Remote sensing of phytoplankton chlorophyll-*a* concentration by use of ridge function fields

Bruno Pelletier and Robert Frouin

A methodology is presented for retrieving phytoplankton chlorophyll-*a* concentration from space. The data to be inverted, namely, vectors of top-of-atmosphere reflectance in the solar spectrum, are treated as explanatory variables conditioned by angular geometry. This approach leads to a continuum of inverse problems, i.e., a collection of similar inverse problems continuously indexed by the angular variables. The resolution of the continuum of inverse problems is studied from the least-squares viewpoint and yields a solution expressed as a function field over the set of permitted values for the angular variables, i.e., a map defined on that set and valued in a subspace of a function space. The function fields of interest, for reasons of approximation theory, are those valued in nested sequences of subspaces, such as ridge function approximation spaces, the union of which is dense. Ridge function fields constructed on synthetic yet realistic data for case I waters handle well situations of both weakly and strongly absorbing aerosols, and they are robust to noise, showing improvement in accuracy compared with classic inversion techniques. The methodology is applied to actual imagery from the Sea-Viewing Wide Field-of-View Sensor (SeaWiFS); noise in the data are taken into account. The chlorophyll-*a* concentration obtained with the function field methodology differs from that obtained by use of the standard SeaWiFS algorithm by 15.7% on average. The results empirically validate the underlying hypothesis that the inversion is solved in a least-squares sense. They also show that large levels of noise can be managed if the noise distribution is known or estimated.   © 2006 Optical Society of America

OCIS codes: 280.0280, 010.0010.

1. Introduction

Standard algorithms for retrieval of phytoplankton chlorophyll-*a* concentration from space[1–7] attempt to correct as accurately as possible the influence of the atmosphere and the surface on the top-of-atmosphere reflectance in the visible. The problem is difficult because only 10% of the signal measured at the ocean-color wavelengths (i.e., the useful signal) may originate from the water body.[1,2] The procedure consists of (1) estimating the aerosol reflectance in the red and near infrared where the ocean can be considered black (i.e., totally absorbing) and (2) extrapolating the estimated aerosol reflectance to shorter wavelengths. The retrieved water reflectance is then related to chlorophyll-*a* concentration by use of a bio-optical model that is semianalytical[8,9] or empirical.[10] This approach has been successful, and it is employed in the operational processing of data from most satellite ocean-color sensors. In coastal regions (~2% of the oceans), however, the assumption of null ocean reflectance in the red and near infrared does not hold,[11] and improvements to the standard algorithms have been proposed. These improvements are based on assumptions of spatial homogeneity for the spectral ratio of the aerosol and water reflectance in the red and near infrared[12] or for the aerosol type.[13]

Other algorithms[14–20] attempt to determine simultaneously aerosol properties and chlorophyll-*a* concentration. Through systematic variation of candidate aerosol models, phytoplankton scattering, chlorophyll-*a* concentration, and aerosol optical thickness, a best fit to the spectral top-of-atmosphere reflectance (visible and near infrared) is obtained in an iterative manner. The advantage of this approach, compared with the standard, two-

step approach, resides in its capability to handle both weakly and strongly absorbing aerosols.[19,20] A drawback is that convergence may not be achieved immediately in some cases, making it impractical to process large amounts of satellite data. The approach is also limited by difficulty in differentiating between aerosol absorption and water constituents such as yellow substances, and the result may be noisy.

The two types of algorithm described above are fairly complicated. They require large lookup tables of aerosol optical properties or aerosol radiance. These tables are used internally as the atmospheric correction (standard procedure or nonlinear optimization) is effected. It may be possible to accomplish and speed up the search for the proper geophysical variables or to perform a mapping between top-of-atmosphere reflectance and aerosol properties by use of nonlinear modeling tools (e.g., artificial neural networks).[21,22] Alternatively, nonlinear regression techniques may be used to perform a direct mapping between top-of-atmosphere reflectance and geophysical variables. This approach has been used with some success in retrieving concentrations of phytoplankton chlorophyll-a, dissolved organic matter, and suspended sediments[23–25] but from spectral marine reflectance. Working with top-of atmosphere reflectance instead of marine reflectance (the information obtained after atmospheric correction) involves a different kind of complexity linked to the strong variation of the former with angular geometry. Owing to the embedded spectral and angular variability, the uncertainty attached to the retrieved value may be significantly larger. Diffuse marine reflectance is much more isotropic, at least over the range of angles used in remote sensing.[26]

The problem of retrieving the chlorophyll-a concentration from top-of-atmosphere reflectance in a regression setting may be introduced as follows: Let y be the chlorophyll-a concentration, let $\mathbf{t}$ be the vector of angular variables that characterize the observation process (Sun zenith angle, view zenith angle, and relative azimuth angle), let $\mathbf{z}$ be a vector of other parameters influencing the radiative transfer in the ocean–atmosphere system (e.g., atmospheric parameters, surface parameters), and let $\mathbf{x}$ be a vector of reflectance at d wavelengths related to $\mathbf{t}$, y, and $\mathbf{z}$ by

$$\mathbf{x} = F(\mathbf{t}, y, \mathbf{z}), \qquad (1)$$

where F is the radiative transfer map. Consider a statistical framework in which the reflectance is measured with error such that, instead of $\mathbf{x}$, the available datum is the vector

$$\tilde{\mathbf{x}} = \mathbf{x} + \tilde{\epsilon}, \qquad (2)$$

where $\tilde{\epsilon}$ is a random vector of null mean and finite variance representing an additive noise. Note that the properties of $\tilde{\epsilon}$ are left completely unspecified in this statement; in particular, $\tilde{\epsilon}$ may be correlated with $\mathbf{x}$. Given a pair $(\tilde{\mathbf{x}}, \mathbf{t})$, for which $\mathbf{t}$ is assumed to

be measured accurately, the problem at hand is to provide a prediction $\hat{y}(\tilde{\mathbf{x}}, \mathbf{t})$ of y.

One classic approach is to consider the prediction of y with mimimum mean-squared error, which is the conditional mean $\mathbb{E}[y|\tilde{\mathbf{x}}, \mathbf{t}]$ of y given $\tilde{\mathbf{x}}$ and $\mathbf{t}$, i.e., when y is considered a function, the regression function of y on $\tilde{\mathbf{x}}$ and $\mathbf{t}$. This proves useful under the assumption that the conditional distribution of y given $\tilde{\mathbf{x}}$ and $\mathbf{t}$ is sufficiently concentrated about its mean value, which may be checked empirically after a model has been fitted. Note that $\mathbb{E}[y|\tilde{\mathbf{x}}, \mathbf{t}]$ depends on the probability distribution of the noise $\tilde{\epsilon}$. So in this setting one obtains the prediction of y by evaluating the regression function at $(\tilde{\mathbf{x}}, \mathbf{t})$. The point is that map F in Eq. (1), which involves integrodifferential operators, is far too complex for an analytical expression of the regression function to be derived. One alternative approach is to generate a statistically significant data set that embeds the major physics of the problem, using Eq. (1), to assume a plausible noise distribution, and next to fit a model to this data. This settles the retrieval problem in the context of regression estimation, leading to a large class of models expressed as follows:

$$y = r(\tilde{\mathbf{x}}, \mathbf{t}) + \sigma(\tilde{\mathbf{x}}, \mathbf{t})\epsilon, \qquad (3)$$

where r and σ are the regression and variance functions, respectively, and where ϵ is independent of $\tilde{\mathbf{x}}$ and $\mathbf{t}$, with $\mathbb{E}[\epsilon] = 0$ and $\mathrm{Var}(\epsilon) = 1$.

This point of view is general and presents some drawbacks, however. Clearly $\mathbf{t}$ is not informative with respect to y, and only $\mathbf{x}$ contains explanatory variables. It turns out that the angular variables in $\mathbf{t}$ act as conditioning variables in the sense that they influence the relationships between the chlorophyll-a concentration and the reflectance. For instance, if $\mathbf{x_t}$ denotes a reflectance acquired in observation geometry $\mathbf{t}$, then y can be predicted independently from measurements of $\mathbf{x}_{\mathbf{t}_1}$ as well as of $\mathbf{x}_{\mathbf{t}_2}$, for some observation geometries $\mathbf{t}_1 \neq \mathbf{t}_2$. Naturally, there exist observation geometries that are less favorable than others, e.g., geometries that yield glitter and geometries that correspond to large Sun and view zenith angles, where the atmosphere contribution becomes predominant in the measured signal. Another fact is that $\mathbf{x}$ and $\mathbf{t}$ are correlated *per se*, inasmuch as the former depends on the latter by means of radiative transfer map F. Correlations between regressors is a common issue in multivariate regression problems that may worsen performance, especially in the case of nonlinear cross dependencies that are difficult to handle.[27]

In this paper, we introduce a novel statement of the retrieval problem that emphasizes the separation of the angular variables from the reflectance. The problem is formulated as a collection indexed by $\mathbf{t}$ of similar inverse problems that we propose to solve in a regression context and under the constraint of continuity of the solutions with respect to $\mathbf{t}$. The underlying idea is to attach to each $\mathbf{t}$, and continuously in $\mathbf{t}$,

a regression model that provides one with a prediction of y. This leads one to express the solution as a function field over the set T of permitted values for the angular variables, i.e., as a map defined on T and valued in some class $\mathcal{M}$ of functions of $\mathbf{x}$, which represents the model space for the regression functions of y on $\mathbf{x_t}$. As no particular shape of regression functions is expected, we focus in this paper mainly on function fields valued in the class spanned by linear combinations of shifted ridge functions, as they generate a nested sequence of function sets, the union of which is dense,[28,29] and possess good approximation properties.[30–33] The density of the set spanned by functions of the ridge form is an important property, as, roughly speaking, it expresses the fact that this set entirely fills the ambient space of continuous functions of $\mathbf{x}$. So, for each continuous function f of $\mathbf{x}$, one may find a linear combination of ridge functions that is arbitrarily close to f. Furthermore, a similar density property holds, in suitable spaces, for function fields valued in that set spanned by ridge functions.[34] However, if ridge functions are emphasized in this paper, the proposed methodology remains general and may easily be adapted to other choices of the class $\mathcal{M}$.

The paper is organized as follows: In Section 2 the basics of approximation by ridge functions and of function fields are first presented. The utilization of ridge function fields in a regression setting is then described. In Section 3, procedures to retrieve chlorophyll-a concentration from space by use of ridge function fields are detailed. The performance of the modeling is examined with synthetic data, including a detailed analysis of how performance varies as a function of the geophysical variables and noise distribution. In Section 4 the methodology is applied to Sea-Viewing Wide Field-of-View Sensor (SeaWiFS) data, and the results are compared with those from the standard SeaWiFS processing algorithm. Finally, in Section 5 a summary of the findings is given, as well as a perspective for future work.

2. Ridge Function Fields

A. Definitions

A ridge function on some set $X \subset \mathbb{R}^d$ is a function of the form $h(\mathbf{a} \cdot \mathbf{x})$, where h is a function on $\mathbb{R}$, where $\mathbf{a}, \mathbf{x} \in \mathbb{R}^d$, and where $\mathbf{a} \cdot \mathbf{x}$ is the standard inner product of $\mathbb{R}^d$. A ridge function approximation refers to an approximation by linear combinations of n ridge functions for some integer n, i.e., by functions of the form

$$f(\mathbf{x}) = \sum_{i=1}^{n} c_i h(\mathbf{a}_i \cdot \mathbf{x}). \tag{4}$$

A slight variant is an approximation by linear combinations of shifted ridge functions of the form

$$f(\mathbf{x}) = \sum_{i=1}^{n} c_i h(\mathbf{a}_i \cdot \mathbf{x} + b_i), \tag{5}$$

where the scalars b_i are the shifts. Function h is called the generator function, vectors $\mathbf{a}_i$ are called the directions, and scalars c_i are the coefficients of expansion. $\mathbf{a}_i$, b_i, and c_i constitute the free parameters of f, summarized by parameter vector θ_n, taking values in the set $\Theta_n = \Pi_{i=1}^{n} \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}$. Note that one may also allow the generator function to vary in some subset of $\mathcal{C}(\mathbb{R})$, but this setting is not considered herein. Function f of the form given by Eq. (2), which corresponds to a given θ_n, will also be denoted $f(.; \theta_n)$. When θ_n varies in all Θ_n, the functions $f(.; \theta_n)$ span a set of continuous functions on X that are denoted $\mathcal{M}_n$. Sets $\mathcal{M}_n$ and Θ_n are generally not homeomorphic, for their elements are generally not in one-to-one correspondance, but there is the continuous surjection $i_n: \Theta_n \to \mathcal{M}_n$, carrying each θ_n to the function $f(.; \theta_n)$. Also useful is the set $\mathcal{M} = \cup_n \mathcal{M}_n$. It is known that $\mathcal{M}$ is dense in $\mathcal{C}(X)$, i.e., its closure $\bar{\mathcal{M}} = \mathcal{C}(X)$ (see Refs. 28 and 29). Additional materials on ridge function approximations may be found in the survey paper by Light[35] and in Refs. 30–33.

Now let X be an open subset of $\mathbb{R}^d$ and let $\mathcal{C}(X)$ be the space of continuous real-valued functions on X with the topology of uniform convergence on compact sets. Let T be a compact subset of $\mathbb{R}^p$. Recall that a set S is said to be compact if every open covering of S has a finite subcovering and that the compact subsets of $\mathbb{R}^p$ are its closed and bounded subsets. A function field over T is an application $\zeta: T \to \mathcal{C}(X)$. If ζ is continuous, then the field is said to be continuous, and the set of all continuous function fields over T is denoted $\mathcal{C}(X)^T$. For all $\mathbf{t} \in T$, $\zeta(\mathbf{t})$ is the function carrying each $\mathbf{x} \in X$ to the scalar $\zeta(\mathbf{t})(\mathbf{x})$. For X and T as above, there is the homeomorphism $\mathcal{C}(X \times T) \stackrel{\sim}{\Rightarrow} \mathcal{C}(X)^T$. By this homeomorphism, there corresponds continuously to each $\zeta \in \mathcal{C}(X)^T$ a unique map $\zeta^* \in \mathcal{C}(X \times T)$ and conversely. Maps ζ and ζ^* are such that $\zeta^*(\mathbf{x}, \mathbf{t}) = \zeta(\mathbf{t})(\mathbf{x})$ for all $\mathbf{x} \in X$ and $\mathbf{t} \in T$.

One may define a ridge function field over T similarly by replacing $\mathcal{C}(X)$ with $\mathcal{M}$. To compute the value of a ridge function field $\zeta \in \mathcal{M}^T$ at some point $\mathbf{x}$ and $\mathbf{t}$, an explicit representation of ζ is needed. Consider the representation of elements of $\mathcal{M}_n^T$ for some fixed integer n. As the functions in $\mathcal{M}_n$ are parameterized by vectors in Θ_n, it is natural to seek a representation of ζ as a continuous map $\xi: T \to \Theta_n$, i.e., such that $\zeta = i_n \circ \xi$. Unfortunately the existence of such a continuous $\xi: T \to \Theta_n$ for each (continuous) $\zeta \in \mathcal{M}_n^T$ is not ensured because Θ_n is not homeomorphic to $\mathcal{M}_n$. However, it is shown in Ref. 34 that the set of all continuous ridge function fields of the form $i_n \circ \xi$, with ξ continuous, is dense in $\mathcal{C}(X)^T$. For each element ζ in this set, its associated map $\zeta^* \in \mathcal{C}(X \times T)$ is expressed more conveniently as

$$\zeta^*(\mathbf{x}, \mathbf{t}) = \sum_{i=1}^{n} c_i(\mathbf{t}) h[\mathbf{a}_i(\mathbf{t}) \cdot \mathbf{x} + b_i(\mathbf{t})], \tag{6}$$

where b_i and c_i are continuous real-valued functions of $\mathbf{t}$ and where $\mathbf{a}_i$ are continuous vector-valued functions of $\mathbf{t}$.

The process by which an arbitrary function field $\bar{\zeta} \in \mathscr{C}(X)^T$ is approximated by a ridge function field that satisfies Eq. (6) may be described as follows: Let $e_{i,\mathbf{t}}(\mathbf{x}) = h[\mathbf{a}_i(\mathbf{t}) \cdot \mathbf{x} + b_i(\mathbf{t})]$. These are basis functions that span at most an n-dimensional vector space $E_\mathbf{t} \subset \mathscr{C}(X)$. Vector spaces $E_\mathbf{t}$ constitute a collection of approximation spaces for the collection of functions $\bar{\zeta}(\mathbf{t}) \in \mathscr{C}(X)$; i.e., each $\bar{\zeta}(\mathbf{t})$ is approximated by an element of $E_\mathbf{t}$. By varying the directions and the shifts, one seeks a collection of $E_\mathbf{t}$ such that, for all $\mathbf{t}$, $E_\mathbf{t}$ contains a good (in some sense) approximation to $\bar{\zeta}(\mathbf{t})$, all of this being performed under the constraint of continuity with respect to $\mathbf{t}$.

Note that directions $\mathbf{a}_i$ and shifts b_i in Eq. (6) may be constants that do not depend on $\mathbf{t}$. In this important special case, the collection of approximation spaces reduces to a single approximation space, common to all $\mathbf{t}$, that is spanned by functions of the form $h(\mathbf{a}_i \cdot \mathbf{x} + b_i)$. This leads to ridge function fields ζ such that

$$\zeta^*(\mathbf{x}, \mathbf{t}) = \sum_{i=1}^n c_i(\mathbf{t})h(\mathbf{a}_i \cdot \mathbf{x} + b_i), \qquad (7)$$

and the set of all ridge function fields satisfying this equation is also dense in $\mathscr{C}(X)^T$.

B. Regression

Let $\mathscr{D} = \{(\tilde{\mathbf{x}}_i, \mathbf{t}_i, y_i; i = 1, \ldots, N\}$ be a sample of N independent and identically distributed observations drawn according to Eqs. (1) and (2), i.e., such that, for all i, there have been outcomes $\mathbf{z}_i$ and $\tilde{\epsilon}_i$ of $\mathbf{z}$ and $\tilde{\epsilon}$ with

$$\tilde{\mathbf{x}}_i = F(\mathbf{t}_i, y_i, \mathbf{z}_i) + \tilde{\epsilon}_i. \qquad (8)$$

Denote by $r(\mathbf{t})$: $\tilde{\mathbf{x}} \to r(\mathbf{t})(\tilde{\mathbf{x}}) \in \mathbb{R}$ the regression function of y on $\tilde{\mathbf{x}}$ acquired in geometry $\mathbf{t}$. With this notation, map $\mathbf{t} \to r(\mathbf{t})$ is a function field over T. We consider the estimation of this field of regression functions by ridge function field ζ introduced above, based on the sample $\mathscr{D}$. The statistical model is expressed as

$$y_i = \zeta(\mathbf{t}_i)(\tilde{\mathbf{x}}_i) + \epsilon_i, \qquad (9)$$

where ϵ_i are independent random variables of null mean and finite variance. Once this model is adjusted to $\mathscr{D}$, the prediction $\hat{y}(\tilde{\mathbf{x}}^*, \mathbf{t}^*)$, for a new measured pair $(\tilde{\mathbf{x}}^*, \mathbf{t}^*)$ is taken to be $\hat{y}(\tilde{\mathbf{x}}^*, \mathbf{t}^*) = \zeta(\mathbf{t}^*)(\tilde{\mathbf{x}}^*)$. Practically, we take $\zeta \in (\mathcal{M}_n)^T$ for some integer n to be selected from the data, such that the statistical model may be rewritten as

$$y_i = \zeta^*(\tilde{\mathbf{x}}_i, \mathbf{t}_i) + \epsilon_i \qquad (10)$$

$$= \sum_{k=1}^n c_k(\mathbf{t}_i)h[\mathbf{a}_k(\mathbf{t}_i) \cdot \tilde{\mathbf{x}}_i + b_k(\mathbf{t}_i)] + \epsilon_i. \qquad (11)$$

Note that, contrary to the general regression model in Eq. (3), the above model involves a special kind of interaction between the noisy reflectance ($\tilde{\mathbf{x}}$) and the angular variables ($\mathbf{t}$).

The free parameters in the above model are the vector-valued functions $\mathbf{a}_k(\mathbf{t})$ and the real-valued functions $b_k(\mathbf{t})$ and $c_k(\mathbf{t})$, plus eventually integer n. Statistical models indexed by infinite-dimensional parameters are called semiparametric models. Without constraints on the functional parameters, these models are actually too general, and an unrestricted nonparametric estimation of their parameters may not be possible, so that one generally imposes restrictions such as that they belong to some given parametric model set or that they are nonparametric but smooth. The latter form of restriction is commonly used and fairly efficient, and we describe it now. It consists in expanding the functional parameters on some set $\{\varphi_1(\mathbf{t}), \ldots, \varphi_K(\mathbf{t})\}$ of basis functions of $\mathbf{t}$ while penalizing the roughness of the curve during the fit of the model. The functional parameters take the form $\sum_{k=1}^K \alpha_k \varphi_k(\mathbf{t})$, where α_k are the coefficients of the expansion, and their roughness is controlled by some functional, typically the integrated squared second derivative. Using this representation of the functional parameters, one may fit a ridge function field by penalized least squares. More specifically, let SSE$(\zeta) = \sum_{i=1}^N [y_i - \zeta(\mathbf{t}_i)(\tilde{\mathbf{x}}_i)]^2$ be the sum of the squared errors for ζ on $\mathscr{D}$. The penalized least-squares criterion ε to be minimized is

$$\varepsilon = \text{SSE}(\zeta) + \sum_k \lambda_k^a \int_T \|\nabla^2 \mathbf{a}_k\|^2(\mathbf{t}) \, \mathrm{d}\mathbf{t}$$

$$+ \sum_k \lambda_k^b \int_T \|\nabla^2 b_k\|^2(\mathbf{t}) \, \mathrm{d}\mathbf{t}$$

$$+ \sum_k \lambda_k^c \int_T \|\nabla^2 c_k\|^2(t) \, \mathrm{d}\mathbf{t}, \qquad (12)$$

where the nonnegative coefficients λ_k^a, λ_k^b, and λ_k^c, called the smoothness parameters, control the trade-off between goodness of fit and parsimony of the model, where functional parameters $\mathbf{a}_k$, b_k, and c_k belong to span $\{\varphi_1(\mathbf{t}), \ldots, \varphi_K(\mathbf{t})\}$, and where ∇^2 denotes the iterated differentiation operator with respect to $\mathbf{t}$. Note that other types of penalty may be used. For given smoothness parameters, this minimization problem may be solved by means of standard minimization techniques. Then the smoothness parameters have to be selected from the data by use of model selection criteria, such as cross validation and generalized cross validation.

It remains to define basis functions $\varphi_1, \ldots, \varphi_K$. Using spline functions associated with a set of knots (i.e., points of T) is generally good for this purpose. Provided that the number of knots is large enough and that the knots cover the data well, the linear span of a spline basis leads to very flexible models, from simple shapes with few degrees of freedom (high

value of the smoothing parameter) to more-complex wiggly shapes with numerous degrees of freedom (low values of the smoothness parameter). As $\mathbf{t}$ is of dimension greater than 1, multidimensional splines must be used, such as thin plate splines and tensor-product splines. This completely defines a general methodology for fitting a function field valued in some parameterized set of functions, in particular, ridge function fields. Note that, even if the number of knots is large, the resultant model is not necessarily of high complexity, as the overall number of degrees of freedom is controlled by the smoothing parameters.

In this work, we adopt another point of view, where few basis functions are used for the representation of the functional parameters but with no roughness penalty, which corresponds to the first kind of restriction described above. More specifically, we consider a small number of knots, positioned on a regular grid covering T, and the basis functions $\{\varphi_k; k = 1, \ldots, K\}$ associated with multilinear interpolation on this grid, such that the functional parameters belong to the parametric model set, which is the linear span of φ_k. This approach is suitable for fitting a proof-of-concept model, whereas the above methodology allows for finer control of the complexity (i.e., the number of degrees of freedom). Nevertheless, the set of ridge function fields constructed in this way is known to span a dense set.[34]

Multilinear interpolation on $\mathbb{R}^p$, $p \geq 1$, is a standard technique for interpolation of multidimensional data that generalizes linear interpolation on the real line. Given a set of K fixed points in $\mathbb{R}^p$ that define a regular grid on some hypercube $D \subset \mathbb{R}^p$, let f be a real-valued map on D. Denote by $f_1, \ldots, f_K$ the values of f on the K points. Note that K is a product of p integers greater than 2. In multilinear interpolation, the value of the reconstructed f at some point $\mathbf{t} \in D$ is a linear combination of the values of f on the 2^p closest points to $\mathbf{t}$ on the grid, where the coefficients of the linear combination depend on $\mathbf{t}$. By an appropriate scaling, the general multilinear interpolation procedure may be defined on the interior of the unit cube $[0; 1]^p$. So assume that $\mathbf{t} \in \text{int}([0; 1]^p)$, denote by $\mathbf{t}_1, \ldots, \mathbf{t}_{2^p}$ the 2^p corners of $[0; 1]^p$ and by t_k^j the jth component of $\mathbf{t}_k$ (so $k = 1, \ldots, 2^p$ and $j = 1, \ldots, p$), and denote by $f_1, \ldots, f_{2^p}$ the values of f on these corners. Then the interpolated value $\hat{f}(\mathbf{t})$ of f at $\mathbf{t} = (t^1, \ldots, t^p)^t$ is given by

$$\hat{f}(\mathbf{t}) = \sum_{k=1}^{2^p} f_k \prod_{j=1}^{p} (1 - |t^j - t_k^j|). \quad (13)$$

When $\mathbf{t}$ belongs to boundary $\partial[0; 1]^p$, $\mathbf{t}$ is in the interior of $[0; 1]^q$ for some $1 \leq q < p$. So, to complete the definition, it suffices to define linear interpolation on the interval $[0; 1]$: for $t \in [0; 1]$, we have $\hat{f}(t) = f_0(1 - t) + f_1 t$. It may be shown that the resultant interpolated function on D may be expressed as $\hat{f}(\mathbf{t}) = \sum_{k=1}^{K} f_k \varphi_k(\mathbf{t})$, and the functions φ_k are called the basis functions associated with multilinear interpolation on the set of K knots. So, given a reduced set of

K fixed knots defining a regular grid covering T, denote by $\mathscr{F}_K$ the set of all such real-valued interpolated maps, i.e., $\mathscr{F}_k = \text{span}\{\varphi_1, \ldots, \varphi_K\}$. Each element of $\mathscr{F}_K$ depends on K parameters. We consider the fitting of a ridge function field ζ such that

$$\zeta^*(\mathbf{x}, \mathbf{t}) = \sum_{k=1}^{n} c_k(\mathbf{t}) h[\mathbf{a}_k(\mathbf{t}) \cdot \tilde{\mathbf{x}} + b_k(\mathbf{t})], \quad (14)$$

where for all $k = 1, \ldots, n$, $\mathbf{a}_k \in \mathscr{F}_K^d$, $b_k \in \mathscr{F}_K$, and $c_k \in \mathscr{F}_K$, by minimizing the sum of the squared errors $\text{SSE}(\zeta)$ on $\mathscr{D}$, without penalty. For notational convenience, the free parameters of the maps $\mathbf{a}_k$, b_k, and c_k are summarized in a unique vector θ, of length $(d + 2)K$. As mentioned above, the minimization of $\text{SSE}(\zeta)$ with respect to θ may be solved by use of standard minimization methodologies. An iterative minimization algorithm performing a stochastic gradient descent is described below. Denote by $\theta^{(n)}$ the value of the parameter vector at the nth iteration. One iteration of the algorithm is composed of the following steps:

1. Draw a sample $(\mathbf{x}, \mathbf{t}, y)$ in $\mathscr{D}$,
2. Compute the prediction error $e(\zeta) = y - \zeta(\mathbf{t})(\mathbf{x})$,
3. Update the parameters of ζ according to

$$\theta^{(n+1)} = \theta^{(n)} - \eta \, \text{grad}(\zeta),$$

where η is a strictly positive scalar.

The algorithm is iterated until convergence. This type of algorithm allows one to simulate the noise, because in practice the data set is obtained from multiple runs of a radiative transfer code. In this case, given noise distribution p_ϵ, step 1 of the algorithm above comprises the following steps:

1(a). Draw a sample $(\mathbf{x}, \mathbf{t}, y)$ in $\mathscr{D}$,
1(b). Draw a noise sample $\tilde{\epsilon}$ according to $p_{\tilde{\epsilon}}$ and compute the noisy reflectance $\tilde{\mathbf{x}} = \mathbf{x} + \tilde{\epsilon}$.

3. Retrieval of Chlorophyll-*a* Concentration

A. Simulated Data Sets

The top-of-atmosphere reflectance in the eight Sea-WiFS spectral bands, centered at 410, 490, 510, 555, 670, 765, and 865 nm, has been simulated by use of the radiative transfer code of Ref. 36 based on the modeling approach, i.e., signal decomposition, of Ref. 37. This decomposition takes into account the essential physics of the problem, namely, scattering by molecules, scattering and absorption by aerosols, molecule–aerosol interactions, reflection by the surface, molecule–surface interactions, aerosol–surface interactions, and backscattering by whitecaps and the water body. Absorption by ozone, oxygen, and water vapor is neglected in the simulations, as its effect on the top-of-atmosphere reflectance can be removed easily, either by a simple correction factor (for ozone and water vapor)[38] or by independent corrections for molecule and aerosol components (for oxy-

gen).[39] This assumes that absorption by ozone and water vapor can be decoupled from scattering by molecules and aerosols. The justification is that ozone is located high in the atmosphere where molecules are rarified; i.e., solar radiation goes through the ozone layer pratically without scattering. For water vapor, absorption occurs at wavelengths where scattering by molecules is small and scattering by aerosols dominates. The scattered radiation field results essentially from primary and secondary scattering events, and, because scattering by aerosols is mostly forward, the photons follow a path that does not deviate much from the direct path, i.e., Sun-to-surface and surface-to-satellite.

The simulations have been performed for a wide range of Sun and view angles, aerosol types and amounts, and wind speeds. Sun and view zenith angles were varied randomly from 0 to 60°, the relative azimuth angle from 0 to 180°, wind speed from 0 to 15 ms^{-1}, and aerosol optical thickness at 550 nm from 0.01 to 0.5. For larger Sun and view zenith angles, the degree of accuracy of the radiative transfer model becomes unacceptable. Various aerosol mixtures have been considered, from purely continental to purely maritime, from purely maritime to purely urban, and from purely continental to purely urban (i.e., three types of mixture). The water body has been assumed to be Lambertian, and its reflectance modeled according to Ref. 40 for chlorophyll-a concentrations of 0 to 30 mg m^{-3}. This model is valid only for case I (open oceans) waters. Case II waters were considered in the simulations. Whitecaps have also been assumed to behave as Lambertian bodies, with a reflectance in the visible and a fractional coverage modeled according to Ref. 41. The spectral dependence of whitecap reflectance was parameterized following Ref. 42. The Sun glint reflectance was modeled according to Ref. 43, assuming isotropic slope distribution. Note that the inverse problem would not be more complicated, or more difficult to solve if the bidirectional characteristics of the water-body reflectance were taken into account, because a unique statistical model is attached to each angular geometry.

In practice, the radiative transfer code was run twice for each case, once with aerosols and a nonblack water body and once without aerosols and with a black water body. The difference between the results of the two runs, i.e., the top-of-atmosphere reflectance corrected for the effects of molecular scattering, molecule–surface interactions, and most of Sun glint (hereafter referred to as corrected reflectance) was used to construct the function fields. All cases with a Sun glint reflectance above 0.04 were discarded, leaving for each type of mixture ~20,000 instances of the corrected reflectance vector for the analysis.

B. Experimental Designs

The simulated data described above were used to construct ridge function fields providing predictions of the chlorophyll-a concentration from the corrected reflectance. Let us introduce some notation. The sim-

Table 1. Statistics for Fields ζ^0, ζ^1, and ζ^2 Evaluated for Data Sets $\mathcal{D}_e^0$, $\mathcal{D}_v^0$, $\mathcal{D}_e^1$, $\mathcal{D}_v^1$, $\mathcal{D}_e^2$, and $\mathcal{D}_v^2$

Data set	Statistic	ζ^0	ζ^1	ζ^2
$\mathcal{D}_e^0$	RMSE	0.450	0.554	0.748
	$b_{\ln}$	−0.000	0.001	0.004
	RMSE$_{\ln}$	0.040	0.049	0.062
$\mathcal{D}_v^0$	RMSE	0.502	0.641	0.812
	$b_{\ln}$	−0.000	0.000	0.003
	RMSE$_{\ln}$	0.043	0.053	0.065
$\mathcal{D}_e^1$	RMSE	1.642	0.899	0.903
	$b_{\ln}$	−0.000	0.001	0.004
	RMSE$_{\ln}$	0.135	0.086	0.087
$\mathcal{D}_v^1$	RMSE	1.739	1.010	0.973
	$b_{\ln}$	−0.001	0.000	0.002
	RMSE$_{\ln}$	0.139	0.090	0.090
$\mathcal{D}_e^2$	RMSE	4.255	1.628	1.284
	$b_{\ln}$	0.003	0.002	0.004
	RMSE$_{\ln}$	0.264	0.150	0.137
$\mathcal{D}_v^2$	RMSE	4.908	1.740	1.375
	$b_{\ln}$	0.002	0.000	0.002
	RMSE$_{\ln}$	0.270	0.156	0.143

ulated data set, composed of 61,990 samples, will be denoted $\mathcal{D}^0$. Note that this large number of points is necessary; keep in mind the high dimensionality of the problem in connection with the curse of dimensionality in regression estimation. Next, $\mathcal{D}^0$ has been randomly split into a data set $\mathcal{D}_e^0$ of 40,896 points used for estimation and a data set $\mathcal{D}_v^0$ of 21,094 points used for validation, i.e., approximately in proportions of 2/3 and 1/3. To assess the robustness of the fields, we generated noisy versions of data sets $\mathcal{D}_e^0$ and $\mathcal{D}_v^0$ with total amounts of noise of 1% and 2%; they will be denoted $\mathcal{D}_e^1$, $\mathcal{D}_v^1$, $\mathcal{D}_e^2$, and $\mathcal{D}_v^2$. The selected noise scheme is decomposed into a sum of correlated and uncorrelated components according to

$$\tilde{\mathbf{x}} = \mathbf{x} + \nu^c \mathbf{x} + \left(\nu_1^{uc} x^1, \ \ldots, \ \nu_8^{uc} x^8 \right)^t, \qquad (15)$$

In Eq. (15), $\tilde{\mathbf{x}}$ is a noisy version of reflectance vector $\mathbf{x}$, and $\nu^c, \nu_1^{nc}, \ldots, \nu_8^{nc}$ are random variables uniformly distributed on the interval $[-\nu/200; \nu/200]$, where ν is the total amount of noise, in percent. To summarize the notation used for data sets and models, a superscript stands for a total amount of noise, in percent, whereas a subscript stands for the type of data set.

The ridge function fields that we consider satisfy Eq. (14) and are defined over the set T of allowable values for the cosines of the angular variables; i.e., we have

$$\mathbf{t} = (\cos \theta_s, \ \cos \theta_v, \ \cos \Delta\varphi)^t, \qquad (16)$$

$$T = [1/2; 1] \times [1/2; 1] \times [-1; 1]. \qquad (17)$$
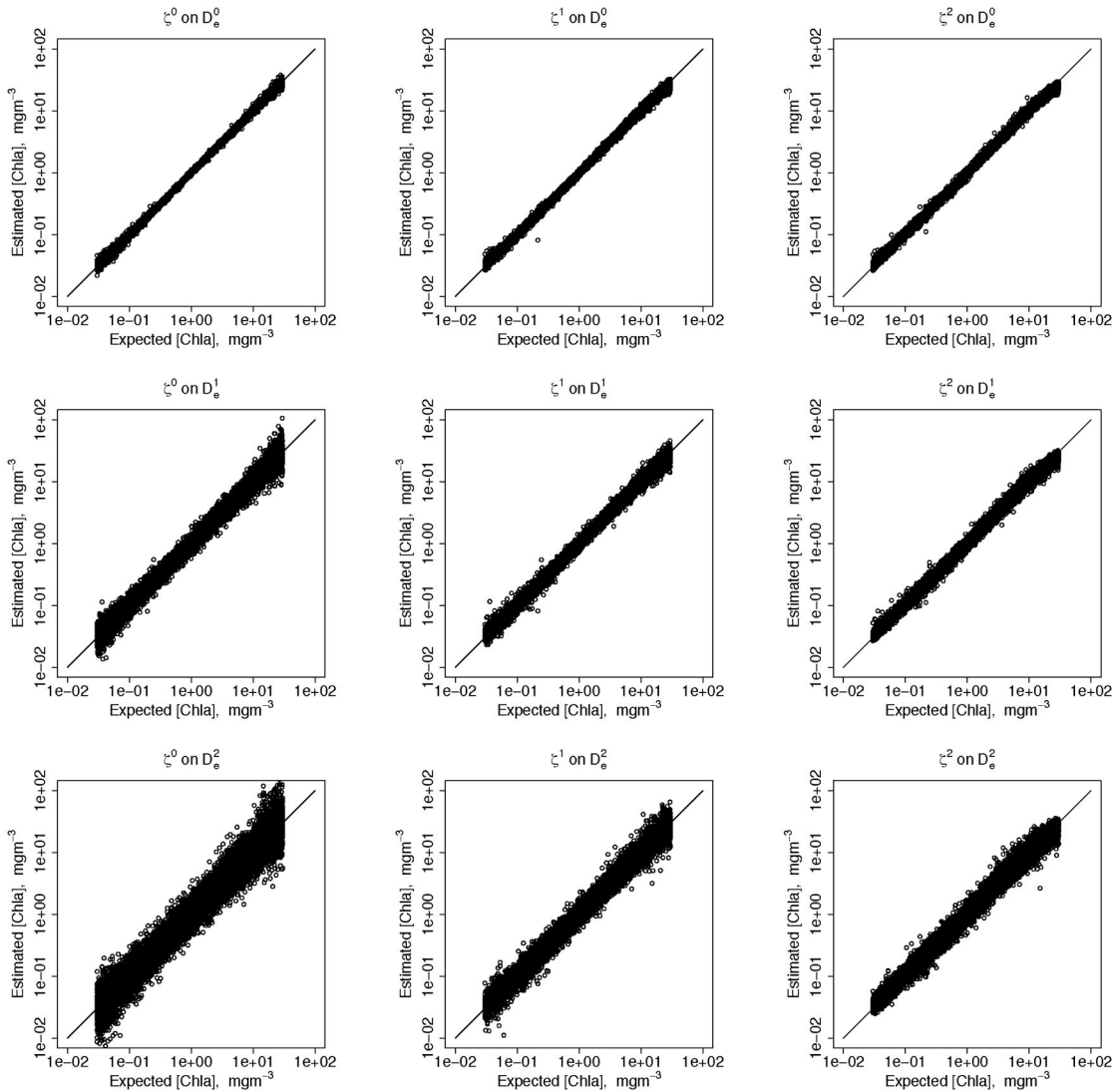
Choosing to work with the cosines of the angular

Fig. 1. Estimated versus expected [Chl-a] for fields ζ^0, ζ^1, and ζ^2 for samples in data sets $\mathscr{D}_e{}^0$, $\mathscr{D}_e{}^1$, and $\mathscr{D}_e{}^2$.

variables is somewhat arbitrary but motivated by the fact that radiation transfer is expressed in terms of trigonometric functions of angular variables. The knots for the functional parameters are the elements of $\{1/2; 1\} \times \{1/2; 1\} \times \{-1; 0; 1\}$. This defines a coarse regular $2 \times 2 \times 3$ grid covering T, thus limiting the number of degrees of freedom, as discussed above. So the functional parameters are valued in the linear span of the basis functions associated with multilinear interpolation on this reduced set of knots. Generator function h has been taken as the hyperbolic tangent, a popular choice in ridge function approximation.

Input vector $\mathbf{x}$ is composed of the top-of-atmosphere reflectance in the eight SeaWiFS spectral bands, corrected for molecular scattering effects, molecule–surface interactions, and Sun glint effects as detailed in Subsection 3.A. Output variable y is the logarithm of the chlorophyll-a concentration.

Three ridge function fields, further denoted ζ^0, ζ^1, and ζ^2, were fitted based on the data in $\mathscr{D}^0$ by the stochastic minimization algorithm described above. Similarly to data set notation, a superscript stands for the total amount of noise, in percent, that has been added during the execution of the algorithm, as detailed above. The noise was generated according to Eq. (15).

The number of basis functions of the fields is taken to be $n = 10$. We determined it empirically by fitting fields for different values of n, starting at a low n. The selected field corresponds to the smallest value of n such that, when n is increased by 1, no noticeable change in the performance of the estimation and validation sets is observed.

C. Results

1. Performance Statistics

We considered the following statistics in evaluating the performance: the root-mean-square error (RMSE), the bias in natural logarithm (b_{ln}), and the root-mean-square error in natural logarithm (RMSE$_{ln}$), which
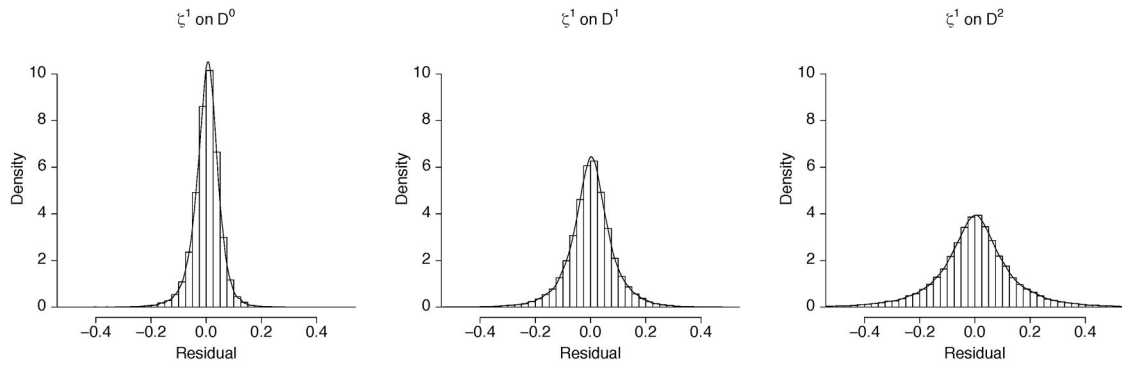
Fig. 2. Histograms of the residuals for field ζ^1 applied to the data set with 0% (left), 1% (middle), and 2% (right) amount of noise. Nonparametric smoothed kernel density estimates are superimposed as curves.

corresponds, up to a second-order term, to the root-mean-square relative error (in fractional units). We define

$$\text{RMSE} = \left[\frac{1}{N} \sum_{i=1}^{N} ([\hat{\text{Chl-}a}]_i - [\text{Chl-}a]_i)^2 \right]^{1/2}, \quad (18)$$

$$b_{\text{ln}} = \frac{1}{N} \sum_{i=1}^{N} (\ln[\hat{\text{Chl-}a}]_i - \ln[\text{Chl-}a]_i), \quad (19)$$

$$\text{RMSE}_{\text{ln}} = \left[\frac{1}{N} \sum_{i=1}^{N} (\ln[\hat{\text{Chl-}a}]_i - \ln[\text{Chl-}a]_i)^2 \right]^{1/2}, \quad (20)$$

where $[\text{Chl-}a]$ and $[\hat{\text{Chl-}a}]$ are, respectively, the expected and estimated chlorophyll-a concentrations, in units of milligrams per cubic meter. The values taken by these statistics for the different ridge function fields are listed in Table 1.

The statistics for ζ^ν take comparable values in estimation and validation sets $\mathscr{D}_e^\nu$ and $\mathscr{D}_v^\nu$, which shows the good generalization performance of the fields. Note that this occurs for all $\nu = 0, 1, 2$. As expected, the best [Chl-a] retrieval is achieved by the field constructed on nonnoisy data (ζ^0) and applied to nonnoisy samples ($\mathscr{D}_e^0$ and $\mathscr{D}_v^0$), for which RMSE_{ln} is of the order of 4%. However, ζ^0 is relatively sensitive to

the presence of noise in the reflectance measurements, as RMSE_{ln} is increased to $\sim$14% and $\sim$27% when reflectances are corrupted by 1% and 2% noise, respectively. This ratio is improved when noise in the fitting of the field is accounted for. For instance, RMSE_{ln} of ζ^1 decreases to less than 9% in $\mathscr{D}_e^1$ and $\mathscr{D}_v^1$, with a slight increase to $\sim$5% in the case of nonnoisy reflectances, and field ζ^2 can process samples with 2% noise slightly better than ζ^1 and with comparable performance for samples with 0% or 1% noise. It appears that ζ^1 achieves a reasonable compromise between accuracy and robustness. It may be noted also that the addition of noise does not introduce a bias. In fact, [Chl-a] retrievals are almost unbiased, with a bias not exceeding 0.4% in all cases.

Plots of estimated versus expected [Chl-a] are given in Fig. 1. The improvement in robustness on passage from ζ^0 to ζ^1 and ζ^2 can be seen from this figure. See, for example, how the points concentrate on the 45° line from the left to right at the bottom of Fig. 1. It may be also noted that the [Chl-a] estimation is accurate in its whole range, i.e., from 0.03 to 30 mg m^{-3}.

2. Analysis of the Residuals

The residual e_i for the ith sample is defined by

$$e_i = \hat{y}_i - y_i$$

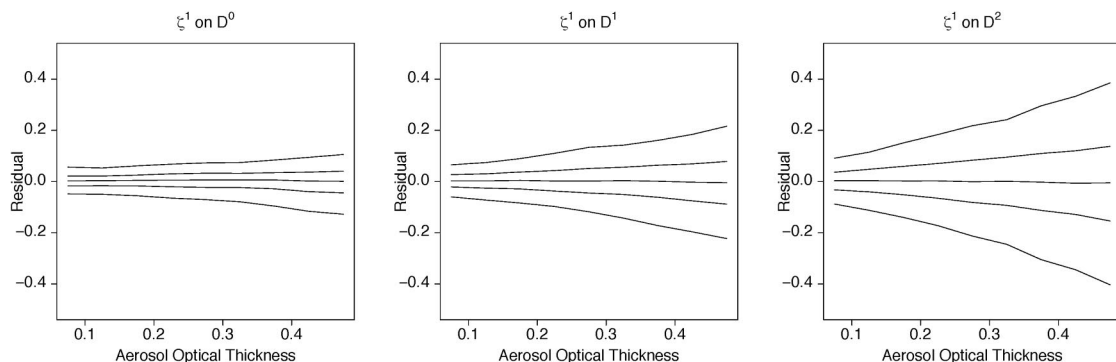$$= \ln([\hat{\text{Chl-}a}]_i) - \ln([\text{Chl-}a]_i),$$

Fig. 3. Conditional quantile curves of orders 5%, 25%, 50%, 75%, and 95% (top to bottom) of the residual given the aerosol optical thickness for field ζ^1 applied to the data set with 0% (left), 1% (middle), and 2% (right) amount of noise.
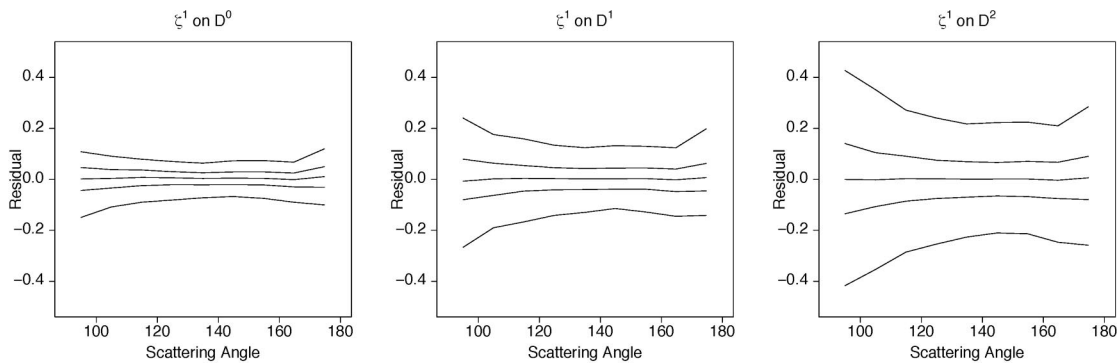
Fig. 4.   Conditional quantiles curves of orders 5%, 25%, 50%, 75%, and 95% (top to bottom) of the residual given the scattering angle for field ζ^1 applied to the data set with 0% (left), 1% (middle), and 2% (right) amount of noise.

i.e., e_i are the errors in natural logarithm. The residuals have an almost null mean, as measured by the statistic $b_{\ln}$. Estimates of the residual distributions for fields ζ_f^1 and ζ_v^1 applied to data sets $\mathscr{D}^\nu$, for $\nu = 0, 1, 2$, are displayed in Fig. 2. These distributions are unimodal, with centered mode, and present symetric shapes with respect to 0.

Additional properties are revealed when one looks at the estimated conditional quantile curves of the residual, given one geophysical parameter. These curves provide a statistical understanding of the influence of one geophysical parameter on the retrieval error. Roughly speaking, the sample quantile of order $\alpha \in (0; 1)$ of N errors is a point q_α such that there are αN points below q_α and $(1 - \alpha)N$ points above q_α (the precise definition has to account for αN not being an integer). The definition of a conditional quantile proceeds in the same way, by considering a variable of the form $e \mid X_g$, i.e., e given X_g, where X_g is a geophysical parameter, say. The conditional quantiles of order 5%, 25%, 50%, 75%, and 95% have been estimated given (separately) the aerosol optical thickness, the scattering angle, and the proportion of one aerosol model in a mixture of two. Conditionally to one geophysical parameter, 90% of the residuals are between $q_{0.05}$ and $q_{0.95}$, and 50% of them are between $q_{0.25}$ and $q_{0.75}$.

The conditional quantile curves given the aerosol optical thickness are plotted in Fig. 3 for fields ζ^1 on

data sets $\mathscr{D}^\nu$; $\nu = 0, 1, 2$. As shown by the leftmost figure, the performance is minimally affected by the aerosol optical thickness in the absence of noise. When the noise level is increased, the influence of the aerosol optical thickness becomes more visible: The spread of the conditional distribution of the residuals increases with the aerosol optical thickness. This occurs at a rate that grows with the noise level, as evidenced by the absolute slopes of the curves.

Plots of conditional quantile curves given the scattering angle are displayed in Fig. 4. They show that the error is slightly larger for lower and higher scattering angles and smaller at intermediate scattering angles where the aerosol phase function exhibit its lowest values, i.e., scattering angles for which the aerosol effect on the top-of-atmosphere reflectance is generally a minimum.

The performance is minimally affected by the aerosol model, as shown by Fig. 5 in plots of the conditional quantile curves given the proportions of one aerosol model in a mixture of two. This is an important result because standard atmospheric algorithms, which extrapolate to the visible aerosol information obtained in the red and the near infrared, generally fail in the presence of absorbing aerosols. The fact that information from all spectral bands, including the bands in the blue that are sensitive to aerosol absorption, is processed by the function fields certainly explains their
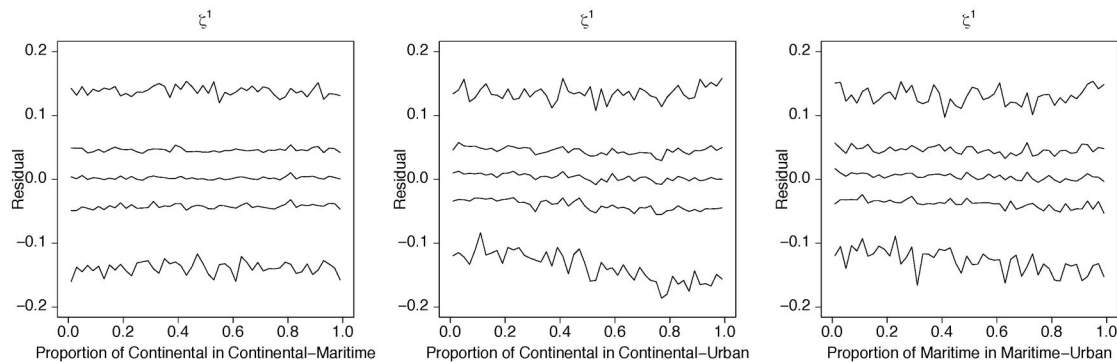

Fig. 5.   Conditional quantiles curves of orders 5%, 25%, 50%, 75%, and 95% (top to bottom) of the residual given the proportion of one aerosol model in a mixture of two of types, Continental, Maritime, and Urban. These curves correspond to 1% noisy data.
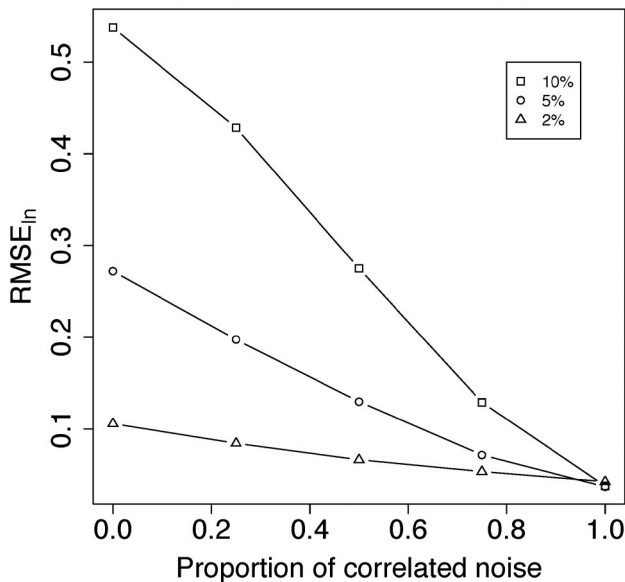
Fig. 6. RMSE in natural logarithm evaluated in $\mathcal{D}^0$ as a function of the proportion of correlated noise for function fields built with 2%, 5%, and 10% total amount of noise.

Statistic	ζ^0	ζ^1	ζ^2
RMSD	0.455	0.434	0.416
MD_{ln}	-1.121	-0.200	-0.275
$RMSD_{ln}$	1.041	0.608	0.549

[a]Comparison statistics with SeaDAS-derived [Chl-a] for fields ζ^ν; $\nu = 0, 1, 2$.

good performance when aerosols are strongly absorbing (urban model).

3. Influence of the Noise Structure
As mentioned in Section 1, the regression function of y on noisy reflectance $\tilde{\mathbf{x}}$ acquired in geometry $\mathbf{t}$ depends, by definition, on the noise distribution. We investigated its influence on the theoretical performance by varying the proportion of correlated noise in a mixture of correlated and uncorrelated components, i.e., by considering the following family of noise schemes:

$$\tilde{\mathbf{x}} = \mathbf{x} + p_c \nu^c \mathbf{x} + (1 - p_c)(\nu_1^{uc} x^1, \ldots, \nu_8^{uc} x^8)^t, \quad (21)$$

where $\tilde{\mathbf{x}}$ is a noisy version of reflectance $\mathbf{x}$, $p_c \in [0; 1]$ is the proportion of correlated noise, and ν^c, $\nu_1^{uc}, \ldots, \nu_8^{uc}$ are random variables following $\mathcal{U}_{[-\nu/100; \nu/100]}$, with ν the total amount of noise, in percent. This family includes as a special case the noise scheme defined by Eq. (15) for $p_c = 1/2$. We constructed several function fields with p_c varying from 0 to 1 in steps of 1/4 and for ν equal to 2%, 5%, and 10%. The RMSE in natural logarithm, evaluated on $\mathcal{D}_e^0$, is plotted in Fig. 6 as a function of p_c for the three levels of ν. It appears that the error is weakly dependent on the noise level when the noise is purely uncorrelated ($p_c = 1$) and that the error rapidly increases with the noise level when the noise is purely uncorrelated ($p_c = 0$), to exceed 50% at 10% of added noise. Also, for a fixed noise level, the error decreases with p_c.

Though in reality the nature of the correlations between the noise components might be more subtle and diverse than in the noise scheme considered here, these results confirm the importance of the noise structure. It is clear that the performance would ben-

efit from an estimation of the noise distribution and of its dependency on the geophysical parameters. The lack of such an estimate may be circumvented by the use of a strongly uncorrelated noise but at the expense of a rapid increase in theoretical error.

4. Application to SeaWiFS Imagery
The function field methodology has been applied to the processing of a typical SeaWiFS image aquired on day 323 (19 November) of year 2002 above southern California. Based on the results of the standard Sea-WiFS processing algorithm (SeaDAS), chlorophyll-a concentration in the image varies from 0.05 to 5 mg m^{-3} (i.e., by 2 orders of magnitude), and aerosol optical thickness from ~0.01 to ~0.1 at 865 nm. Surface wind is weak everywhere, with a speed less than 5 ms^{-1}, and the viewing zenith angle varies from 20° to 55°, Sun zenith angle from 51° to 55°, and relative azimuth angle from 20 to 70 degrees. Owing to this angular geometry, Sun glint is not present in the image.

The three function fields ζ^ν, $\nu = 0, 1, 2$, constructed above were applied to this image, and the [Chl-a] predictions were compared with the SeaDAS predictions. Table 2 summarizes the comparison statistics. There is an important negative mean difference (above 20%), and the root-mean-square differences in natural logarithm ($RMSD_{ln}$) are all greater than 55%. These discrepancies in [Chl-a] are explained as being due to differences between the top-of-atmosphere reflectance measured by SeaWiFS and computed by the radiative transfer code or, equivalently, between the respective reflectances corrected for molecular scattering effects, molecule–surface interactions, and Sun glint; see Subsection 3.A. The differences may be due to several factors, such as accuracy of the radiative transfer code; the aerosol models used in the simulations, which for a given angstrom coefficient may not yield the same single-scattering albedo and phase function as the aerosol models used in SeaDAS; and the bio-optical model. One can quantify these differences, which can be considered noise in the simulated data, by plugging into the radiative transfer code the SeaDAS estimates of [Chl-a] and aerosol parameters and comparing the SeaWiFS reflectance with the simulated reflectance; this was accomplished for 2000 pixels of the image selected randomly from a total of 78,298 pixels, and the results are summarized in Table 3. The rms difference between the two types of reflectance varies from 0.005 at 412 to 0.001 at 865. Compared with the corrected reflectance, the rms dif-

	λ (nm)							
Parameter	412	443	490	510	555	670	765	865
RMSD	0.0052	0.0037	0.0045	0.0038	0.0013	0.0018	0.0017	0.0010
Mean	0.0139	0.0133	0.0130	0.0097	0.0068	0.0019	0.0009	0.0007
r	0.7683	0.8642	0.7844	0.4265	0.8226	0.8993	0.9276	0.9696

ference is large, especially at wavelengths above 555 nm, where on average the corrected reflectance is similar to or less than the rms difference. But the two types of reflectance are relatively well linearly correlated. Therefore the major part of the differences may be explained by an affine map of the corrected reflectance. Hence if the noise level is large, it is, conditionally to the corrected reflectance, strongly correlated. The results for the influence of the noise structure (Subsection 3.D) suggest that such a type of noise (high level, strongly correlated) might be handled with accuracy by an adequately fitted function field, i.e., when the appropriate noise scheme is used during fitting. This possibility was investigated as described next.

The differences e_i between SeaWiFS and simulated reflectance for the $n = 2000$ pixels selected randomly in the image are considered independent samples of noise and are used to estimate the noise distribution nonparametrically, i.e., without assuming any special parametric form of the distribution. There are several classic techniques for this purpose, which we do not review; let us just consider the empirical distribution, defined by

$$\mathbb{P}_n = \frac{1}{n} \, \delta_{e_i}, \qquad (22)$$

where δ_x is the probability distribution that is degenerate at x. This distribution assigns a mass $1/n$ to each sample error e_i. Compared with a smoothed kernel estimate, it has the advantage of providing a computationally fast and fairly simple way of generating noise samples: One observation drawn from $\mathbb{P}_n$ is obtained by uniform random selection of one of the e_i.

A function field, further denoted ζ^n, was constructed with $\mathbb{P}_n$ used as the noise distribution during the execution of the stochastic fitting algorithm. The performance of ζ^n on $\mathscr{D}^0$ is comparable to that of ζ^1, with $b_{\mathrm{ln}} = -0.005$ and $\mathrm{RMSE}_{\mathrm{ln}} = 0.069$. When they are applied to the SeaWiFS image, the [Chl-a] predictions by ζ^n differ from the SeaDAS by 15.7% on average ($\mathrm{RMSD}_{\mathrm{ln}} = 0.157$) and are weakly biased, with an average difference in natural log of -0.019. The SeaDAS and ζ^n [Chl-a] predictions are strongly correlated with a correlation coefficient of 0.931 (0.977 when measured in natural logarithm). The [Chl-a] images are given in Fig. 7. The chlorophyll-a concentration is predicted by ζ^n with lower values than by the SeaDAS in the case of high concentration values, as may also be seen from the histograms in Fig. 8. Nevertheless, these histograms exhibit similar shapes, with a corresponding principal mode and a second lower mode in the upper tail, indicating relatively good agreement between the two distributions. In fact, most of the differences in chlorophyll distribution are located in the upper and lower parts of the

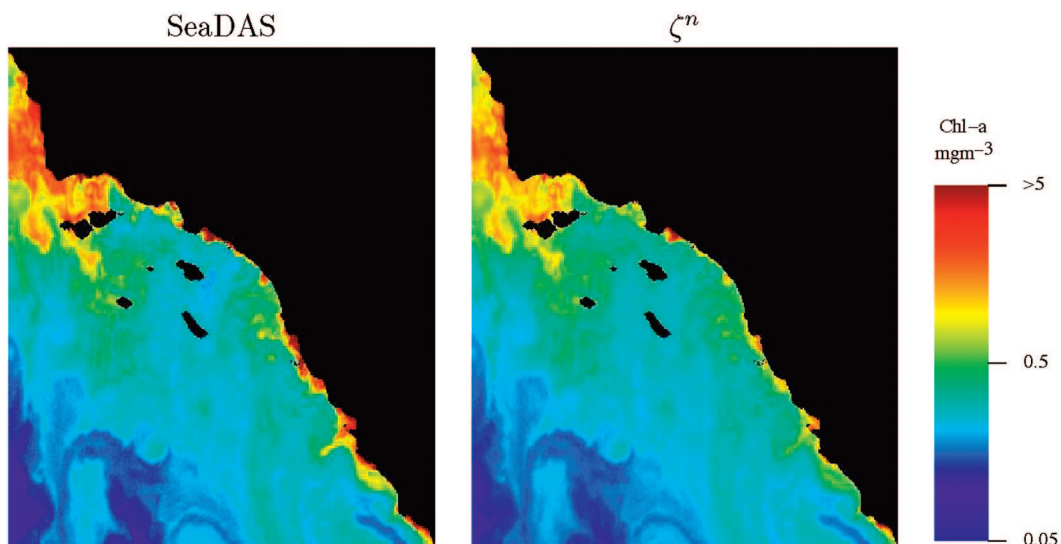Fig. 7. (Color online) Chlorophyll-a concentration predicted by the SeaDAS and ζ^n. They differ by $\sim$16% on average.
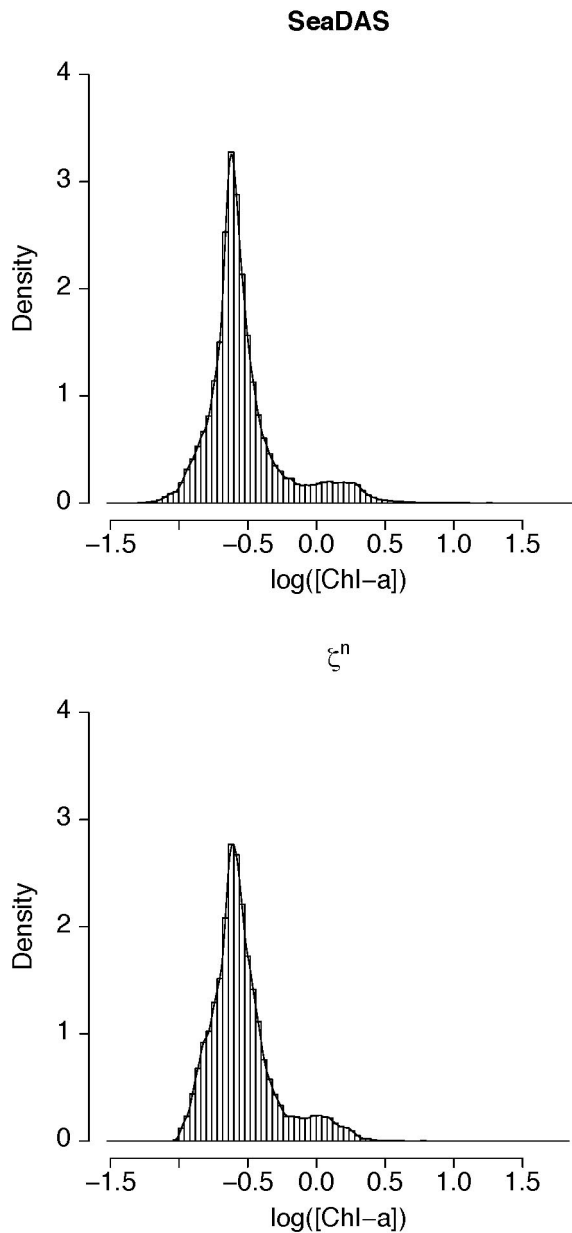
SeaDAS

ζ^n

Fig. 8. Histograms of $\log_{10}([\text{Chl-}a])$ values estimated by the Sea-DAS and ζ^n, with a smoothed kernel density estimate superimposed. They exhibit similar shapes, with a dominant mode and a second lower mode in the upper tail.
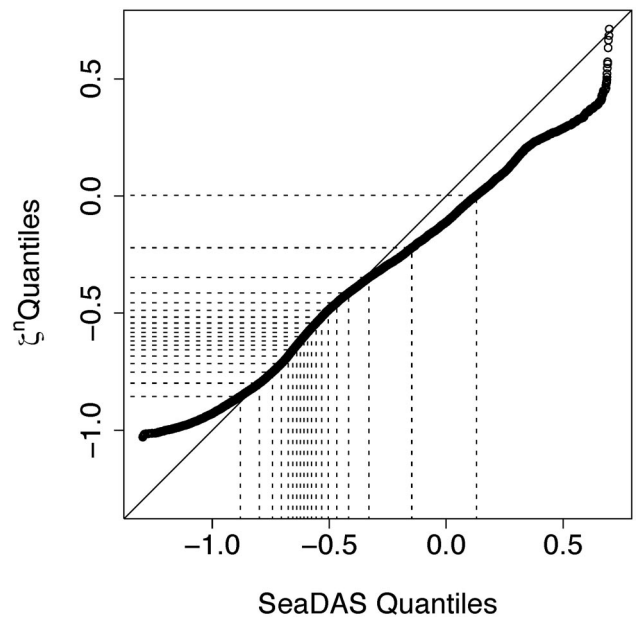

Fig. 9. Field ζ^n versus SeaDAS quantiles of the $\log_{10}([\text{Chl-}a])$ distribution, with a 45° line added. The dashed lines give quantiles of order 5% to 95% in steps of 5% (see Table 4 for the precise values). There is good agreement between the two distributions in the domain that comprises quantiles of orders 5% through 85%, which represents 80% of the points. So most of the differences are located in the tails.

These results suggest that large levels of noise may be handled efficiently by function fields, provided that the noise distribution on the top-of-atmosphere reflectance is known or estimated. As mentioned above, this can be achieved by use of measurements of geophysical variables to compute the top-of-atmosphere reflectance (with the radiative transfer code used in the function field construction), but for generalization the data would have to include the expected range of geophysical variables and angular geometry. The results also show that, compared with existing schemes, the methodology has potential for improving the [Chl-a] retrievals, in terms of accuracy and robustness, especially in the presence of absorbing aerosols and for high [Chl-a] values. Naturally, a more comprehensive assessment would require evaluation of several images representative of the various conditions that may be encountered in the real world. Here, the focus is on theory and methodology, and only one application example is given as a demonstration of feasibility.

5. Conclusions and Perspectives

In the ocean-color inversion problem, the collected information (reflectance) is influenced by some conditioning variables (angular geometry), and this is the reason why this problem may be defined as being a collection of similar inverse problems. Embedding this peculiarity in a regression model, which performs direct mapping between the reflectance and a geophysical variable such as the chlorophyll-a concentration, leads to the setting of function fields. In this case, function fields may show improvement over

chlorophyll range in the image, as revealed by the quantile-versus-quantile plot in Fig. 9. In a quantile-versus quantile plot, the quantiles of one distribution are plotted versus the quantiles of the other, and thus a graph close to the $y = x$ line indicates that the two distributions are similar. The quantiles of orders 5% to 95% in steps of 5% are marked by dashed lines in Fig. 9, and their values are listed in Table 4. They take comparable values for orders from 5% to 85%, indicating good agreement between the two distributions, so the SeaDAS and ζ^n chlorophyll-a predictions are similarly distributed in approximately 80% (in probability) of its range; the differences are located in the lower (below $q_{0.05}$) and upper (above $q_{0.85}$) tails.

Table 4. Quantiles of Orders 5% to 95% In Steps of 5% of the Distribution of $\log_{10}[$Chl-$a]$ Predicted by SeaDAS and ζ^n

	Predicted Quantile	
Order (%)	SeaDAS	ζ^n
5	−0.880	−0.855
10	−0.798	−0.798
15	−0.743	−0.752
20	−0.704	−0.715
25	−0.675	−0.682
30	−0.655	−0.656
35	−0.638	−0.636
40	−0.623	−0.618
45	−0.608	−0.600
50	−0.592	−0.582
55	−0.575	−0.563
60	−0.555	−0.542
65	−0.532	−0.517
70	−0.504	−0.488
75	−0.467	−0.455
80	−0.416	−0.413
85	−0.330	−0.348
90	−0.147	−0.220
95	0.129	0.002

a general multivariate regression technique that uses explanatory and conditionning variables as covariates, because the correlations between the two types of variable are accounted for by construction, yet this possibility remains to be investigated in detail. One may obtain a continuous field of regression models, such as the ridge function fields introduced in this paper, by varying the coefficients of a parametric regression model with respect to the angular variables. The formalism of function fields is much more general, however, as the definition of a function field valued in a set $\mathcal{M}$ of functions does not depend on some special parameterization of $\mathcal{M}$ Thus there is the possibility of studying and developing appropriate methodologies for this type of model in an intrinsic manner, i.e., one that does not depend on the parameterization. This is an important perspective, as the invariance by reparameterization of statistical methodologies has long been recognized as being a desirable goal, even if it is generally difficult to reach, because it requires knowledge of the geometry of the family of models under consideration, including, at least, a bijective parameterization. When such a bijection is missing, one may consider instead a given continuous surjection and study the properties of the induced function fields. This is the methodology that we have investigated for fields of ridge functions, which have been shown to span a dense set. The density property is interesting in the problem, first because little is known about the regression function to be estimated and second because it has not been proved that a function of the corrected reflectance may provide an accurate prediction of [Chl-a], i.e.,

that the problem may be solved accurately in a least-squares sense.

The ridge function fields designed in this study to predict chlorophyll-a concentrations from corrected reflectance have shown promising theoretical performance. First, the chlorophyll-a concentration is predicted accurately in the whole range 0.03–30 mg m^{-3}, with relative rms errors as low as 4% in the absence of noise and increasing with noise to 14% in the presence of 2% noise. This is a definite improvement with respect to standard two-step algorithms, which generally exhibit degraded performance at high chlorophyll-a concentrations. Second, the prediction is fairly independent of aerosol type, even for strongly absorbing aerosols. In such aerosol situations the standard atmospheric correction fails, especially in the blue where the retrieved marine reflectance is too low or even negative, leading to overestimation of [Chl-a]. Thus the methodology should enable more accurate estimates of chlorophyll-a concentration, hence primary production, to be made in dust-contaminated or polluted oceanic areas. Because the areas involved, including the coastal zone, tend to be highly productive, the gain in accuracy would be significant. Third, [Chl-a] predictions are robust to noise. Yet, and naturally, this robustness is strongly tied to the nature of the correlations, especially for high noise levels. Indeed, low levels of noise can be handled efficiently in the design of a function field by simulation of purely uncorrelated noise, whereas for larger noise levels an estimation of the noise distribution might be necessary, as suggested by the application to the SeaWiFS image. So these theoretical results empirically validate the underlying assumption according to which this inverse problem is solved in a least-squares sense, even in a noisy case. This implies the existence of a special kind of geometrical localization of the classes of corrected reflectances that correspond to different [Chl-a] values: two classes overlap only when they corresponding to close [Chl-a] values, and the classes vary smoothly with [Chl-a] by continuity of the radiation transfer map. The noise in the reflectance has the effect of enlarging these classes and thus of increasing their overlap and, in turn, the prediction error. Intuitively, this effect will be even larger as the noise is distributed along directions that correspond to a [Chl-a] gradient, and this probably explains why it has been found, in the simulations and in the application to the SeaWiFS image, that there are large but strongly anisotropic noises that may be handled without a significant increase in the prediction error.

The application to a SeaWiFS image has led to [Chl-a] predictions comparable with the SeaDAS predictions, i.e., a RMSD of 15.7% and a bias of 1.9%, when the noise distribution is approximately known, which we accomplished by analyzing, for a small subset of pixels in the image, the differences between the SeaWiFS reflectance and simulations from SeaDAS estimates of [Chl-a] and aerosol parameters. Further investigations are required for precise determination of the noise distribution with respect to the radiation

transfer model and the bio-optical model used in constructing the function fields. One possibility is to use concomitant measurements of satellite reflectance and of the intervening atmospheric and oceanic parameters. Naturally, for robust inferences to be drawn, the data set needs to be statistically significant and representative of real-world variability. This might not be feasible because of the large number of parameters involved plus the difficulty in collecting a large amount of *in situ* data. In this respect, the plug-in approach represents an interesting alternative: It may be applied systematically and independently of the radiation transfer code that is used. In fact, the plug-in approach provides one with an upper bound on the difference between measured and computed reflectance. So, when it is applied to a significant number of samples, the plug-in approach does not immediately lead to an estimation of the noise distribution; instead, incomplete information is collected this way. How to account for this kind of information when one is fitting a regression model constitutes an interesting statistical perspective.

The demonstration was made on one SeaWiFS image. Other images acquired in various oceanic and atmospheric situations need to be processed to permit a definitive conclusion about the practical suitability of the methodology. In other respects, the methodology is fast in application, i.e., well adapted to the processing of large amounts of satellite data. As explained above, the methodology developed is rather general and can easily be adapted to other sets of models. One can also extend it to the simultaneous estimation of several, eventually correlated variables, by designing fields of vector-valued maps. In optically complex waters, one may attempt to retrieve not only [Chl-*a*] but also the concentrations of yellow substances and inorganic material. In this case the function field methodology would be interesting under the assumption that the least-squares solution to the inversion problem matches the required accuracy for scientific applications. The relevance of this hypothesis would then be tested, as has been done in the present study for case I waters. Other applications include the retrieval of spectral marine reflectance as well as of the spectral optical properties of individual water constituents.

References

1. H. Gordon, "Removal of atmospheric effects from satellite imagery of the oceans," Appl. Opt. **17,** 1631–1636 (1978).
2. M. Viollier, D. Tanré, and P.-Y. Deschamps, "An algorithm for remote sensing of water color from space," Boundary-Layer Meteorol. **18,** 247–267 (1980).
3. H. Gordon and M. Wang, "Retrieval of water-leaving radiance and aerosol optical thickness over the oceans with SeaWiFS: a preliminary algorithm," Appl. Opt. **33,** 443–452 (1994).
4. H. Gordon, "Atmospheric correction of ocean color imagery in the Earth Observing System era," J. Geophys. Res. **102,** 17,081–17,118 (1997).
5. H. Fukushima, A. Higurashi, Y. Mitomi, T. Nakajima, T. Noguchi, T. Tanaka, and M. Toratani, "Correction of atmospheric effects on ADEOS/OCTS ocean color data: algorithm description and evaluation of its performance," J. Oceanogr. **54,** 417–430 (1998).
6. D. Antoine and A. Morel, "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 (1999).
7. B.-C. Gao, M. Montes, Z. Ahmad, and C. Davis, "Atmospheric correction algorithm for hyperspectral remote sensing of ocean color from space," Appl. Opt. **39,** 887–896 (2000).
8. H. Gordon, O. Brown, R. Evans, J. Brown, R. Smith, and K. Baker, "A semi-analytical radiance model of ocean color," J. Geophys. Res. **93,** 10,909–10,924 (1988).
9. A. Morel and S. Maritorena, "Bio-optical properties of oceanic waters: a reappraisal," J. Geophys. Res. **106,** 7163–7180 (2002).
10. J. O. Reilly, S. Maritorena, B. Mitchelle, D. Siegel, K. Carder, S. Garver, M. Kahru, and C. Mc-Clain, "Ocean colour chlorophyll algorithms for SeaWiFS," J. Geophys. Res. **103,** 24,937–24,953 (1998).
11. D. Siegel, M. Wang, S. Maritorena, and W. Robinson, "Atmospheric correction of satellite ocean color imagery: the black pixel assumption," Appl. Opt. **39,** 3582–3591 (2000).
12. K. Ruddick, F. Ovidio, and M. Rijkeboer, "Atmospheric correction of SeaWiFS imagery for turbid and inland waters," Appl. Opt. **39,** 897–912 (2000).
13. C. Hu, K. Carder, and F. Muller-Karger, "Atmospheric correction of SeaWiFS imagery over turbid coastal waters: a practical method," Remote Sens. Environ. **74,** 195–206 (2000).
14. J.-M. André and A. Morel, "Atmospheric corrections and interpretation of marine radiances in CZCS imagery, revisited," Oceanol. Acta **14,** 3–22 (1991).
15. P. Land and J. Haigh, "Atmospheric correction over case 2 waters using an iterative fitting algorithm," Appl. Opt. **35,** 5443–5451 (1996).
16. R. Fraser, S. Mattoo, E.-N. Ych, and C. McClain, "Algorithm for atmospheric and glint corrections of satellite measurements of ocean pigment," J. Geophys. Res. **102,** 17,107–17,118 (1997).
17. H. Gordon, T. Du, and T. Zhang, "Remote sensing of ocean color and aerosol properties: resolving the issue of aerosol absorption," Appl. Opt. **36,** 8670–8684 (1997).
18. F. Zhao and T. Nakajima, "Simultaneous determination of water reflectance and aerosol optical thickness from Coastal Zone Color Scanner measurements," Appl. Opt. **36,** 6949–6956 (1997).
19. R. Chomko and H. Gordon, "Atmospheric correction of ocean color imagery: use of the Junge power-law aerosol size distribution with variable refractive index to handle aerosol absorption," Appl. Opt. **37,** 5560–5572 (1998).
20. C. Moulin, H. Gordon, R. Chomko, V. Banzon, and R. Evans, "Atmospheric correction of ocean color imagery through thick layers of Saharan dust," Geophys. Res. Lett. **28,** 5–8 (2001).
21. T. Schroeder, J. Fisher, M. Shaale, and F. Fell, "Artificial neural network based atmospheric correction algorithm: application to MERIS data," in *Ocean Remote Sensing and Applications,* R. Frouin, Y. Yuan, and H. Kawamura, eds., Proc. SPIE **4892,** 124–132 (2003).
22. C. Jamet, "Inversion neuro-variationelle des images de la couleur de l'océan. Restitution des paramètres optiques des aérosols et de la concentration en chlorophylle pour les eaux du cas," Ph.D. dissertation (Université Pierre et Marie Curie, 2004).
23. L. Keiner and C. Brown, "Estimating oceanic chlorophyll con-

centrations with neural networks," Int. J. Remote Sens. **20**, 189–194 (1999).

24. L. Gross, S. Thiria, R. Frouin, and B. Mitchell, "Artificial neural networks for modeling the transfer function between marine reflectance and phytoplankton pigment concentration," J. Geophys. Res. **105**, 3483–3495 (2000).

25. H. Schiller and R. Doerffer, "Neural network for emulation of an inverse model—operational derivation of Case II properties from MERIS data," Int. J. Remote Sens. **20**, 1735–1746 (1999).

26. A. Morel and B. Gentilli, "Diffuse reflectance of oceanic waters," Appl. Opt. **32**, 6864–6879 (1993).

27. W. Stahel, "Purposes and strategies in regression analysis," J. Statist. Planning Inference **122**, 175–186 (2004).

28. V. Y. Lin and A. Pinkus, "Fundamentality of ridge functions," J. Approx. Theory **75**, 295–311 (1993).

29. B. Vostrecov and M. Kreines, "Approximation of continuous functions by superpositions of plane waves," Dokl. Akad. Nauk SSSR **2**, 1237–1240 (1961).

30. A. Barron, "Universal approximation bounds for superpositions of a sigmoidal function," IEEE Trans. Inform. Theory **39**, 930–945 (1993).

31. V. Maiorov, "On best approximation by ridge functions," J. Approx. Theory **99**, 68–94 (1999).

32. Y. Makovoz, "Random approximants and neural networks," J. Approx. Theory **85**, 98–109 (1996).

33. M. Burger and A. Neubauer, "Error bounds for approximation with neural networks," J. Approx. Theory **112**, 235–250 (2001).

34. B. Pelletier, "Approximation by ridge function fields over compact sets," J. Approx. Theory **29**, 230–239 (2004).

35. W. Light, "Ridge functions, sigmoidal functions and neural networks," in *Approximation Theory VII* (Academic, 1992), pp. 163–206.

36. E. Vermote, D. Tanré, J.-L. Deuze, M. Herman, and J.-J. Morerette, "Second simulation of the satellite signal in the solar spectrum: an overview," IEEE Trans. Geosci. Remote Sens. **35**, 675–686 (1997).

37. D. Tanré, M. Herman, P.-Y. Deschamps and A. de Leffe, "Atmospheric modeling of space measurements of ground reflectances, including bidirectional properties," Appl. Opt. **18**, 3587–3594 (1979).

38. P.-Y. Deschamps, M. Herman, and D. Tanré, "Modélisation du rayonnement solaire réfléchi par l'atmosphère et la terre entre 0.35 et 4 μm," European Southern Observatory contract Rep. 4393/80/F/DD(SC) (University of Lille, 1983).

39. K. Ding and H. Gordon, "Analysis of the influence of $O_2 A$-band absorption on atmospheric correction of ocean-color imagery," Appl. Opt. **34**, 2068–2080 (1995).

40. A. Morel and S. Maritorena, "Bio-optical properties of oceanic waters: a reappraisal," J. Geophys. Res. **106**, 7173–7180 (2001).

41. P. Kocpke, "Effective reflectance of oceanic whitecaps," Appl. Opt. **21**, 2845–2854 (1984).

42. R. Frouin, M. Schwindling, and P.-Y. Deschamps, "Spectral reflectance of sea foam in the visible and near infrared: *in situ* measurements and remote sensing implications," J. Geophys. Res. **101**, 14,361–14,371 (1996).

43. C. Cox and W. Munk, "Measurements of the roughness of the sea surface from photographs of the Sun's glitter," J. Opt. Soc. Am. **44**, 838–850 (1954).