University of California San Diego, 8810 La
Jolla Shores Drive, La Jolla, California, USA*

*Institut de Mathématiques et de Modélisation de Montpellier, Université
Montpellier II, CC 051 Place Eugène Bataillon, 34095 Montpellier Cedex 5,
France*

Abstract

Remote sensing of ocean color from space, a problem that consists in retrieving spectral marine reflectance from spectral top-of-atmosphere reflectance, is considered as a collection of similar inverse problems continuously indexed by the angular variables influencing the observation process. A general solution is proposed in the form of a field of non-linear regression models over the set T of permitted values for the angular variables, i.e., as a map from T to some function space. Each value of the field is a regression model that performs a direct mapping from the top-of-atmosphere reflectance to the marine reflectance. Since the spectral components of the field take values in the same variable vector space, the retrievals in individual spectral bands are not independent, i.e., the solution is not just a juxtaposition of independent models for each spectral band. A scheme based on ridge functions is developed to approximate this solution to an arbitrary accuracy, and is applied to

the retrieval of marine reflectance in Case 1 waters, which optical properties are only governed by biogenic content. The statistical models are evaluated on synthetic data as well as actual data originating from the SeaWiFS instrument, taking into account noise in the data. Theoretical performance is good in terms of accuracy, robustness, and generalization capabilities, suggesting that the function field methodology might improve atmospheric correction in the presence of absorbing aerosols and provide more accurate estimates of marine reflectance in productive waters. When applied to SeaWiFS imagery acquired off California, the function field methodology gives generally higher estimates of marine reflectance than the standard SeaDAS algorithm, but the values are more realistic.

Key words: Ocean Color, Remote Sensing, Function Fields, Statistical Inverse Problems.

1 Introduction

Standard algorithms to invert remotely sensed data over the oceans basically proceed in two steps: (i) atmospheric correction that yields an estimation of the marine reflectance (Gordon, 1978; Viollier et al., 1980; Gordon and Wang, 1994; Gordon, 1997; Fukushima et al., 1998; Antoine and Morel, 1999; Gao et al., 2000), and (ii) inversion of the marine reflectance using bio-optical models (Gordon et al., 1988; O.Reilly et al., 1998; Morel and Maritorena, 2002). The atmospheric correction has long been recognized as being a difficult and critical operation, since the contribution of the water body may represent only a small fraction ($< 10\%$) of the measured signal (Gordon, 1978; Viollier et al.,

Email addresses: RFrouin@ucsd.edu, Pelletier@math.univ-montp2.fr

(Robert Frouin and Bruno Pelletier).

1980). In some situations (i.e., absorbing aerosols, high chlorophyll content), the expected errors in marine reflectance may go beyond the acceptable 5% limit for biological applications. The uncertainties in atmospheric correction propagate and yield, assuming that the bio-optical model is known, errors in chlorophyll-a concentration that may reach 20% in the presence of little-absorbing aerosols, and larger errors in the case of strongly absorbing aerosols (Gordon, 1997).

The presence of measurement noise and of uncertainties in optical models, associated with the fact that elements of the radiative transfer models are based on inferences from in-situ data sets, anchors the inversion problem in a statistical setting. Several algorithms using nonlinear regression methods have been developed to perform a direct mapping between geophysical and exogenous observable conditions. Such approaches have been used to retrieve the concentrations of oceanic constituents, including phytoplankton chlorophyll-a, dissolved organic matter, and suspended sediment, from the marine reflectance (Keiner and Brown, 1999; Schiller and Doerffer, 1999; Gross et al., 2000), and they have shown improvement over classic algorithms (e.g., band ratios). Applying regression methodologies to drive directly inference about oceanic variables from top-of-atmosphere reflectance is however more complex, due, of course, to the variability induced by the atmosphere, but also to the dependence of the reflectance on the angular geometry. This results in correlations, trends, or even nonlinear dependencies between the regressors (i.e., the top-of-atmosphere reflectance and the angular variables) that are generally difficult to handle for the purpose of statistical modeling (Stahel, 2004, p. 181).

In a recent note, (Pelletier and Frouin, 2004) introduced a general statistical model for inverting satellite data, that fits the ocean color problem as a special

case, and where the peculiarity that the *explanatory variables* are influenced by other, less informative *conditioning variables* is accounted for. The top-of-atmosphere reflectance is considered separately from the angular variables in the modeling. This approach is motivated by the following observation, and we first introduce some notations. Let y be a geophysical variable of interest (e.g., phytoplankton chlorophyll-a concentration, marine reflectance), let \mathbf{t} be the vector of angular variables that characterize the observation process (Sun zenith angle, view zenith angle, and relative azimuth angle), and let \mathbf{x} be a vector of reflectance at d wavelengths. If we let $\mathbf{x}_{\mathbf{t}}$ denote a reflectance acquired in the angular geometry \mathbf{t} , then one may seek to retrieve the value of y independently from measurements of $\mathbf{x}_{\mathbf{t}_1}$ or $\mathbf{x}_{\mathbf{t}_2}$, for any two geometries $\mathbf{t}_1 \neq \mathbf{t}_2$. Thus inverting the top-of-atmosphere reflectance may be considered as a *continuum* of inverse problems, one for each $\mathbf{x}_{\mathbf{t}}$. This leads to a representation of the solution as a *function field* over the set T of permitted values for the angular variables. The underlying idea is to attach an inverse model to each \mathbf{t} which provides an estimation of y from $\mathbf{x}_{\mathbf{t}}$, and where we demand that the attachment be continuous.

The paper is organized as follows. Function fields and fields of nonlinear regression models are defined in Section 2, with an emphasis on fields of linear combinations of shifted ridge functions, selected for reasons of approximation theory. The methodology is applied to the retrieval of marine reflectance, Case 1 waters, in Section 3. It is tested on actual satellite imagery from the Sea-Viewing Wide Field-of-view Sensor (SeaWiFS) in Section 4. Conclusions and a perspective for future work are given in Section 5.

2 Fields of nonlinear regression models

The notations introduced above will be used here and in the remainder of the paper, and y will designate indifferently a real variable or a vector of variables.

Consider a family of parameterized statistical models $f(\mathbf{x}; \theta_n)$ performing a mapping from \mathbb{R}^d to some real vector space, where θ_n is either a scalar parameter or a vector of scalar parameters valued in a set Θ_n . Next consider using one of these models to retrieve y from $\mathbf{x}_{\mathbf{t}}$, for some fixed \mathbf{t} . This can be written as

$$y = f(\mathbf{x}_{\mathbf{t}}; \theta_n) + \epsilon, \quad (1)$$

where ϵ represents the residual of the modeling. At this stage, the task is to infer a value of θ_n from several observations of $\mathbf{x}_{\mathbf{t}}$ and y , typically by minimizing a quadratic loss criterion. Now consider varying \mathbf{t} by a slight amount $\Delta\mathbf{t}$, such that $\mathbf{x}_{\mathbf{t}+\Delta\mathbf{t}}$ is being measured. Then if the above model is satisfactory, one might seek to retrieve y from $\mathbf{x}_{\mathbf{t}+\Delta\mathbf{t}}$ by

$$y = f(\mathbf{x}_{\mathbf{t}+\Delta\mathbf{t}}; \theta_n + \Delta\theta_n) + \epsilon, \quad (2)$$

for some change $\Delta\theta_n$ in the parameters, that is to be inferred from the data.

The extension to all \mathbf{t} in the set T leads to the following model, valid for all $\mathbf{t} \in T$:

$$y = f(\mathbf{x}_{\mathbf{t}}; \theta_n(\mathbf{t})) + \epsilon, \quad (3)$$

where $\theta_n(\mathbf{t})$ is a function of \mathbf{t} , defined on T and valued in Θ_n . For some $\mathbf{t}_0 \in T$, the map $f(\cdot; \theta_n(\mathbf{t}_0))$ is the inverse model *attached* to \mathbf{t}_0 ; it is specified by the *parameter map* $\theta_n(\mathbf{t})$ taken at $\mathbf{t} = \mathbf{t}_0$.

The above display is in fact a particular case of a more general situation. The idea of attaching an inverse model to each \mathbf{t} , be it parameterized or not, can

be made precise by the concept of a function field. A *function field* over a set T is a map defined on T and valued in a space of functions, like for instance the space $\mathcal{C}(X)$ of continuous real-valued functions on the set X . The space of continuous fields over T of continuous functions on X is denoted by $(\mathcal{C}(X))^T$. The value of a function field $\zeta \in \mathcal{C}(X)^T$ is the function $\zeta(\mathbf{t})$, which to each $\mathbf{x} \in X$ associates the real number $\zeta(\mathbf{t})(\mathbf{x})$, i.e., $\zeta(\mathbf{t})$ is the model *attached* to \mathbf{t} . Under mild assumptions on the spaces X and T , there is the homeomorphism $\mathcal{C}(X \times T) \xrightarrow{\cong} \mathcal{C}(X)^T$. This means that, (i) to each $\zeta \in \mathcal{C}(X)^T$, there corresponds the unique map $\zeta^* \in \mathcal{C}(X \times T)$ such that, for all $\mathbf{t} \in T$ and $\mathbf{x} \in X$, we have:

$$\zeta^*(\mathbf{x}, \mathbf{t}) = \zeta(\mathbf{t})(\mathbf{x}), \quad (4)$$

and conversely, and that (ii) the relations $\zeta \rightarrow \zeta^*$ and $\zeta^* \rightarrow \zeta$ are continuous.

For modeling purposes, we shall be led to consider function fields valued in some subset \mathcal{M} of $\mathcal{C}(X)$, or in some nested sequence $\mathcal{M}_0 \subset \dots \subset \mathcal{M}_n \subset \mathcal{M}_{n+1} \subset \dots \subset \mathcal{C}(X)$. The set of such continuous function fields is denoted by \mathcal{M}^T , or by \mathcal{M}_n^T . Consider a set \mathcal{M}_n of functions parameterized by a vector θ_n valued in Θ_n . There is the map i_n carrying a parameter vector θ_n to a function in \mathcal{M}_n , say $f(\cdot; \theta_n) \stackrel{def}{=} i_n(\theta_n)$. A map $\xi_n : T \rightarrow \Theta_n$ defines a function field ζ by $\zeta = i_n \circ \xi_n$. Using a parameter map ξ_n , the model in Eq. (3) may be written as a function field by letting:

$$f(\mathbf{x}; \theta_n(\mathbf{t})) \stackrel{def}{=} f(\mathbf{x}; \xi_n(\mathbf{t})) \quad (5)$$

$$= (i_n \circ \xi_n)(\mathbf{t})(\mathbf{x}) \quad (6)$$

$$= \zeta(\mathbf{t})(\mathbf{x}). \quad (7)$$

The function field formalism allows one to study the mathematical properties of these models, including density and parameterization. In particular, it is important to note that for an arbitrary continuous function field $\zeta \in \mathcal{M}_n^T$,

the existence and uniqueness of a continuous map $\xi_n : T \rightarrow \Theta_n$ such that $\zeta = i_n \circ \xi_n$ is not guaranteed. This depends on the set \mathcal{M}_n , on how \mathcal{M}_n is parameterized, i.e., on the properties of i_n . For details, see (Pelletier, 2004).

In the absence of information on the nature of the inverse relationship between y , \mathbf{x} , and \mathbf{t} , there is an interest in considering dense sequences of sets, i.e., such that $\overline{\cup_n \mathcal{M}_n} = \mathcal{C}(X)$, since in this case, we have $\overline{\cup_n \mathcal{M}_n^T} = (\mathcal{C}(X))^T$ (Pelletier, 2004). The sets spanned by functions of the ridge form have this property. A *ridge function* on \mathbf{R}^d is a function of the form $h(\mathbf{a} \cdot \mathbf{x})$, where h is a function on \mathbf{R} , and where $\mathbf{a} \cdot \mathbf{x}$ is the standard inner product on \mathbf{R}^d . Ridge function approximation refers to approximation by linear combinations of n ridge functions, for some integer n , i.e., by functions of the form

$$f(\mathbf{x}) = \sum_{i=1}^n c_i h(\mathbf{a}_i \cdot \mathbf{x}), \quad (8)$$

or by linear combinations of *shifted ridge functions*:

$$f(\mathbf{x}) = \sum_{i=1}^n c_i h(\mathbf{a}_i \cdot \mathbf{x} + b_i), \quad (9)$$

where the scalars b_i are the shifts. The \mathbf{a}_i , b_i , and c_i , constitute the parameter vector θ_n , taking values in the set $\Theta_n \stackrel{\text{def}}{=} \prod_{i=1}^n \mathbf{R}^d \times \mathbf{R} \times \mathbf{R}$. The set spanned by functions of the form given by Eq. (9) will be denoted by \mathcal{M}_n , and the function fields taking values in it will be called *ridge function fields*. This set is over-parameterized (i_n is not injective), but ridge function fields have the following interesting property (Pelletier, 2004): the union over n of ridge function fields induced by continuous parameter maps $\xi_n : T \rightarrow \Theta_n$ is dense in $\mathcal{C}(X)^T$.

Therefore, ridge function fields are used to define statistical models explaining y from \mathbf{x} and \mathbf{t} . Introducing Eq. (9) in Eq. (3) leads to the following statistical model, expressed in its expanded form:

$$y = \zeta(\mathbf{t})(\mathbf{x}) + \epsilon \quad (10)$$

$$= \sum_{i=1}^n c_i(\mathbf{t})h(\mathbf{a}_i(\mathbf{t}) \cdot \mathbf{x} + b_i(\mathbf{t})) + \epsilon. \quad (11)$$

Adjusting the statistical model to a data set $\mathcal{D} = \{(\mathbf{x}_i, \mathbf{t}_i, y_i); i = 1, \dots, N\}$ consists in estimating the free parameters of the field, i.e., the maps $\mathbf{a}_i(\mathbf{t})$, $b_i(\mathbf{t})$, and $c_i(\mathbf{t})$. Since these maps have infinitely many degrees of freedom, we proceed by multi-linear interpolation on a regular grid covering T for their constructions. The adjustment is considered in the least square sense, and the minimization of the mean squared error is carried out by use of a dedicated stochastic gradient descent algorithm.

3 Retrieval of marine reflectance

A statistically significant data set of about 62,000 joint samples of top-of-atmosphere (TOA) and marine reflectance was generated using a coupled ocean-atmosphere radiative transfer code (Vermote et al., 1997). This code takes into account the essential physics of the problem, namely gaseous absorption, scattering by molecules, scattering and absorption by aerosols, molecule-aerosol interactions, reflection by the surface, molecule-surface and aerosol-surface interactions, and backscattering by whitecaps and the water body. The top-of-atmosphere reflectance was simulated in spectral bands centered at 412, 443, 490, 510, 555, 670, 765, and 865 nm (case of SeaWiFS) and corrected for molecular effects. The marine reflectance, ρ_w , defined as the product of π and water-leaving radiance normalized by surface irradiance, was simulated according to (Morel and Maritorena, 2002) in the first 6 spectral bands (visible). It was assumed to be equal to zero in the other spectral bands (near-infrared). The reflectance model, valid for Case 1 waters, only depends on

chlorophyll-a concentration, which was varied from 0.03 to 30 mgm^{-3} . The marine reflectance was considered isotropic in the simulations. This is not limiting since a unique statistical model is attached to each angular geometry (see below). In other words, the inverse problem would not be more complicated to solve if bidirectional characteristics were included. The data set encompasses the major sources of variability and includes three aerosol models in varied mixtures (maritime, continental, urban). Specifically, Sun and view zenith angles were varied from 0 and 60 degrees and 0 to 50 degrees, respectively, relative azimuth angles from 0 to 180 degrees, wind speed of 0 to 15 ms^{-1} , and aerosol optical thickness from 0.05 to 0.5. All the cases with Sun glint reflectance above 0.04 were discarded. For each remaining case, the coupled radiative transfer code was run once with molecules and aerosols and a non-black water body and once with only molecules and a black body. The difference between the output of the two runs, i.e., a TOA reflectance corrected for molecular effects, was used to construct the function fields.

The synthetic data set was randomly split into two sub data sets \mathcal{D}_e^0 and \mathcal{D}_v^0 , used for adjusting and validating the statistical inverse models, respectively. Noisy versions of the data sets, denoted by \mathcal{D}_e^ν and \mathcal{D}_v^ν , of \mathcal{D}_e^0 and \mathcal{D}_v^0 were also created for robustness assessments, by adding some amount of noise to the TOA reflectance. The selected noise is expressed as the sum of correlated and uncorrelated components, and is defined by:

$$\tilde{\mathbf{x}} = \mathbf{x} + \nu^c \mathbf{x} + \left(\nu_1^{uc} x^1, \dots, \nu_8^{uc} x^8 \right)^t, \quad (12)$$

where $\tilde{\mathbf{x}}$ is a noisy version of a TOA reflectance \mathbf{x} , and where $\nu^c, \nu_1^{nc}, \dots, \nu_8^{nc}$ are random variables uniformly distributed in the interval $[-\frac{\nu}{200}; \frac{\nu}{200}]$, where ν is the total amount of noise, in percent.

Now let $\mathbf{x} \in \mathbb{R}^d$ and $\mathbf{y} \in \mathbb{R}^{d'}$ be vectors of top-of-atmosphere and marine reflectance, respectively. Let $\mathbf{t} = (\cos \theta_s, \cos \theta_v, \cos \Delta\varphi)^t$ be the vector of cosines of the angular variables, taking values in the set $T = [\frac{1}{2}; 1] \times [\frac{1}{2}; 1] \times [-1; 1]$. Let

$$E(\mathbf{a}_1, \dots, \mathbf{a}_n, b_1, \dots, b_n) = \text{span} \{h(\mathbf{a}_1 \cdot \mathbf{x} + b_1), \dots, h(\mathbf{a}_n \cdot \mathbf{x} + b_n)\}, \quad (13)$$

where $h : \mathbb{R} \rightarrow \mathbb{R}$, i.e., $E(\mathbf{a}_1, \dots, \mathbf{a}_n, b_1, \dots, b_n)$ is the vector space spanned by the linear combinations of the $h(\mathbf{a}_i \cdot \mathbf{x} + b_i)$. For the retrieval of \mathbf{y} from \mathbf{x} , we consider function fields $\zeta : T \rightarrow E^{d'}(\mathbf{a}_1, \dots, \mathbf{a}_n, b_1, \dots, b_n)$ such that

$$\zeta(\mathbf{t})^j(\mathbf{x}) = \sum_{i=1}^n c_i^j(\mathbf{t}) h(\mathbf{a}_i \cdot \mathbf{x} + b_i), \quad (14)$$

for $j = 1, \dots, d'$, and the statistical model relative to the j^{th} component of \mathbf{y} is written as

$$y^j = \zeta(\mathbf{t})^j(\mathbf{x}) + \epsilon^j. \quad (15)$$

The functions c_i^j are defined by multi-linear interpolation on a $2 \times 2 \times 3$ regular grid covering T , i.e., they are piecewise-differentiable functions, each of whose depends on 12 scalars. These scalars, together with the \mathbf{a}_i and the b_i constitute the free parameters of the model. The model is adjusted by minimizing the sum of the squared errors \mathcal{E} on \mathcal{D}_e^0 defined by

$$\mathcal{E} = |\mathcal{D}_e^0|^{-1} \sum_{(\mathbf{x}_k, \mathbf{t}_k, \mathbf{y}_k) \in \mathcal{D}_e^0} \|\zeta(\mathbf{t}_k)(\mathbf{x}_k) - \mathbf{y}_k\|^2, \quad (16)$$

where $\|\cdot\|$ is the usual Euclidean norm on $\mathbb{R}^{d'}$. In practice, the components of \mathbf{y} are normalized linearly between -1 and 1 . One might think of ζ as a d' -tuple of fields of real-valued functions over T . But note that since the d' components of ζ are taking values in the same variable vector space (i.e., in $E(\mathbf{a}_1, \dots, \mathbf{a}_n, b_1, \dots, b_n)$), this approach is not equivalent to d' independent retrievals, on a component per component basis. The minimization of \mathcal{E} is

carried out using a stochastic gradient descent procedure.

Via the simulations, the sufficient number of $n = 15$ basis functions was obtained, and three fields ζ^0 , ζ^1 , and ζ^2 of this kind, i.e., with $n = 15$, were constructed. The first one, ζ^0 , was adjusted on \mathcal{D}^0 . In the case of ζ^1 and ζ^2 , we added a noise amount of 1% and 2%, respectively, to the TOA reflectance during the execution of the minimization procedure, where the added noise satisfies Eq. (12). Note that this is not equivalent to adjusting the fields on \mathcal{D}_e^1 and \mathcal{D}_e^2 . The noise amount corresponds typically to 10 and 20% of the marine signal, i.e., of the signal to retrieve.

The theoretical performance of the regression models is summarized in Table. 1, which gives, on a component-per-component basis, the Root Mean Squared error (RMS) and the Relative Root Mean Squared error (RMSR) of ζ^ν evaluated on $\mathcal{D}_e^{\nu'}$ and $\mathcal{D}_e^{\nu''}$, for all combinations of $\nu, \nu', \nu'' = 0, 1, 2$. These statistics show that the fields present good validation properties, and that the robustness to input noise is improved by the addition of some amount of noise during their construction. In parallel, it may be remarked that, for a noise level of $\nu\%$ on the TOA reflectance, the best retrievals are achieved by ζ^ν . This illustrates the importance of the noise distribution, defined prior to the construction of the models. This point will be further discussed in the next section. The marine reflectance estimations are accurate over the whole range of values, as depicted by Fig. (1) and Fig. (2), which give plots of estimated versus expected reflectance. For the model ζ^0 adjusted on \mathcal{D}^0 , the RMSR ranges from 0.7% at 510 nm to 3.3% at 412 nm, and for the model ζ^1 adjusted on \mathcal{D}^1 from 1.0% and 4.1%, respectively. The average error increases with the aerosol optical thickness (τ_a). More precisely, the residuals spread themselves with increasing τ_a , as shown in the plots of several conditional quantiles of

the residuals distributions as a function of τ_a (Fig. 3). However, the errors are weakly dependent on the aerosol type and no trends are revealed, as evidenced in Fig. (4).

Ratios of marine reflectance are commonly used to estimate chlorophyll-a concentration (e.g., O'Reilly et al., 1998). The theoretical relation between ratios of reflectance at 443 and 555 nm and at 490 and 555 nm is slightly degraded when using the reflectance estimated by the function fields (Fig. 5). The resulting RMS in chlorophyll-a concentration, computed in natural logarithm scale, which gives approximately the fractional error, is only 0.07 (7%) and 0.04 (4%), respectively, when using ζ^0 on \mathcal{D}^0 , and 0.10 (10%) and 0.09 (9%) when using ζ^1 on \mathcal{D}^1 (Fig. 6). Unlike the ratio at 443 and 555 nm, which yields larger relative errors at high chlorophyll-concentration, the ratio at 490 and 555 nm exhibits similar accuracy in the entire range of chlorophyll-a concentration. This ratio, therefore, is preferred when using function field methodology to retrieve marine reflectance.

4 Application to SeaWiFS imagery

The function field methodology was applied to a SeaWiFS image acquired on November 19, 2002 (day 323) above Southern California. For this image, which contains above 78,000 useful pixels, the chlorophyll-a concentration ([Chl-a]) ranged from 0.05 to 5 mgm^{-3} , as retrieved by the SeaDAS algorithm, and was spatially distributed in typical marine structures, with relatively high values near the coast and low values offshore. The SeaDAS aerosol optical thickness was generally low, with a maximum value of about 0.1 at 865 nm. Surface wind was calm everywhere, with speed values of less than 5 ms^{-1} . The exper-

imental procedure included the following steps. First, the marine reflectance was retrieved from the SeaWiFS TOA reflectance with a function field, of the kind described in Section 2. Second, the retrieved marine reflectance was used to derive [Chl-a] by applying the SeaDAS bio-optical algorithm based on reflectance ratios. Finally, the [Chl-a] values obtained in step 2 were compared with those obtained from marine reflectance retrieved with the SeaDAS atmospheric correction algorithm.

In a first bunch of experiments, the three fields ζ^{ν} presented above, constructed with up to 2% of noise, were used to retrieve the marine reflectance. The function field values were substantially different, generally higher than the corresponding SeaDAS values, yielding unrealistic [Chl-a] estimates. More specifically, the root-mean-squared differences with respect to SeaDAS were above 78% for those fields (best result obtained for ζ^2). This is explained by the large level of noise on the TOA reflectance, which was estimated by plugging into the radiative transfer code the SeaDAS values for both [Chl-a] and aerosols parameters, and by comparing the results with the actual (i.e., measured) SeaWiFS TOA reflectance. Since the SeaDAS retrievals are consistent with the measured TOA reflectance, the differences in TOA reflectance reflect differences in radiative transfer modeling. At 412 nm, for example, the root-mean-squared difference between simulated and measured TOA reflectance exceeded 14%. Such large differences obviously cannot be compensated by a general noise scheme, suggesting that the noise distribution should be taken into account in the construction of the function fields, which was accomplished as follows.

Based on 2000 pixels randomly selected in the image, an estimation of the noise distribution was inferred by using the plug-in approach just described. More

precisely, the empirical distribution $\mathbb{P}_n = \frac{1}{n} \sum_{i=1}^n \delta_{\mathbf{e}_i}$ was considered, where $\mathbf{e}_i \in \mathbb{R}^d$ is the vector of difference between the measured and computed TOA reflectance for the i^{th} pixel, and where $\delta_{\mathbf{x}}$ is the probability distribution that is degenerate at \mathbf{x} . Using \mathbb{P}_n as the noise distribution, a function field ζ^* of the same characteristics as the fields ζ^ν was constructed. The marine reflectance retrieved by ζ^* is displayed in Fig. (7). In the productive waters North of the Channel Islands (upper left part of the images, values are lower than in the surrounding waters at wavelengths below 510 nm, where phytoplankton absorption is effective, and higher above 510 nm, where particle backscattering dominates. Compared with the marine reflectance retrieved by SeaDAS, displayed in Fig. (8) with the same color table, the marine reflectance retrieved by ζ^* is generally higher. The SeaDAS values are sometimes negative at 412 and 443 nm, especially near the coast, and spatial variability is larger for SeaDAS at 510 nm. The differences between the marine reflectance obtained from SeaDAS and ζ^* may also be seen on the density estimates in Fig. (9). On average the marine reflectance obtained from ζ^* is higher by 0.0164 at 412 nm, 0.0093 at 443 nm, 0.0037 at 490 nm, 0.0010 at 510 nm, 0.0009 at 555 nm, and 0.0002 at 670 nm. The values are more narrowly distributed with ζ^* , especially at 510 nm, which may be explained, at least partly, by the averaged bio-optical model used in the construction of ζ^* , which does not include phytoplankton type variability. The SeaDAS atmospheric correction algorithm, on the other hand, does not make any assumption about bio-optical properties of the water body. Plots of 1,000 marine spectra retrieved by SeaDAS (top panel) and by ζ^* (bottom panel) from the same TOA reflectance are reproduced in Fig. (10). Clearly ζ^* improves the marine reflectance estimation over SeaDAS, as the marine spectra are much more realistic than the SeaDAS ones. More specifically, they agree with the model of (Morel and Maritorena, 2002),

known to be valid in the study region (Mitchell and Kahru, 1998; Kahru and Mitchell, 1999). Nevertheless, the SeaDAS [Chl-a] is globally validated and may be reasonably trusted, even in cases of large errors in marine reflectance (Hooker and McClain, 2000; McClain et al., 2004).

The resulting [Chl-a] image is given in Fig. (11). In comparison with SeaDAS, the [Chl-a] values obtained by ζ^* are lower on average, with a negative bias of -13.8% . The two spatial fields differ by 19.6% when measured with the root-mean-squared error in natural logarithm, which corresponds approximately to the root-mean-squared fractional error. The relative differences are large in magnitude in the clearest waters (higher values with ζ^*) and in the most productive waters (lower values with ζ^*). The 19.6% difference, however, does not exceed the root-mean-squared uncertainty attached to the SeaDAS processing chain, recently evaluated at 33% (Werdell and Bailey, 2005). It may also be noted that the coherence of the retrieved spatial fields is rather good, not only individually, but also when comparing the locations of the marine structures in the image. Furthermore, the spatial fields of marine reflectance retrieved by the non-linear regression models are less noisy than those obtained from the SeaDAS algorithm, especially at 555 and 670 nm (see Fig. (7) and Fig. (8)). Hence, in this case of large levels of noise, using an appropriate noise distribution in the construction of a function field allows one to obtain a realistic marine reflectance that leads to a chlorophyll-a concentration comparable with the SeaDAS value.

5 Conclusions

Fields of non-linear regression models emerge as solutions to a continuum of similar statistical inverse problems. This formulation of the remote-sensing problem, that does involve explanatory covariates influenced by the angular geometry, allows one to de-correlate the useful information from the effect of the observation process. In this sense, the function fields presented herein well match the physical characteristics of the problem, while taking benefit from an efficient statistical modeling technique, namely nonlinear regression. This probably explains the good theoretical results, in terms of accuracy, robustness, and in particular generalization, obtained with this methodology for the retrieval of the marine reflectance. In particular, the retrievals are accurate over the entire range of marine reflectance and in the presence of strongly absorbing aerosols.

When applied to actual SeaWiFS data, the function fields significantly improved the marine reflectance estimations with respect to SeaDAS, providing more realistic values, especially in the blue. As described above, the function field construction required an estimation of the noise distribution. Indeed this is necessary only because the noise level is large in this case. For noise levels up to several percent, resorting to an estimation of the noise distribution may be avoided by using a general noise scheme, with correlated and uncorrelated components, as revealed by the experiments on synthetic data. Nevertheless in any case, using an estimated noise distribution would certainly improve performance, as the fields approximate the conditional means of one geophysical variable (or vector) given the TOA reflectance. Thus studying the noise statistical properties, where noise is to be understood as overall differ-

ence between measurements from one given instrument and predictions from one given radiative transfer code, constitutes an interesting and important perspective. In this direction, one idea would be to make an inference on the basis of concomitant in-situ measurements of atmospheric and oceanic parameters and satellite reflectance. Naturally, this would require the collected data to be statistically significant, and therefore a plug-in approach may represent a reasonable alternative.

In this study, we considered the simultaneous retrieval of the marine reflectance at 6 wavelengths by adjusting a field of nonlinear regression models on a simulated data set. In the generation of this synthetic data, we used a marine reflectance model that solely depends on the chlorophyll-a concentration, so the complexity of the problem could have been reduced by directly estimating the chlorophyll-a concentration, as addressed in our earlier work (Pelletier and Frouin, 2005). There is however an interest in developing estimation methodologies for the marine reflectance, even in this simple case. The reason is well known: the marine reflectance might give access to other variables than chlorophyll-a concentration, for example yellow substance absorption and sediment concentration. In this direction, the function field methodology appears to have the potential for improving the quality of the marine reflectance retrievals in optically complex, i.e., Case 2 waters compared with existing techniques. Of course, accuracy will be governed by the bijective character of the relation between TOA reflectance and marine reflectance. In this context, there is the especially interesting question to determine whether or not a direct estimation of one geophysical variable of interest is more efficient than a 2-step estimation, i.e., by passing via the marine reflectance.

Acknowledgements

This work was supported by the National Aeronautics and Space Administration under Grant No. NNG05GR20G, by the National Science Foundation under Grant No. OCE-0417748, and by the Institut de Mathématiques et de Modélisation de Montpellier, Université Montpellier II, France. We are grateful to Mr. John McPherson, Scripps Institution of Oceanography, for technical assistance.

References

- Antoine, D., Morel, A., 1999. A multiple scattering algorithm for atmospheric correction of remotely sensed ocean colour (meris instrument): Principle and implementation for atmospheres carrying various aerosols including absorbing ones. *Int. J. Remote Sens.* 20, 1875–1916.
- Fukushima, H., Higurashi, A., Mitomi, Y., Nakajima, T., Noguchi, T., Tanaka, T., Toratani, M., 1998. Correction of atmospheric effects on adeos/octs ocean color data: Algorithm description and evaluation of its performance. *J. Oceanogr.* 54, 417–430.
- Gao, B.-C., Montes, M., Ahmad, Z., Davis, C., 2000. Atmospheric correction algorithm for hyper-spectral remote sensing of ocean color from space. *Appl. Opt.* 39, 887–896.
- Gordon, H., 1978. Removal of atmospheric effects from satellite imagery of the oceans. *Appl. Opt.* 17, 1631–1636.
- Gordon, H., 1997. Atmospheric correction of ocean color imagery in the earth observing system era. *J. Geophys. Res.* 102, 17081–17118.
- Gordon, H., Brown, O., Evans, R., Brown, J., Smith, R., Baker, K., 1988. A

- semi-analytical radiance model of ocean color. *J. Geophys. Res.* 93, 10909–10924.
- Gordon, H., Wang, M., 1994. Retrieval of water-leaving radiance and aerosol optical thickness over the oceans with seawifs: a preliminary algorithm. *Appl. Opt.* 33, 443–452.
- Gross, L., Thiria, S., Frouin, R., Mitchell, B., 2000. Artificial neural networks for modeling the transfer function between marine reflectance and phytoplankton pigment concentration. *J. Geophys. Res.* 105, 3483–3495.
- Hooker, S., McClain, C., 2000. The calibration and validation of seawifs data. *Prog. Oceanogr.* 47, 427–465.
- Kahru, M., Mitchell, B., 1999. Empirical chlorophyll algorithm and preliminary seawifs validation for the california current. *Int. J. Remote Sen.* 20, 3423–3430.
- Keiner, L., Brown, C., 1999. Estimating oceanic chlorophyll concentrations with neural networks. *Int. J. Rem. Sen.* 20, 189–194.
- McClain, C., Feldman, G., Hooker, S., 2004. An overview of the seawifs project and strategies for producing a climate research quality global ocean bio-optical time series. *Deep-Sea Res.*, In press.
- Mitchell, B., Kahru, M., 1998. Algorithms for seawifs standard products developed with the calcofi bio-optical data set. *Tech. Rep. 30*, California Cooperative Fisheries Investigation.
- Morel, A., Maritorena, S., 2002. Bio-optical properties of oceanic waters: A reappraisal. *J. Geophys. Res.* 106, 7163–7180.
- O.Reilly, J., Maritorena, S., Mitchell, B., Siegel, D., Carder, K., Garver, S., Kahru, M., McClain, C., 1998. Ocean colour chlorophyll algorithms for seawifs. *J. Geophys. Res.* 103, 24937–24953.
- Pelletier, B., 2004. Approximation by ridge function fields over compact sets.

- Journal of Approximation Theory 129, 230–239.
- Pelletier, B., Frouin, R., 2004. Fields of nonlinear regression models for inversion of satellite data. *Geophys. Res. Lett.* 31, 116304.
- Pelletier, B., Frouin, R., 2005. Remote sensing of phytoplankton chlorophyll-a concentration via ridge function fields. *Applied Optics*, In Press.
- Schiller, H., Doerffer, R., 1999. Neural network for emulation of an inverse model - operational derivation of case ii properties from meris data. *Int. J. Rem. Sen.* 20, 1735–1746.
- Stahel, W., 2004. Purposes and strategies in regression analysis. *Journal of Statistical Planning and Inference* 122, 175–186.
- Vermote, E., Tanré, D., Deuzé, J.-L., Herman, M., Morcrette, J.-J., 1997. Second simulation of the satellite signal in the solar spectrum: An overview. *IEEE Trans. Geosci. Remote Sens.* 35, 675–686.
- Viollier, M., Tanré, D., Deschamps, P.-Y., 1980. An algorithm for remote sensing of water color from space. *Boundary-Layer Meteor.* 18, 247–267.
- Werdell, P., Bailey, S., 2005. An improved bio-optical data set for ocean color algorithm development and satellite data product validation. *Rem. Sen. Environ.* 98 (1), 122–140.

		FIELD ζ^0					
	λ (nm)	412	443	490	510	555	670
\mathcal{D}_e^0	RMS	$4.3 \cdot 10^{-4}$	$2.3 \cdot 10^{-4}$	$0.7 \cdot 10^{-4}$	$0.3 \cdot 10^{-4}$	$0.5 \cdot 10^{-4}$	$0.2 \cdot 10^{-4}$
	RMSR (%)	2.73	1.71	0.64	0.38	0.81	2.04
\mathcal{D}_v^0	RMS	$4.4 \cdot 10^{-4}$	$2.3 \cdot 10^{-4}$	$0.7 \cdot 10^{-4}$	$0.3 \cdot 10^{-4}$	$0.5 \cdot 10^{-4}$	$0.2 \cdot 10^{-4}$
	RMSR (%)	2.92	1.81	0.67	0.40	0.80	2.04
\mathcal{D}_e^1	RMS	$10.6 \cdot 10^{-4}$	$6.3 \cdot 10^{-4}$	$2.3 \cdot 10^{-4}$	$1.4 \cdot 10^{-4}$	$1.6 \cdot 10^{-4}$	$0.7 \cdot 10^{-4}$
	RMSR (%)	5.86	4.46	2.54	1.74	2.22	5.81
\mathcal{D}_v^1	RMS	$10.7 \cdot 10^{-4}$	$6.4 \cdot 10^{-4}$	$2.4 \cdot 10^{-4}$	$1.4 \cdot 10^{-4}$	$1.6 \cdot 10^{-4}$	$0.8 \cdot 10^{-4}$
	RMSR (%)	6.21	4.69	2.67	1.81	2.24	5.59
\mathcal{D}_e^2	RMS	$19.9 \cdot 10^{-4}$	$12.0 \cdot 10^{-4}$	$4.5 \cdot 10^{-4}$	$2.7 \cdot 10^{-4}$	$3.0 \cdot 10^{-4}$	$1.5 \cdot 10^{-4}$
	RMSR (%)	10.89	8.41	4.86	3.37	4.25	12.64
\mathcal{D}_v^2	RMS	$19.6 \cdot 10^{-4}$	$11.8 \cdot 10^{-4}$	$4.6 \cdot 10^{-4}$	$2.8 \cdot 10^{-4}$	$3.1 \cdot 10^{-4}$	$1.5 \cdot 10^{-4}$
	RMSR (%)	11.34	8.68	5.05	3.52	4.23	12.16
		FIELD ζ^1					
\mathcal{D}_e^0	RMS	$5.9 \cdot 10^{-4}$	$3.2 \cdot 10^{-4}$	$1.0 \cdot 10^{-4}$	$0.5 \cdot 10^{-4}$	$0.7 \cdot 10^{-4}$	$0.3 \cdot 10^{-4}$
	RMSR (%)	3.00	2.05	0.99	0.64	1.04	2.43
\mathcal{D}_v^0	RMS	$6.0 \cdot 10^{-4}$	$3.2 \cdot 10^{-4}$	$1.0 \cdot 10^{-4}$	$0.5 \cdot 10^{-4}$	$0.7 \cdot 10^{-4}$	$0.3 \cdot 10^{-4}$
	RMSR (%)	3.33	2.22	1.05	0.68	1.04	2.39
\mathcal{D}_e^1	RMS	$8.4 \cdot 10^{-4}$	$4.8 \cdot 10^{-4}$	$1.6 \cdot 10^{-4}$	$0.8 \cdot 10^{-4}$	$1.2 \cdot 10^{-4}$	$0.5 \cdot 10^{-4}$
	RMSR (%)	4.08	3.02	1.62	1.02	1.68	4.00
\mathcal{D}_v^1	RMS	$8.5 \cdot 10^{-4}$	$4.9 \cdot 10^{-4}$	$1.7 \cdot 10^{-4}$	$0.9 \cdot 10^{-4}$	$1.2 \cdot 10^{-4}$	$0.5 \cdot 10^{-4}$
	RMSR (%)	4.15	3.09	1.69	1.07	1.71	3.92
\mathcal{D}_e^2	RMS	$13.5 \cdot 10^{-4}$	$8.0 \cdot 10^{-4}$	$2.8 \cdot 10^{-4}$	$1.4 \cdot 10^{-4}$	$2.0 \cdot 10^{-4}$	$0.9 \cdot 10^{-4}$
	RMSR (%)	6.29	4.89	2.71	1.67	2.85	6.77
\mathcal{D}_v^2	RMS	$13.2 \cdot 10^{-4}$	$7.8 \cdot 10^{-4}$	$2.8 \cdot 10^{-4}$	$1.4 \cdot 10^{-4}$	$2.1 \cdot 10^{-4}$	$0.9 \cdot 10^{-4}$
	RMSR (%)	7.18	5.27	2.82	1.74	2.82	6.71
		FIELD ζ^2					
\mathcal{D}_e^0	RMS	$6.7 \cdot 10^{-4}$	$3.7 \cdot 10^{-4}$	$1.3 \cdot 10^{-4}$	$0.7 \cdot 10^{-4}$	$0.9 \cdot 10^{-4}$	$0.4 \cdot 10^{-4}$
	RMSR (%)	3.82	2.56	1.32	0.89	1.31	3.62
\mathcal{D}_v^0	RMS	$6.7 \cdot 10^{-4}$	$3.7 \cdot 10^{-4}$	$1.3 \cdot 10^{-4}$	$0.7 \cdot 10^{-4}$	$1.0 \cdot 10^{-4}$	$0.4 \cdot 10^{-4}$
	RMSR (%)	4.01	2.68	1.40	0.95	1.34	3.57
\mathcal{D}_e^1	RMS	$8.5 \cdot 10^{-4}$	$4.9 \cdot 10^{-4}$	$1.7 \cdot 10^{-4}$	$0.9 \cdot 10^{-4}$	$1.3 \cdot 10^{-4}$	$0.5 \cdot 10^{-4}$
	RMSR (%)	4.60	3.27	1.75	1.13	1.76	4.49
\mathcal{D}_v^1	RMS	$8.6 \cdot 10^{-4}$	$5.0 \cdot 10^{-4}$	$1.8 \cdot 10^{-4}$	$0.9 \cdot 10^{-4}$	$1.3 \cdot 10^{-4}$	$0.6 \cdot 10^{-4}$
	RMSR (%)	4.76	3.40	1.86	1.20	1.83	4.59
\mathcal{D}_e^2	RMS	$12.5 \cdot 10^{-4}$	$7.4 \cdot 10^{-4}$	$2.7 \cdot 10^{-4}$	$1.3 \cdot 10^{-4}$	$1.9 \cdot 10^{-4}$	$0.9 \cdot 10^{-4}$
	RMSR (%)	6.42	4.83	2.64	1.62	2.72	6.63
\mathcal{D}_v^2	RMS	$12.3 \cdot 10^{-4}$	$7.4 \cdot 10^{-4}$	$2.7 \cdot 10^{-4}$	$1.4 \cdot 10^{-4}$	$2.0 \cdot 10^{-4}$	$0.9 \cdot 10^{-4}$
	RMSR (%)	6.72	5.02	2.78	1.71	2.70	6.49

Table 1
Root Mean Squared error (RMS) and Root Mean Squared Relative error (RMSR) for the models ζ^0 , ζ^1 , and ζ^2 , evaluated on the construction and validation sets (\mathcal{D}_e^0 and \mathcal{D}_v^0), and on 1% and 2% noisy versions of them (\mathcal{D}_e^1 , \mathcal{D}_v^1 , \mathcal{D}_e^2 , and \mathcal{D}_v^2).

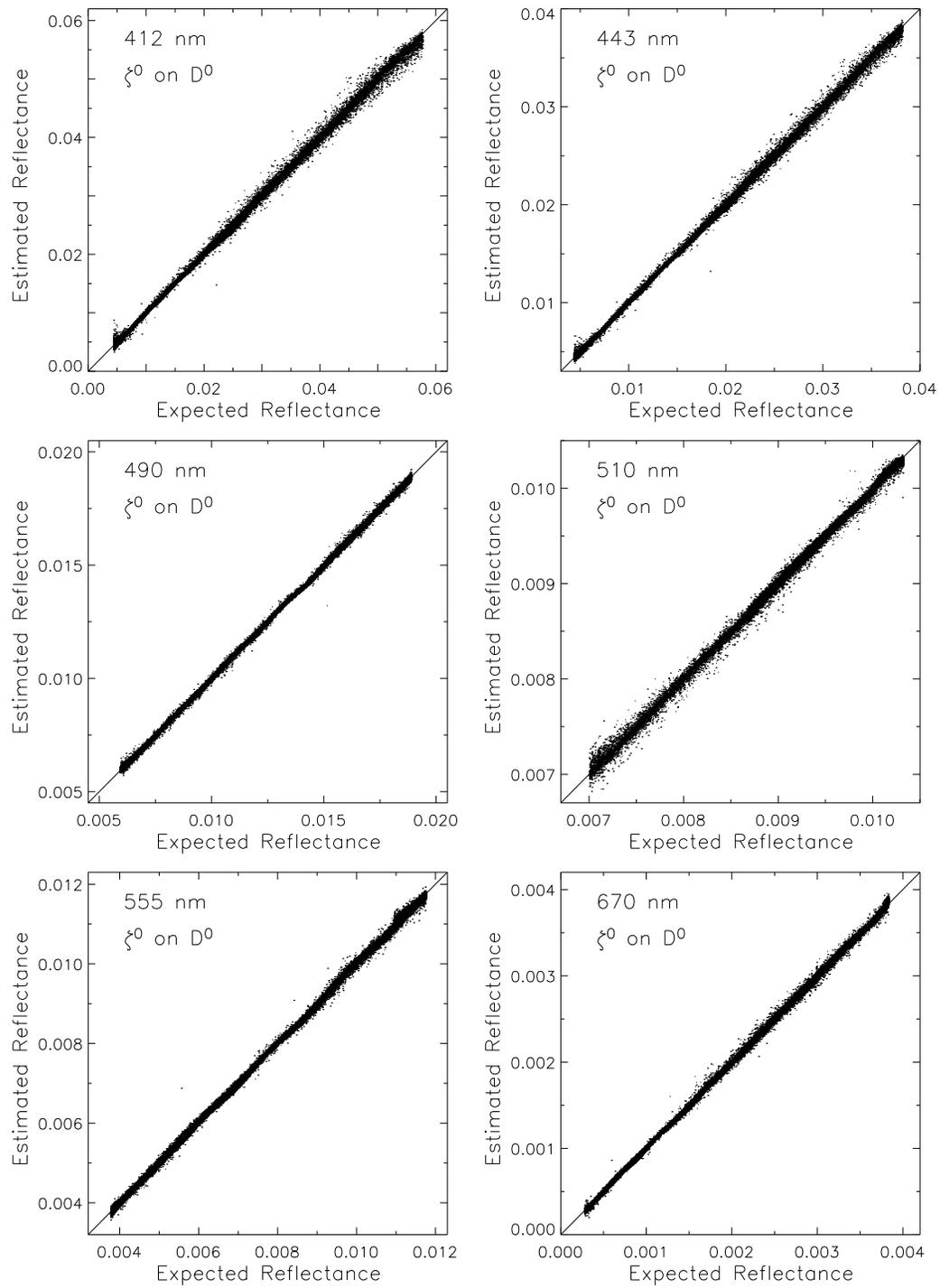


Fig. 1. Expected versus estimated marine reflectance for model ζ^0 on non-noisy data.

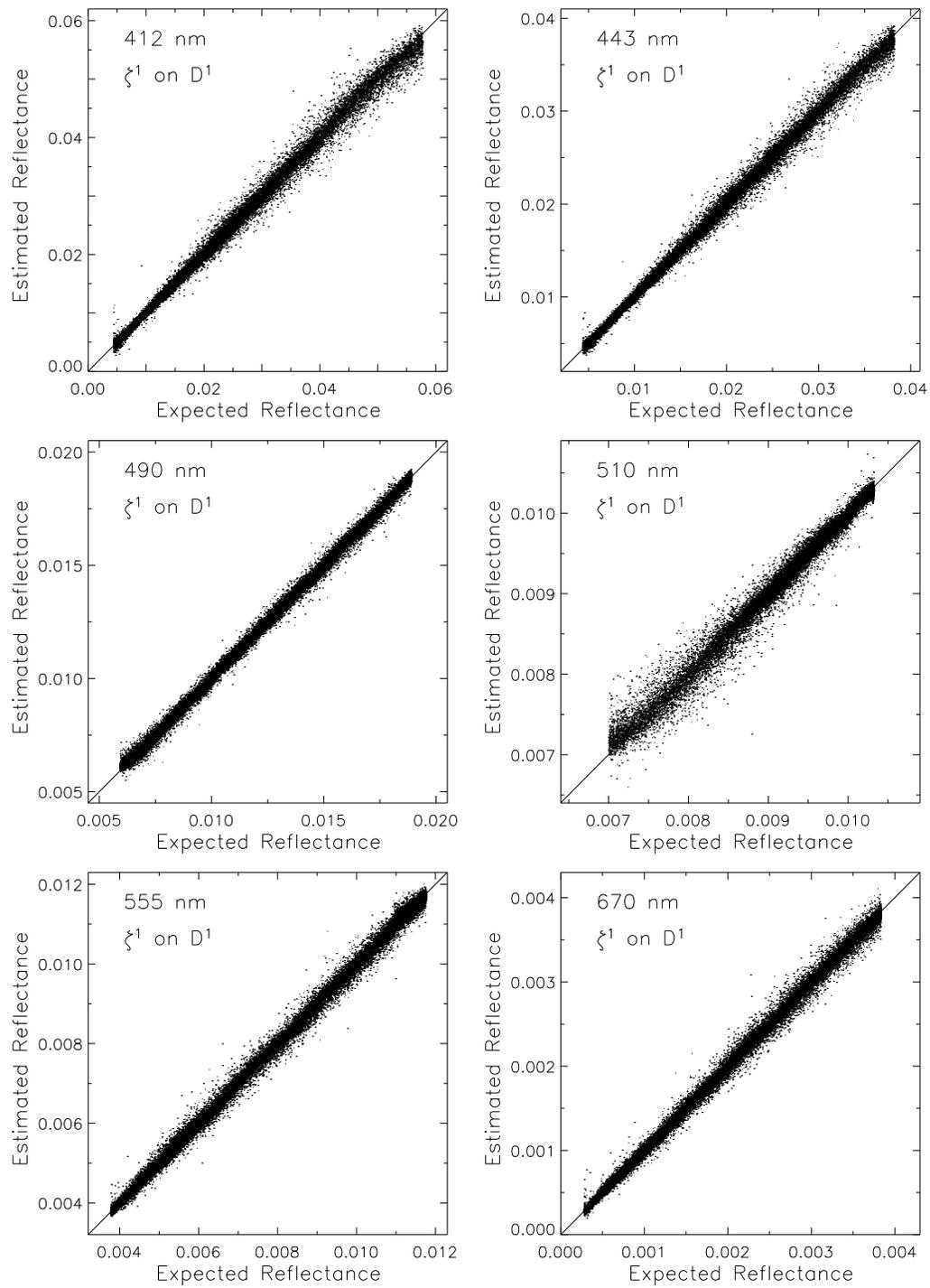


Fig. 2. Expected versus estimated marine reflectance for model ζ^1 on 1%-noisy data.

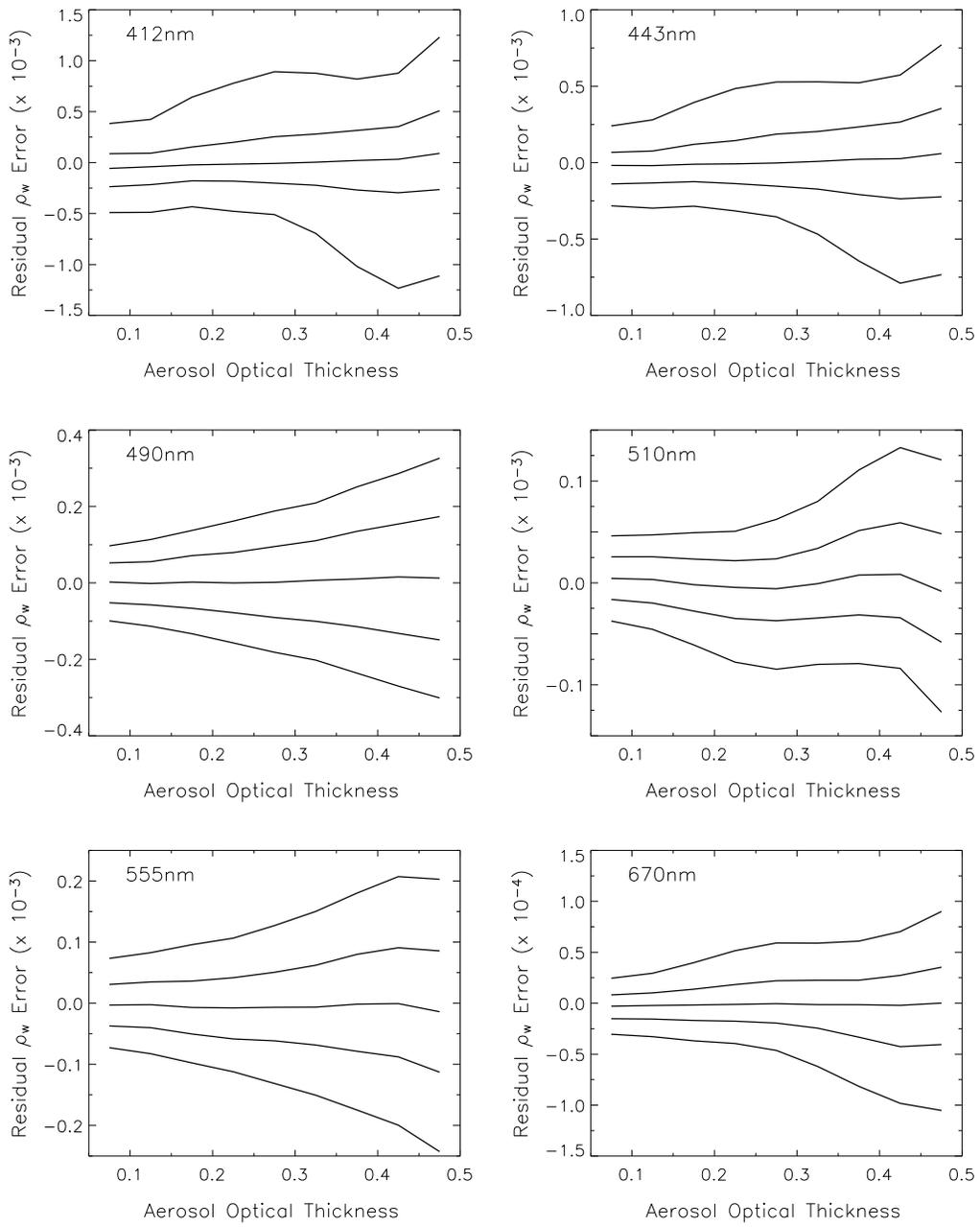


Fig. 3. Conditional quantiles (of order 0.1, 0.25, 0.5, 0.75, and 0.9) of the residual distributions as a function of the aerosol optical thickness for model ζ^1 applied on 1%-noisy data.

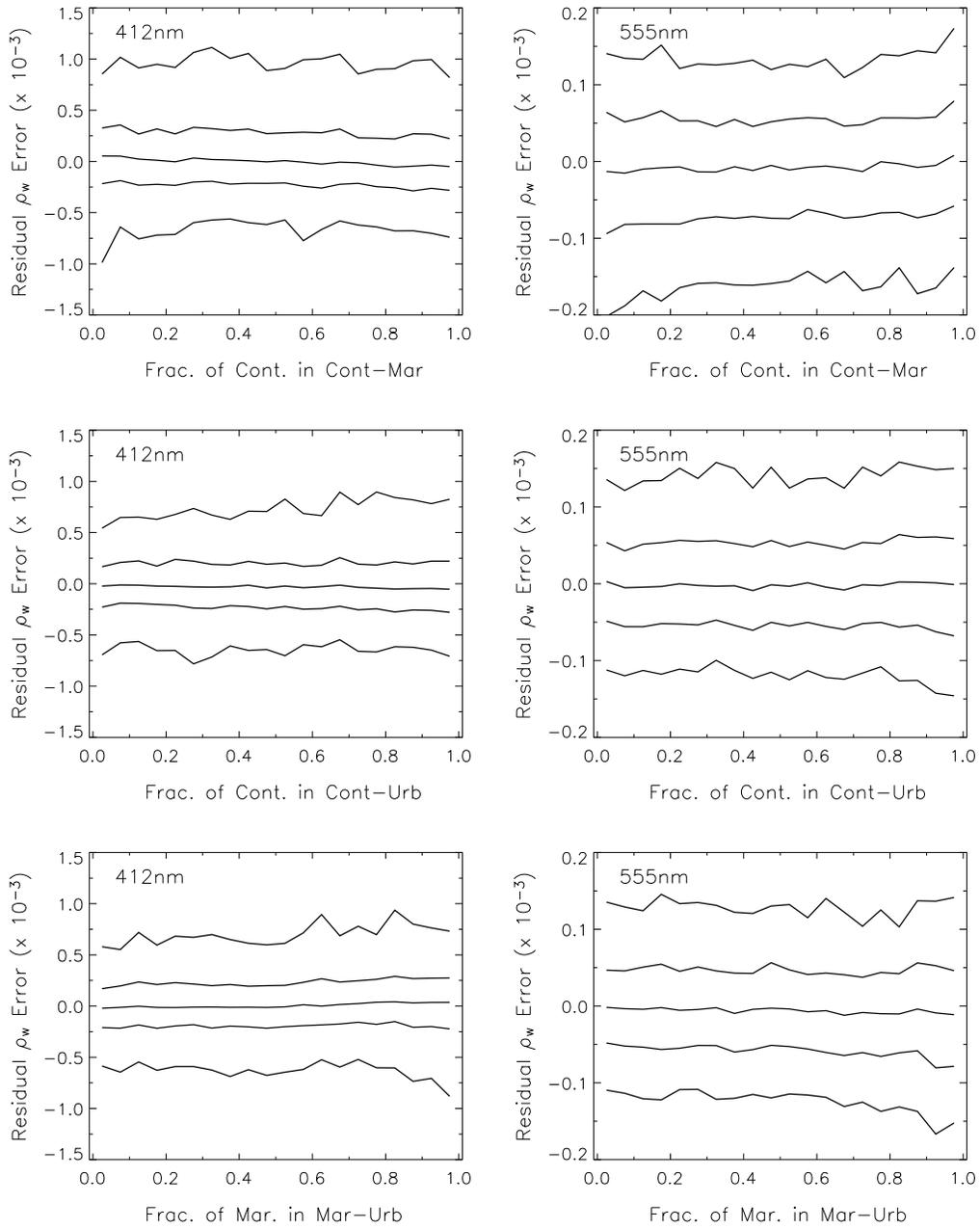


Fig. 4. Conditional quantiles (of order 0.1, 0.25, 0.5, 0.75, and 0.9) of the residual distributions at 412 and 555 nm as a function of the proportion of one aerosol model in a mixture of two for model ζ^1 applied on 1%-noisy data.

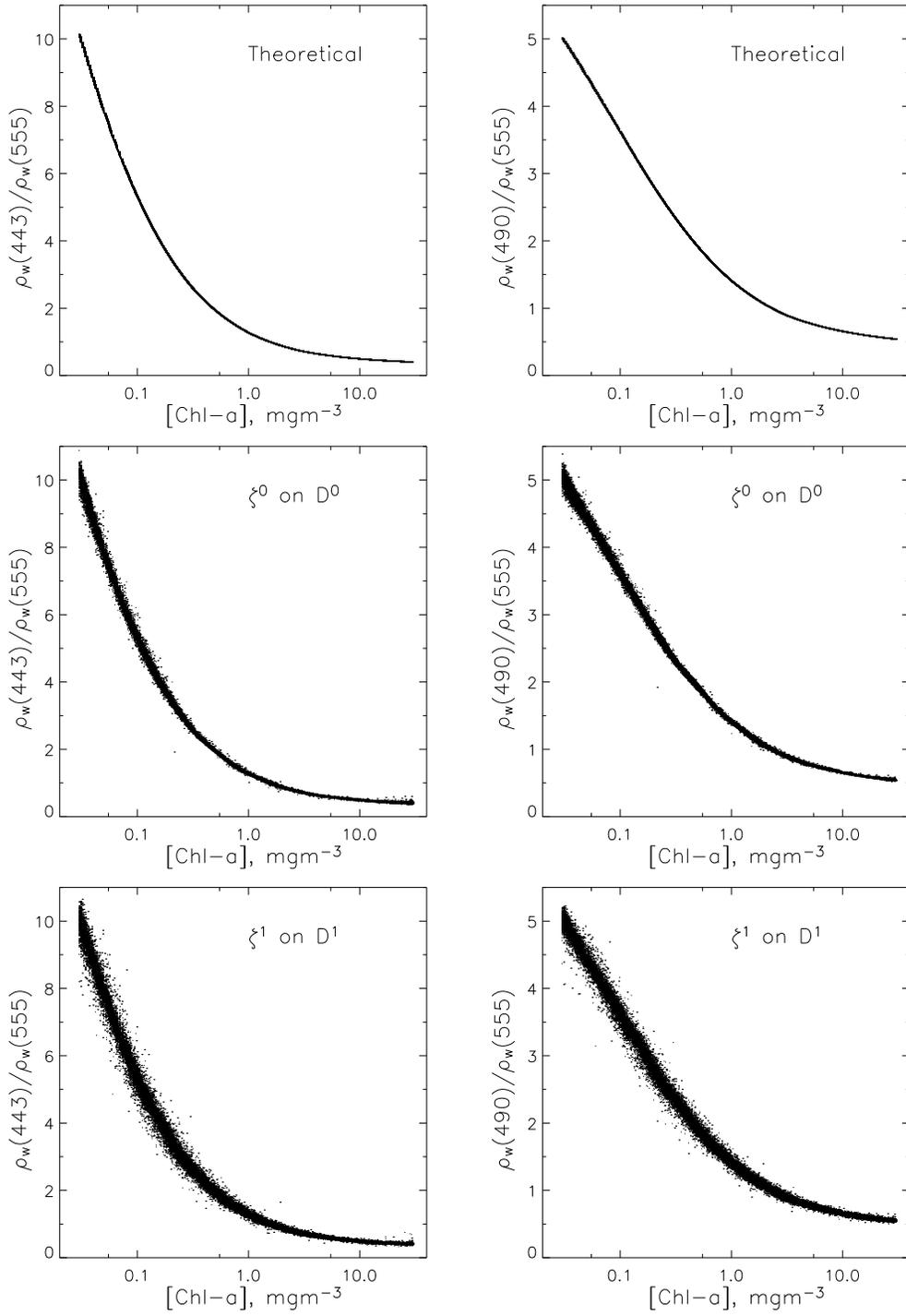


Fig. 5. Ratios $\rho_w(443)/\rho_w(555)$ (top row) and $\rho_w(490)/\rho_w(555)$ (bottom row) as a function of chlorophyll-a concentration for theoretical reflectance (left column), for reflectance estimated by ζ^0 from non-noisy data (middle column), and for reflectance estimated by ζ^1 from 1%-noisy data.

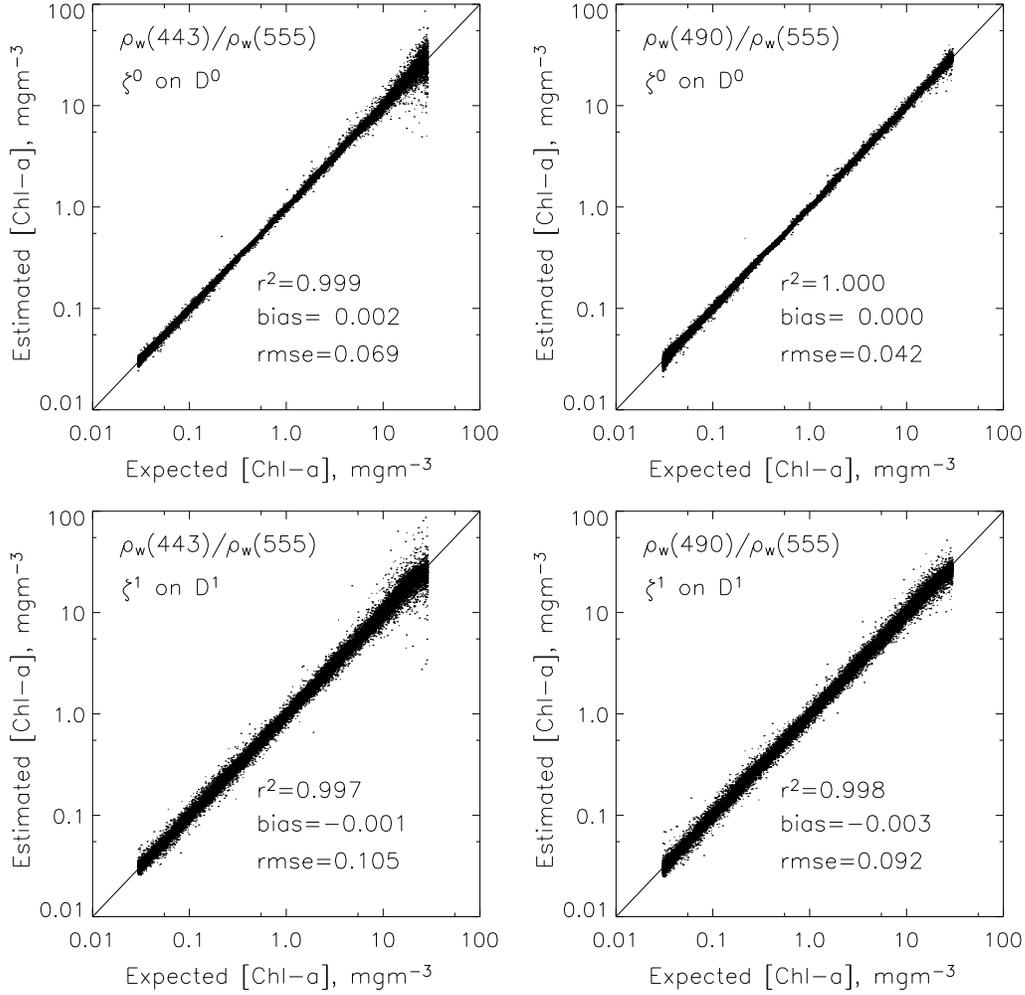


Fig. 6. Estimated versus expected chlorophyll-a concentration for ratios $\rho_w(443)/\rho_w(555)$ (left column) and $\rho_w(490)/\rho_w(555)$ (right column), when marine reflectance is estimated by ζ^0 from non-noisy data (top row) and by ζ^1 from 1%-noisy data (bottom row).

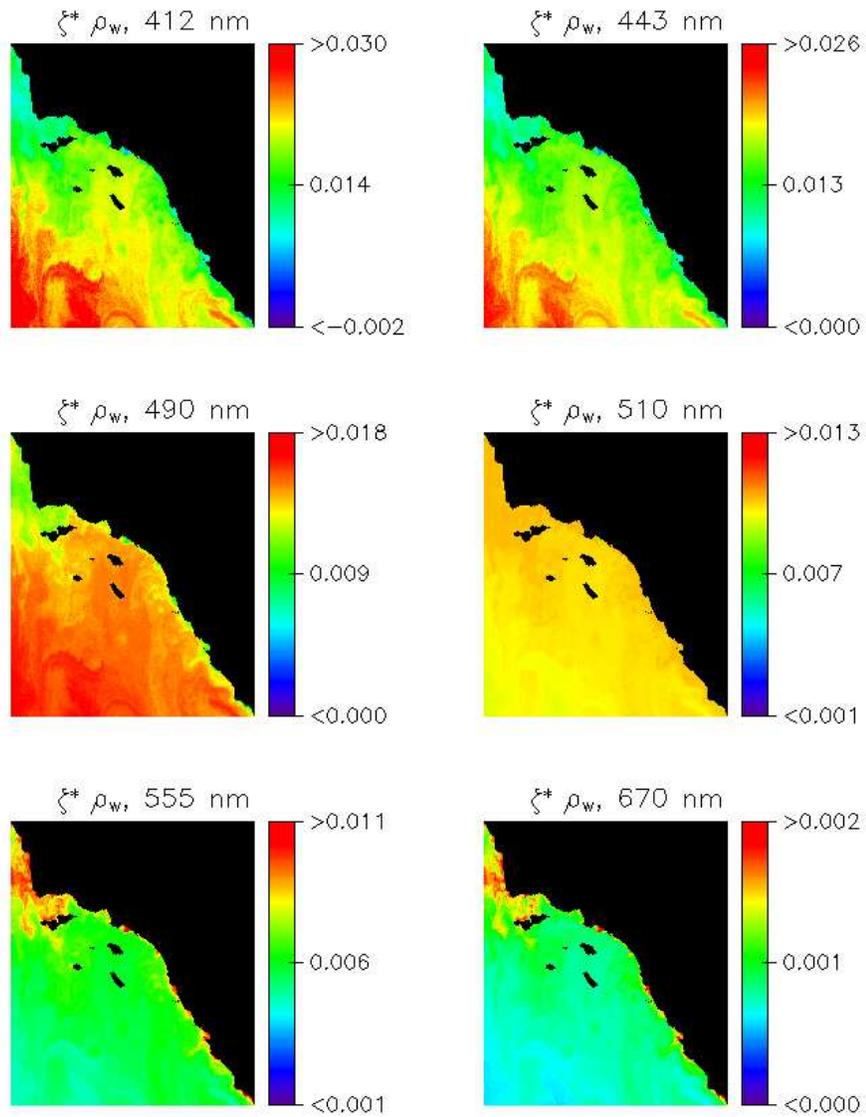


Fig. 7. Marine reflectance ρ_w estimated by ζ^* for SeaWiFS imagery acquired on November 19, 2002 over Southern California.

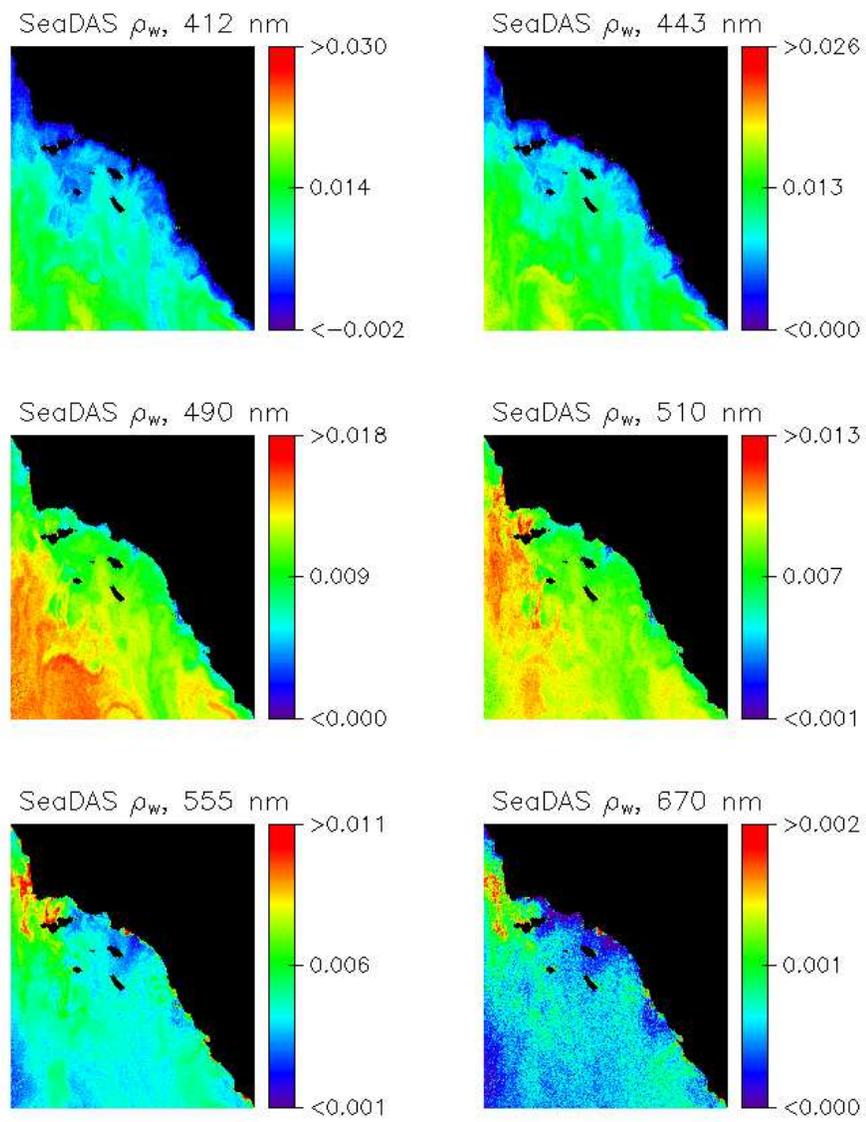


Fig. 8. Same as Fig. 7, but marine reflectance ρ_w estimated by SeaDAS.

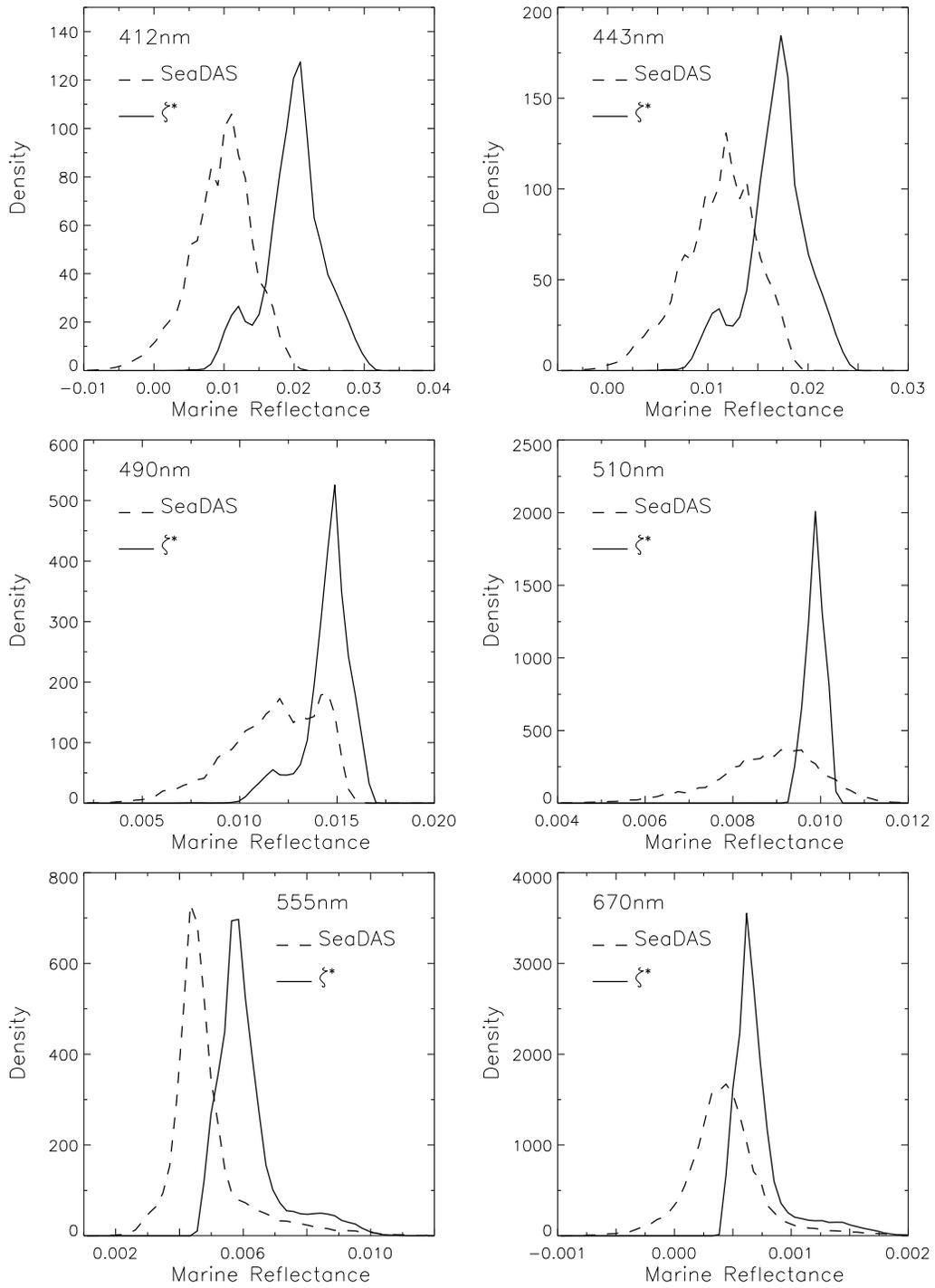


Fig. 9. Density estimates of marine reflectance retrieved by SeaDAS (dashed lines) and by ζ^* (solid lines).

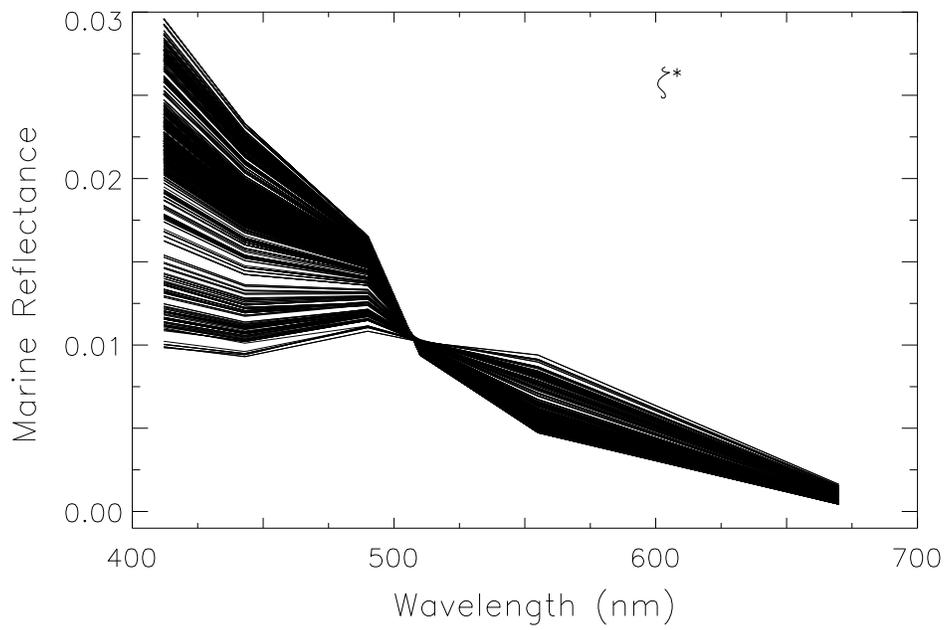
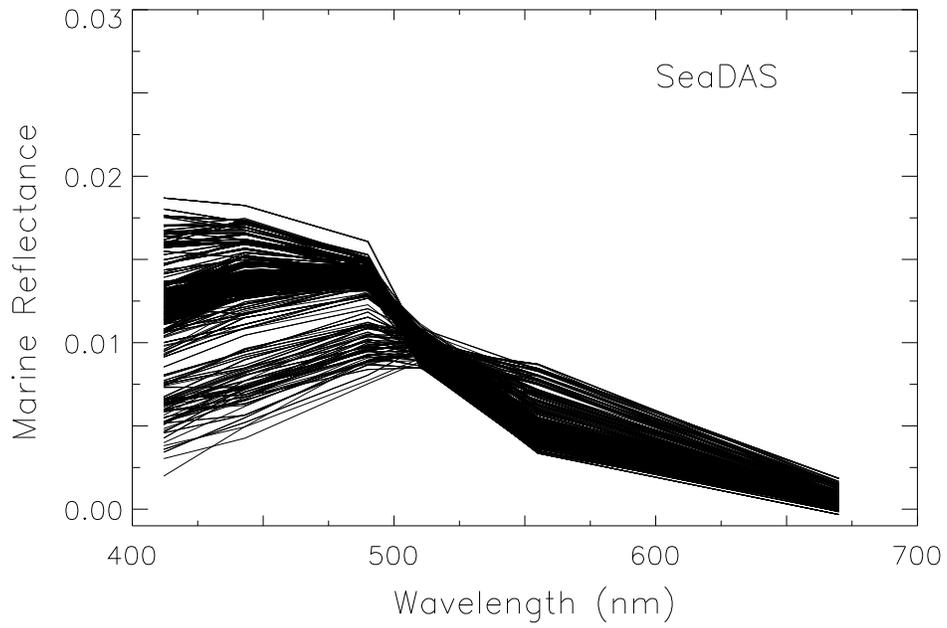


Fig. 10. 1000 randomly selected marine reflectance spectra retrieved by SeaDAS (top) and ζ^* (bottom).

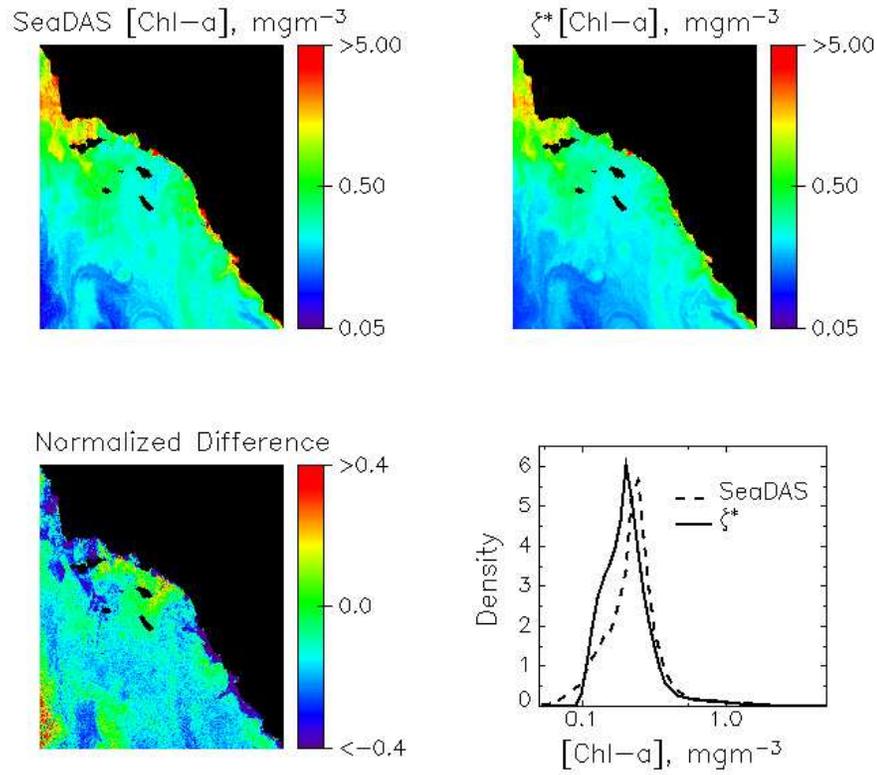


Fig. 11. Chlorophyll-a concentration [Chl-a] from SeaDAS (top left) and ζ^* (top right), normalized difference between ζ^* and SeaDAS [Chl-a] (lower left), and density estimates of SeaDAS [Chl-a] (dashed line) and ζ^* [Chl-a] (lower right).