

Fields of nonlinear regression models for inversion of satellite data

Bruno Pelletier

Laboratoire de Mathématiques Appliquées, Université du Havre, Le Havre, France

Robert Frouin

Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California, USA

Received 27 February 2004; accepted 16 July 2004; published 19 August 2004.

[1] A solution is provided to a common inverse problem in satellite remote sensing, the retrieval of a variable y from a vector \mathbf{x} of explanatory variables influenced by a vector \mathbf{t} of conditioning variables. The solution is in the general form of a field of nonlinear regression models, i.e., the relation between y and \mathbf{x} is modeled as a map from some space to a subset of a function space. Elementary yet important mathematical results are presented for fields of shifted ridge functions, selected for their approximation properties. These fields are shown to span a dense set and to inherit the approximation properties of shifted ridge functions. A serious mathematical difficulty regarding the practical construction of continuous fields of shifted ridge functions is pointed out; it is circumvented while providing grounding to a large class of construction methodologies. Within this class, a construction scheme that builds upon multilinear interpolation is described. When applied to the retrieval of upper-ocean chlorophyll-a concentration from space, the solution shows potential for improved accuracy compared with existing algorithms. *INDEX TERMS:* 3260 Mathematical Geophysics: Inverse theory; 4275 Oceanography: General: Remote sensing and electromagnetic processes (0689); 3210 Mathematical Geophysics: Modeling; 4847 Oceanography: Biological and Chemical: Optics; 4855 Oceanography: Biological and Chemical: Plankton. **Citation:** Pelletier, B., and R. Frouin (2004), Fields of nonlinear regression models for inversion of satellite data, *Geophys. Res. Lett.*, 31, L16304, doi:10.1029/2004GL019840.

1. Introduction

[2] A statistical model aims at explaining an exogenous variable y from several explanatory variables x_1, \dots, x_n . In the case where x_1, \dots, x_n are deterministic variables, it expresses the dependence of the expected value $E[y]$ on the explanatory variables and an unknown parameter vector ω , as a function $f(x_1, \dots, x_n; \omega)$. In the random case, the model is written conditionally to the observations, i.e., $E[y]$ is replaced by the conditional expected value $E[y|x_1, \dots, x_n]$. The function f is called the link function between y and x_1, \dots, x_n and, depending on its expression, defines a linear or non-linear regression statistical model. Models such as perceptrons, falling in the class of so-called ridge constructions, achieve this statistical modeling goal with several well-known interesting properties. Let us just mention the density or universal approximation property [Cybenko, 1989; Lin and Pinkus, 1993], and the results related to the approximation rate, including the dimension-independent

upper bound [Barron, 1993; Burger and Neubauer, 2001; Makovoz, 1998], and the asymptotic expression obtained by Maiorov [1999].

[3] In this vein, we focus on a slightly different regression problem, for which we propose a modified solution, based on ridge function approximants, that inherits the interesting mathematical properties mentioned above. This problem still consists in explaining y from x_1, \dots, x_n , but with the difference that, in fact, only some of the x_i , say x_1, \dots, x_d ($d < n$), convey information about y , while the remaining variables act as parameters, or conditioning variables, in the sense that they influence the link function between y and the true informative variables x_1, \dots, x_d .

[4] Typical examples of this kind of problem are found in geosciences, where the observed data may depend on several angular variables that define the geometry of the observation process. They include the retrieval of ocean color and aerosols from reflectance measurements in the visible and near infrared, and the retrieval of wind speed, salinity, and sea surface temperature from brightness temperature measurements at microwave wavelengths. In ocean color remote sensing, the objective is to estimate the concentration of oceanic constituents, such as phytoplankton chlorophyll-a. The informative variables x_1, x_2, \dots, x_d , in this case the top-of-atmosphere reflectance measurements, depend continuously on the angular variables that characterize the positions of the observing satellite and of the Sun relatively to the target on the Earth's surface. Hence these angular variables, which obviously do not carry any information about the chlorophyll-a concentration, have to be taken into account, for the link function between chlorophyll-a concentration and x_1, x_2, \dots, x_d depends on them.

[5] For this kind of problem, it seems natural to separate the variables being effectively informative with respect to y , from the conditioning variables. We shall denote by \mathbf{x} the d -dimensional vector of informative variables, and by \mathbf{t} the p -dimensional vector of conditioning variables. The proposed solution consists in attaching to \mathbf{t} a nonlinear regression model explaining y from \mathbf{x} , and where we demand that the attachment vary smoothly in \mathbf{t} . This approach yields a field of nonlinear regression models over the set of permitted values for \mathbf{t} .

[6] The paper is organized as follows. In section 2, the problem of interest is stated more formally, and fields of nonlinear regression models are defined. In section 3, construction schemes of such a model from scattered data are presented. In section 4, results obtained by applying this methodology to ocean color remote sensing are discussed.

Finally, conclusions are given, as well as perspectives on future work.

2. Function Fields and Nonlinear Regression

Model Fields

[7] Let \mathbf{x} be a vector of explanatory variables, let \mathbf{t} be a vector of conditioning variables, and let y be the real variable to be explained. Let X and T be the sets of permitted values for \mathbf{x} and \mathbf{t} , respectively. We consider statistical models of the following form:

$$y = f_t(\mathbf{x}) + \epsilon, \tag{1}$$

where for each $\mathbf{t} \in T$, f_t is an element of a subset \mathcal{M} of $\mathcal{C}(X)$, the set of continuous real valued functions on X , and ϵ is a random variable of null mean and finite variance σ^2 that is not correlated with \mathbf{x} . Hence in this model, \mathbf{x} carries information about y , while \mathbf{t} does not, but the link function between y and \mathbf{x} depends on \mathbf{t} . The definition of the set \mathcal{M} will be stated later.

[8] To study the dependence of f_t on \mathbf{t} , including continuity and regularity, we introduce the notion of a function field over T . We shall assume that X is locally compact and Hausdorff, and that T is compact, metric and Hausdorff. A space S is said to be compact if every open covering of S has a finite subcovering, and called Hausdorff if, for any two points $x \neq y$, there exists two disjoint open sets U and V with $x \in U$, and $y \in V$. The compact subsets of \mathbf{R}^n are well characterized; they are its closed and bounded subsets. We define a *function field* over T as being a map $T \rightarrow \mathcal{C}(X)$. The set of all continuous function fields over T will be denoted by $(\mathcal{C}(X))^T$. The natural topology on $(\mathcal{C}(X))^T$ is the compact-open topology, which is equivalent to the topology of uniform convergence on compact sets, under the above assumptions on X and T . Furthermore, there is the homeomorphism $\mathcal{C}(X \times T) \cong (\mathcal{C}(X))^T$ [see, e.g., Bredon, 1993, pp. 437–440]. This fact tells us that the elements of $(\mathcal{C}(X))^T$ are in one-to-one correspondence with the elements of $\mathcal{C}(X \times T)$, and that this correspondence is continuous in both directions. Consequently, for each $\zeta \in (\mathcal{C}(X))^T$, there corresponds the unique map ζ_* of $\mathcal{C}(X \times T)$ such that $\zeta_*(\mathbf{x}, \mathbf{t}) = \zeta(\mathbf{t})(\mathbf{x})$, for all $\mathbf{x} \in X$ and $\mathbf{t} \in T$, and conversely. Similarly, the set of all \mathcal{M} -valued continuous function fields over T will be denoted by \mathcal{M}^T , for \mathcal{M} a subset of $\mathcal{C}(X)$.

[9] Returning to the initial problem, equation (1) may be rewritten equivalently as:

$$y = \zeta(\mathbf{t})(\mathbf{x}) + \epsilon, \tag{2}$$

or as

$$y = \zeta_*(\mathbf{x}, \mathbf{t}) + \epsilon, \tag{3}$$

where ζ belongs to \mathcal{M}^T . Hence equation (2) defines a *field of regression models* over T . One may show that if \mathcal{M} is dense in $\mathcal{C}(X)$ and if T is as above, then \mathcal{M}^T is dense in $(\mathcal{C}(X))^T$.

[10] Herein, we shall be interested in the case where the model set \mathcal{M} is the set spanned by functions of the

ridge form. A ridge function on \mathbf{R}^d is a function of the form $h(\mathbf{a}\mathbf{x} + b)$, where $h \in \mathcal{C}(\mathbf{R})$, $\mathbf{a} \in \mathbf{R}^d$ and $b \in \mathbf{R}$. Hence we consider the set $\mathcal{M} = \cup_n \mathcal{M}_n$, where

$$\mathcal{M}_n = \left\{ \sum_{i=1}^n c_i h(\mathbf{a}_i \mathbf{x} + b_i), b_i, c_i \in \mathbf{R}, \mathbf{a}_i \in \mathbf{R}^d \right\}. \tag{4}$$

As mentioned in the introduction, \mathcal{M} is dense in $\mathcal{C}(\mathbf{R}^d)$, in the topology of uniform convergence on compact subsets [Lin and Pinkus, 1993]. Let us introduce some notations. Each element of \mathcal{M}_n depends on parameters c_i, \mathbf{a}_i, b_i , for $i = 1, \dots, n$, that we shall summarize by a vector θ_n . The elements of \mathcal{M}_n will be denoted by $f(\cdot; \theta_n)$. Let Θ_n be the set of allowable values for θ_n , i.e., $\Theta_n = \prod_{i=1}^n \mathbf{R} \times \mathbf{R}^d \times \mathbf{R}$, and let $i_n: \Theta_n \rightarrow \mathcal{M}_n$ be the continuous map carrying a parameter vector θ_n to the corresponding model of \mathcal{M}_n . So $i_n(\theta_n)$ is the function $f(\cdot; \theta_n)$, which associates to each $\mathbf{x} \in \mathbf{R}^d$ the real number $f(\mathbf{x}; \theta_n)$.

[11] At this point, a function field ζ is a relatively abstract object: to evaluate the value $\zeta(\mathbf{t})(\mathbf{x})$ at some \mathbf{t} and \mathbf{x} , an explicit representation of ζ is necessary. Consider a function field $\zeta \in \mathcal{M}^T$. Since T is compact, we may assume, without loss of generality, that ζ belongs to \mathcal{M}_n^T for some integer n . We intend to build a continuous function field $\zeta \in \mathcal{M}_n^T$ via a parameter map $\xi: T \rightarrow \Theta_n$ such that $\zeta = i_n \circ \xi$. Let us mention the following difficulties, arising because the map i_n is only a continuous surjection. First, for each $\zeta \in \mathcal{M}_n^T$, there might not exist a continuous map $\xi: T \rightarrow \Theta_n$ such that $\zeta = i_n \circ \xi$. Second, if we proceed conversely by building ζ according to $\zeta = i_n \circ \xi$, where ξ is continuous, we are not sure to get all of \mathcal{M}_n^T when ξ is allowed to vary in all of $\mathcal{C}(T, \Theta_n)$. However, it is easy to prove that the set of continuous function fields $\zeta \in \mathcal{M}^T$ such that

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

for some integer n , $c_i \in \mathcal{C}(T)$, $\mathbf{a}_i \in \mathcal{C}(T, \mathbf{R}^d)$, and $b_i \in \mathcal{C}(T)$, is dense in $(\mathcal{C}(X))^T$. It simply follows from the fundamentality in $\mathcal{C}(X \times T)$ of the set of functions of the ridge form on $X \times T$.

3. Construction Schemes

[12] Let \mathcal{D} be a data set of N samples $(\mathbf{x}_i, \mathbf{t}_i, y_i)$. Based on \mathcal{D} , we are willing to represent the link between y , \mathbf{x} and \mathbf{t} through a field of nonlinear regression models of the form defined in equation (2) where $\zeta \in \mathcal{M}_n^T$ and where \mathcal{M}_n is as in (4). In light of the results stated in the previous section, we present below two methods for constructing ζ via a parameter map $\xi: T \rightarrow \Theta_n$. In both of them, we hypothesize normally distributed residuals, i.e., we assume $\epsilon \sim \mathcal{N}(0, \sigma^2)$, and use the averaged sum of the squared errors $\mathcal{E} = \frac{1}{N} \sum_{i=1}^N (y_i - \zeta(\mathbf{t}_i)(\mathbf{x}_i))^2$ as the natural criterion to be minimized for selecting a function field from the data. The main difference with traditional parametric modeling techniques is that here, the parameters for ζ are maps $T \rightarrow \Theta_n$.

[13] The first method consists in taking a parameterized subset \mathcal{F}_ρ of $\mathcal{C}(T, \Theta_n)$, that contains the constant and affine functions of \mathbf{t} (for the reason previously mentioned), where

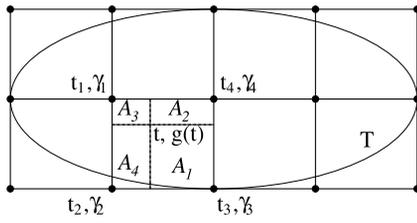


Figure 1. Construction of a real valued piecewise-differentiable map g by multilinear interpolation. The set T is a 2-dimensional ellipsoidal domain covered by a regular 5×3 grid. The value $g(\mathbf{t})$ of g at some point \mathbf{t} in the square with corners located at $\mathbf{t}_1, \dots, \mathbf{t}_4$ is defined by $g(\mathbf{t}) = \sum_{i=1}^4 \frac{A_i}{A_1 + A_2 + A_3 + A_4} \gamma_i$, where the A_i 's are the areas depicted in the diagram. Note that the A_i depend on t . The γ_i are “control points” in the sense that $g(\mathbf{t}_i) = \gamma_i$.

ρ is the parameter vector. Hence the problem reduces to the one of minimizing \mathcal{E} with respect to ρ , which may be achieved, for instance, by means of a stochastic gradient descent algorithm or simulated annealing. However this method may suffer from an inappropriate choice of \mathcal{F}_ρ , which may yield a much more larger n than necessary. The second method, described below, allows one to cope with this issue.

[14] This second method consists in building a sample of a continuous map $\xi: T \rightarrow \Theta_n$ such that the induced field $\zeta := i_n \circ \xi$ minimizes \mathcal{E} . The sketch is as follows. We start by defining a set \mathcal{F}_K of real valued continuous and piecewise-differentiable functions whose domain is a set containing T ; the value of any of these functions is obtained by multilinear interpolation. Next those maps are used to define a function field over T . More precisely assume T is a compact subset of \mathbf{R}^p . Let $\mathbf{t}_1^{\Xi}, \dots, \mathbf{t}_K^{\Xi}$ be K points of \mathbf{R}^p , and the vertices of a regular grid of \mathbf{R}^p , such that T is included in the smallest p -dimensional cube Ξ containing all of the \mathbf{t}_k^{Ξ} . Note that K is the product of p integers $k_i \geq 2$. Hence $T \subset \Xi$, and $\mathbf{t}_k^{\Xi} \in \Xi$, for all $k = 1, \dots, K$. Let $\gamma_1, \dots, \gamma_K$ be K real numbers, and consider those continuous and piecewise-differentiable maps $g \in \mathcal{C}(\Xi)$ such that $g(\mathbf{t}_k^{\Xi}) = \gamma_k$ for all $k = 1, \dots, K$, and defined for all $\mathbf{t} \in \Xi$ such that $\mathbf{t} \neq \mathbf{t}_k^{\Xi} \forall k$ by:

$$g(\mathbf{t}) = \sum_{i=1}^{2^p} \alpha_i(\mathbf{t}) g(\mathbf{t}_{k_i}^{\Xi}). \quad (6)$$

In this equation, the $\mathbf{t}_{k_i}^{\Xi}$ are the 2^p immediate neighbors of \mathbf{t} on the grid, i.e., they are the vertices of a p -cube containing \mathbf{t} , and the coefficients $\alpha_i(\mathbf{t})$ are the coefficients of the standard p -dimensional interpolation procedure on a p -cube. We shall denote by \mathcal{F}_K the set of all such maps. This construction method is illustrated in Figure 1, in the case where T is of dimension 2. Next consider those function fields ζ over T , being the restrictions to T of function fields $\tilde{\zeta}$ over Ξ , of the form defined in equation (5), where the b_i , the c_i , and the components of the \mathbf{a}_i , belong to \mathcal{F}_K . Hence such a function field is parameterized by $n(d+2)K$ real numbers γ_j^i , where $1 \leq i \leq K$ and $1 \leq j \leq n(d+2)$, since $\dim(\Theta_n) = n(d+2)$.

[15] The minimization of \mathcal{E} with respect to them may be performed as follows. First, pick randomly a sample $(\mathbf{x}_t, \mathbf{t}_t, y_t)$

from \mathcal{D} and compute the error e_i of the model for that sample. If \mathbf{t}_t falls on one of the vertices of the grid, say on $\mathbf{t}_{k_i}^{\Xi}$, then the error e_i depends on γ_j^i , $1 \leq j \leq n(d+2)K$. Otherwise, \mathbf{t}_t is different from all the vertices, and if we let $\mathbf{t}_{k_l}^{\Xi}$, $1 \leq l \leq 2^p$, be the 2^p immediate neighbors of \mathbf{t}_t on the grid then the error e_i depends on γ_j^l for $1 \leq l \leq 2^p$ and $1 \leq j \leq n(d+2)K$. So in a second step, modify the appropriate γ_k^j according to the rule:

$$\gamma_k^{j(new)} = \gamma_k^{j(old)} - \eta \frac{\partial e_i}{\partial \gamma_k^j} e_i, \quad (7)$$

where η is a strictly positive scalar. Finally, repeat those steps until convergence. By this method, we obtain from the sets $\mathcal{D}_j^{\Xi} := \{(\mathbf{t}_k, \gamma_k^j); k = 1, \dots, K\}$ a sample of a map ξ yielding the function field $\zeta = i_n \circ \xi$. The advantages with respect to the first method are i) that the grid may be refined during the execution of the algorithm, and ii) that the resulting sample may be used to choose an appropriate model set for ξ that achieves, for example, a higher degree of regularity. Indeed this algorithm performs a stochastic gradient descent and, when the number N of samples tends towards infinity, the resulting field ζ of nonlinear regression models is expected to be a good approximation to the field $E_t[y|\mathbf{x}]$ over T of the $(\mathbf{t}$ dependent) conditional means of y given \mathbf{x} .

4. Application to Ocean Color Remote Sensing

[16] For this problem, we have let $\mathbf{t} = {}^t(\cos \theta_s, \cos \theta_v, \cos \Delta\varphi)$, where θ_s , θ_v , $\Delta\varphi$ are the Sun, view, and relative azimuth angles, respectively, and the set T of permitted values for \mathbf{t} is $[0, 1] \times [0, 1] \times [-1, 1]$. The vector \mathbf{x} is composed of top-of-atmosphere (TOA) reflectances in spectral bands in the visible and near-infrared, centered at 412, 443, 490, 510, 555, 670, 765, and 865 nm (case of the Sea-viewing Wide Field-of-view Sensor). The reflectances are corrected for molecular scattering effects. The variable y is the near surface chlorophyll-a concentration ([Chl-a]). A statistically significant ensemble of about 62,000 realizations of \mathbf{x} , encompassing the major sources of variability (mostly due to the atmosphere) and including maritime, continental, and urban aerosols in varied mixtures, has been generated via intensive use of simulation (multiple runs of a coupled ocean-atmosphere radiative transfer code [Vermote et al., 1997]). In the code, the reflectance of the water body has been modeled according to Morel and Maritorena [2001], with no variability due to phytoplankton type. The vector \mathbf{x} ensemble has been split into data sets \mathcal{D}_v^0 and \mathcal{D}_v^v , used for model construction and validation, respectively. Two fields F and F^v of nonlinear regression models have been built according to the second method, on a $2 \times 2 \times 3$ regular grid, i.e., with vertices belonging to $\{0; 1\} \times \{0; 1\} \times \{-1; 0; 1\}$. The nonlinear regression models attached to \mathbf{t} are elements of \mathcal{M}_n , with $n = 10$, where \mathcal{M}_n is as in (4). The generator function h has been taken as the hyperbolic tangent, a popular choice in the field of ridge approximation. The number $n = 10$ of basis functions has been determined experimentally, by starting at building models with a small value of n and increasing the value gradually until no significant improvement in the quality is noticed. This construction procedure

leads to ridge function fields of the form given by equation (5), where the $\mathbf{a}_i(\mathbf{t})$, $b_i(\mathbf{t})$, and $c_i(\mathbf{t})$, are continuous functions of the angular variables (8 dimensional vector-valued for the former one, and real-valued for the latter two). The fields have been built both on \mathcal{D}_l^0 , but in the case of F^ν , some amount of noise has been added to the data during the execution of the stochastic algorithm. To be as general as possible, this noise scheme has been defined as the sum of spectrally independent and correlated noises, with a total amount of 1%.

[17] To measure the quality of the [Chl-a] estimation, the root mean squared error (*RMS*), the bias in natural logarithm (b_{ln}), and the root mean squared error in natural logarithm (RMS_{ln}), have been used. Their values for fields F and F^ν are summarized in Table 1. The error in [Chl-a] estimation is on the order of 4.2% over the range 0.03–30 mgm^{-3} , in the case of non-noisy data, and 10% in the case of realistic noisy data, which illustrates the efficiency and the robustness of this modeling, as well as its generalization ability. Plots of estimated versus expected [Chl-a] are given in Figure 2, for fields F and F^ν in the non-noisy and noisy cases, showing that the estimation is accurate in the whole range 0.03–30 mgm^{-3} . Performance is minimally affected by aerosol optical thickness and type. These results suggest that the methodology based on ridge approximants has potential for improving the accuracy of [Chl-a] retrievals, since current processing techniques yield, without noise, theoretical errors that may reach 20% in the presence of non- or little absorbing aerosols and larger errors when aerosols are strongly absorbing [Gordon, 1997].

5. Conclusions

[18] Fields of nonlinear regression models allow one to deal with composite data, where some variables are effec-

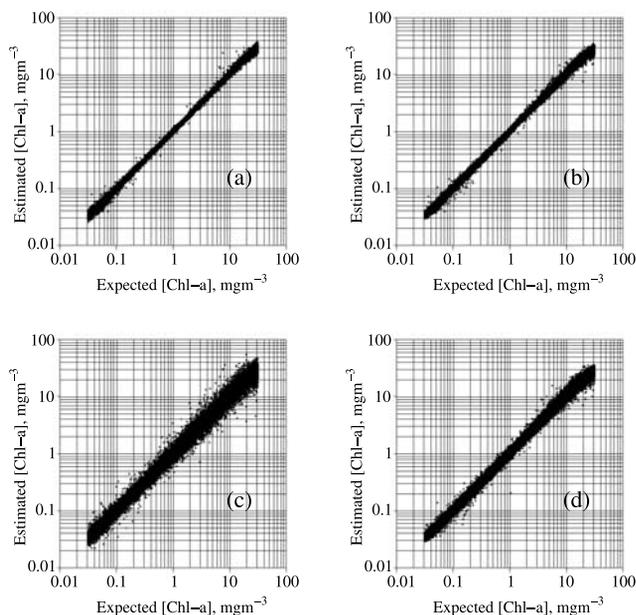


Figure 2. [Chl-a] estimated by fields F and F^ν versus expected [Chl-a] in the case of non noisy TOA reflectance (a for F and b for F^ν) and noisy TOA reflectance (c for F and d for F^ν).

Table 1. Mean Relative Error in [Chl-a] Estimation Evaluated on Data Sets \mathcal{D}_l^0 , \mathcal{D}_v^0 , \mathcal{D}_l^1 , \mathcal{D}_v^1 , for Models F (Built on Non-noisy Data) and F^ν (Built on Noisy Data), Where \mathcal{D}_l^1 and \mathcal{D}_v^1 are Noisy Versions of \mathcal{D}_l^0 , \mathcal{D}_v^0 , Respectively, With a Total Amount of Noise of 1%

	F	F^ν
	\mathcal{D}_l^0	
RMS	0.520	0.836
b_{ln}	-0.000	0.000
RMS_{ln}	0.042	0.068
	\mathcal{D}_v^0	
RMS	0.534	0.859
b_{ln}	-0.001	-0.002
RMS_{ln}	0.042	0.070
	\mathcal{D}_l^1	
RMS	1.672	1.091
b_{ln}	0.000	0.000
RMS_{ln}	0.151	0.104
	\mathcal{D}_v^1	
RMS	1.650	1.118
b_{ln}	-0.001	-0.002
RMS_{ln}	0.151	0.105

tively explanatory, while the others are conditioning, and without having recourse to the product space $X \times T$, which in some cases, such as the ocean color remote sensing problem, may be meaningless. They distinguish themselves from classical parameterized models by the fact that their parameters are functions of the conditioning variables. From this peculiarity follows a mathematical difficulty at constructing dense sets of continuous function fields for some choices of \mathcal{M}_n and \mathcal{M} , namely those subsets being not homeomorphic to some arcwise-connected subset of an Euclidean space, including the sets spanned by at most n functions of the ridge form, and their union. In this particular case, the difficulty may be circumvented, and a large class of methodologies, comprising the presented method based on multilinear interpolation, apply for the practical generation of dense sets of continuous fields of ridge functions.

[19] In this nonlinear regression context, fields of shifted ridge functions are especially interesting when no external, problem-related, knowledge is available since, as shown above, they inherit their interesting approximation properties. The developed methodology is rather general and could be adapted to another set \mathcal{M}_n , the choice of which could be driven by application specific requirements. Extension to the simultaneous explanation of several, eventually correlated variables is possible, by choosing a set \mathcal{M}_n of vector-valued functions. Remote sensing of ocean color is a multi-variate problem, especially complex in coastal and estuarine waters, and one may attempt to retrieve the concentrations of other constituents than phytoplankton, such as yellow substances and inorganic material, or their inherent optical properties. But it would be worth exploring the question of whether or not this approach would be optimal, in the statistical sense of exhaustive description. Intuitively, a preliminary de-correlation of the variables to be explained might lead to an explanation problem of lower complexity.

[20] **Acknowledgment.** This work was supported by the National Aeronautics and Space Administration, by the National Science Foundation, and by the Applied Mathematics Laboratory of the University of Le Havre, France.

References

- Barron, A. (1993), Universal approximation bounds for superpositions of a sigmoidal function, *IEEE Trans. Inf. Theory*, 39(3), 930–945.
- Bredon, G. (1993), *Topology and Geometry*, *Grad. Texts Math.*, vol. 139, Springer-Verlag, New York.
- Burger, M., and A. Neubauer (2001), Error bounds for approximation with neural networks, *J. Approx. Theory*, 112, 235–250.
- Cybenko, G. (1989), Approximation by superpositions of a sigmoidal function, *Math. Control Signals Syst.*, 2, 303–314.
- Gordon, H. (1997), Atmospheric correction of ocean color imagery in the Earth observing system era, *J. Geophys. Res.*, 102, 17,081–17,118.
- Lin, V. Y., and A. Pinkus (1993), Fundamentality of ridge functions, *J. Approx. Theory*, 75, 295–311.
- Maïorov, V. (1999), On best approximation by ridge functions, *J. Approx. Theory*, 99, 68–94.
- Makovoz, Y. (1998), Uniform approximation by neural networks, *J. Approx. Theory*, 95, 215–228.
- Morel, A., and S. Maritorena (2001), Bio-optical properties of oceanic waters: A reappraisal, *J. Geophys. Res.*, 106, 7173–7180.
- Vermote, E., D. Tanré, J.-L. Deuzé, M. Herman, and J.-J. Morcrette (1997), Second simulation of the satellite signal in the solar spectrum: An overview, *IEEE Trans. Geosci. Remote Sens.*, 35, 675–686.

R. Frouin, Scripps Institution of Oceanography, La Jolla Shores Drive, La Jolla, CA 92037, USA. (rfrouin@ucsd.edu)

B. Pelletier, Université du Havre, Laboratoire de Mathématiques Appliquées, F-76063 Le Havre, France.