Chlorophyll profile estimation in ocean waters by a set of artificial neural networks

Fabio Dall Cortivo[1], Ezzat S. Chalhoub[2],
Haroldo F. Campos Velho[2], Milton Kampel[2]
[1] *Institute of Aeronautical Technology (IEFM/ITA)*
   São José dos Campos, São Paulo, Brazil
[2] *National Institute for Space Research (INPE)*
   São José dos Campos, São Paulo, Brazil
   e-mail: ezzat.chalhoub@inpe.br

In this work, we propose a methodology to estimate the profile of chlorophyll concentration from the upwelling radiation at the ocean surface, using a system of artificial neural networks (ANNs). The input patterns to train the networks are obtained from the resolution of the radiative transfer equation, where the absorption and scattering coefficients are represented by bio-optical models, with the profile of chlorophyll concentrations based on a shifted-Gaussian model. In the performed analysis, we used 14 720 profiles of chlorophyll that were generated by attributing two values to the biomass quantity, and by considering two sets of wavelengths and three sets containing the directions in which the radiation emitted at the surface is measured. To be able to recover the chlorophyll profile, we need to use a system of networks that works in a "cascade mode". The first one performs an analysis on the features of the chlorophyll profile from the upwelling radiation and determines which profiles can be recovered. The second and third ANNs act only on those profiles that can be recovered. The second ANN performs estimation of the standard deviation from the upwelling radiation and the chlorophyll concentration at the surface. Finally, the third ANN performs an estimation of the peak depth from the upwelling radiation, the chlorophyll concentration at the surface and the standard deviation estimated by second network. The stopping criteria we adopted was the cross-validation process. The obtained results show that the proposed methodology is quite promising.

Keywords: radiative transfer equation, inverse problems, artificial neural networks, chlorophyll profile concentration, bio-optics, phytoplankton.

1. INTRODUCTION

Ocean color variations are directly connected with the absorption of sunlight by chlorophyll, especially in the blue and green bands. The measurements of ocean color, obtained from orbital sensors on satellites help us to estimate the phytoplankton biomass [18, 33]. Ocean color is a useful source of information regarding some of its features and its study allows determination of the chemical, physical and biologic properties, and helps the selection of better locations for fishing.

As the chlorophyll pigments are inside of the phytoplankton cell, it becomes possible to obtain the amount of these cells from the estimation of chlorophyll concentration, and from this estimation it is possible to get an approximation of the phytoplankton biomass. Marine phytoplanktons are photosynthesizing microscopic organisms that inhabit the upper sunlight layer of almost all oceans and they form the trophic base of the chain food. About one-half of the global primary production is generated by these organisms [17, 45], they affect the abundance and diversity of marine organisms, drive marine ecosystem functionality, and set the upper limits to fishery yields [8]. Furthermore, marine phytoplankton has a strong influence on the climate process [34] and the bio-geochemical cycles, especially in the carbon cycle [42, 43]. Phytoplankton is also the connection between the sunlight energy and the production of marine biological resources that depends on other trophic

levels [9]. The photosynthetic activity of phytoplankton is the first step for the fixation of inorganic carbon into particulate organic carbon [25]. Furthermore, as the phytoplankton is in the trophic base of the oceans chain food, it feeds the species ranging from small crustaceans to large predators. So, we can consider chlorophyll concentration level as parameter that can be used to measure the health of our planet.

Due to the importance of phytoplankton, the scientific community, supported by the governments, has applied their efforts to monitor the oceans color. This monitoring can be done at local scale, by means of the scientific cruises or, globally, by means of the passive sensors on board satellites. The ocean color variations obtained from satellites enable us to know the global distribution of phytoplankton, allowing to produce maps with synoptic, spatial and temporal variabilities of the phytoplankton biomass, partially known by samplings studies done by boats/cruises [2, 16, 28, 30, 51, 52].

The estimation of chlorophyll concentration at the surface from the radiance and/or the reflectance, measured by orbital sensors, is performed by specific algorithms obtained empirically from data sets that contain *in situ* measurements. The first algorithm proposed [15, 16, 19] to process the CZCS (coastal zone color scanner) data was obtained from Nimbus experiment team [1], and it consisted of less than sixty measurements. With an increase of *in situ* measurements and the launch of new orbital sensors, other, more robust algorithms were proposed in order to convert the upwelling radiation at the surface into maps of chlorophyll concentration [27, 36–38]. Those algorithms, in general, produce good estimations for the chlorophyll concentration at the surface for natural waters of Case 1 (offshore). However, they overestimate the chlorophyll concentration for Case 2 waters. This occurs because of the presence of inorganic matter in suspension and dissolved organic matter in Case 2 waters [23].

On the other hand, alternative methodologies, such as artificial neural networks (ANNs), have been used to perform the estimation of chlorophyll concentration at the surface, for Case 1 [20] and Case 2 [44] waters, and they proved to be efficient. Estimation of the chlorophyll concentration close to the surface in Case 1 waters, from the reflectance measured by SeaWiFS, using ANNs of multilayer perceptron (MLP) type, which were trained by back propagation algorithm, can be found in Gross et al. [20]. According to Gross et al. [20] the estimation obtained by ANNs shows a significant reduction of the relative and the absolute errors, when compared with the results obtained by the OC2 algorithm. Applications of ANNs for the estimation of chlorophyll concentration, inorganic matter in suspension and dissolved organic matter, for Case 2 waters, can be found in Schiller and Doerffer [44]. Another application of ANNs to cluster the chlorophyll profiles based on their features was performed by Richardson et al. [41].

Although the proposed algorithms produce accurate maps of the chlorophyll concentration for Case 1 waters, these maps provide information only for the water near the surface layer of. For cases, where the chlorophyll concentration is uniform along the depth, these maps are valid to extract information about the vertical structure of the chlorophyll concentration. However, in most cases, the vertical profiles have a maximum peak along the depth and, in some cases, they are similar to Gaussian curves [26]. A shifted Gauss model to represent the chlorophyll profiles was suggested by Platt et al. [39]. Their model is composed of the Gaussian model proposed by Lewis et al. [26], and a background of chlorophyll concentration is added to the model to represent low concentration of chlorophyll in the water column. In this work, we adopt a shifted Gauss model to represent the behavior of the profiles of the chlorophyll concentration.

Due to the variation of the chlorophyll concentration along the depth and the impossibility of estimating these profiles from proposed algorithms, an alternative methodology to solve the inverse problem was proposed by Souto [46]. His methodology is based on a minimization of a functional, given by the squared differences between the observed and modeled data, and for which a minimum value is sought. The observed data in his methodology were obtained from the resolution of the radiative transfer equation (RTE), where bio-optical models that depend on the profile of chlorophyll concentration were used to represent the absorption [32] and scattering [18] coefficients.

The modeled data were obtained from the model adopted for the formulation of the inverse problem. Good results were obtained by Souto [46] for chlorophyll concentrations down to 20 meters depth and poor results below 20 meters. Another problem associated with his methodology is the extensive computational time spent to solve the inverse problem, even when using parallel codes based on MPI (message passing interface). It becomes difficult to implement his methodology in order to process data for large areas of the ocean. Some improvements in order to get better results in the estimation of chlorophyll profiles were proposed in Souto et al. [47, 48], as well as some improvements were proposed in Souto et al. [49] in order to reduce the computational time spent to solve the problem. However, in the latter case, some mathematical simplifications were considered in the mathematical model.

In this work, we propose a new methodology to estimate the profile of chlorophyll concentration from the upwelling radiation at the surface in Case 1 waters. Our methodology is based on ANN of a MLP type and is able to produce good results with a low computational cost. The ANNs are trained by the quasi-Newton method and three networks that work in a cascade mode are used. The first network, labeled as ANN_{Cte}, is trained especially to perform a pre-classification of the problems considered in this work. The classification, which is based on the upwelling radiation at the surface of the water, is performed in order to separate the problems into two classes. The first class is composed of those problems for which it is possible to determine the chlorophyll profile. The second one is composed of those problems for which it is not possible to determine their profiles. However, for the problems in the second class it is possible to estimate an average value for the chlorophyll profile along the depth. This pre-classification is necessary because there are some features, associated with the physics of the problems and/or our adopted mathematical methodology, which affect the resolution of some problems. These features will be explained in Subsec. 4.2. The input patterns for the network ANN_{Cte} are the upwelling radiation at the surface, and the output patterns are the average value of the profile of the chlorophyll concentration along the depth. So, from the upwelling radiation, the network predicts an average value for the concentration and if this value is greater than a threshold value, the problems are then considered belonging to the first class, otherwise they are in the second class.

As we consider the shifted Gaussian model only two parameters will be estimated by the networks: the standard deviation and the depth of the peak. The other parameters which are present in that model are considered as constant values and they will be explained in Sec. 4.

From the upwelling radiation and the chlorophyll concentration at the surface, the second network, labeled as ANN_{σ}, is used in order to perform the estimation of the standard deviation of the profile of the chlorophyll concentration. The third network, labeled as ANN_{τ_m}, performs the estimation of the depth of the peak considering, as input patterns, the upwelling radiation and the chlorophyll concentration at the surface and the standard deviation estimated by the second network.

The results shown in this work are valid for profiles that respect the behavior of this model. For other oceanic regions, in which that kind of profile is not valid, for example, close to Japan [24], the results obtained by our methodology may not be adequate. A useful study regarding different types of chlorophyll profiles can be found in Uitz et al. [50].

2. Mathematical model

We consider the radiative transfer equation (RTE) to represent the interaction of a light beam (photons) within a body of water. In this case, we need to define the boundary conditions, internal sources and the inherent optical properties to solve the RTE, in order to determine the upwelling radiation at the surface, after the interaction of the light beam within the water body. In many radiative transfer applications, it is reasonable to consider that the variations of water characteristics can only occur vertically within the water column under consideration. Thus, the analyzed geometry can be defined as a plane-parallel geometry. The RTE for this type of geometry

(see Fig. 1), with polar and azimuthal dependency, wavelength dependency, anisotropic scattering and multiregions (which are associated with the depth) is expressed by

$$\mu\frac{\partial}{\partial z}L_r(\tau,\mu,\varphi,\lambda) + L_r(\tau,\mu,\varphi,\lambda)$$

$$= \frac{b_r(\tau,\lambda)}{c_r(\tau,\lambda)}\int\limits_{-1}^{1}\int\limits_{0}^{2\pi}\int\limits_{\lambda} p_r(\cos\Theta,\lambda)L_r(\tau,\mu',\varphi',\lambda')d\lambda'd\varphi'd\mu' + S_0(\tau,\lambda), \qquad (1)$$

subject to the boundary conditions

$$L_1(\tau_0,\mu,\varphi,\lambda) = G^+(\mu,\varphi,\lambda) \qquad (2)$$

and

$$L_R(\tau_R,-\mu,\varphi,\lambda) = G^-(\mu,\varphi,\lambda) \qquad (3)$$

and to the interface conditions, for $r = 1, 2, \ldots, R-1$,

$$L_r(\tau_r,\pm\mu,\varphi,\lambda) = L_{r+1}(\tau_r,\pm\mu,\varphi,\lambda). \qquad (4)$$
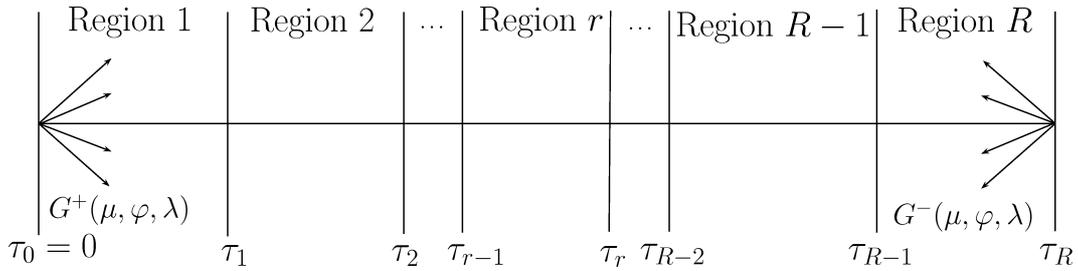

Fig. 1. A system of R regions in a multiregion geometry.

Here, $L_r(\tau,\mu,\varphi,\lambda)$ is the intensity (radiance) of the radiation field in a region r, $\tau \in (0,\zeta)$ the optical variable, with ζ the optical thickness of the medium, $\mu \in [-1,1]$ and $\varphi \in [0,2\pi]$ are, respectively, the cosine of the polar angle and the azimuthal angle, which specify the direction of propagation Θ of the radiation in the medium, and λ is the photon wavelength. In addition, $c_r(\tau,\lambda) = a_r(\tau,\lambda) + b_r(\tau,\lambda)$ is the attenuation coefficient, where $a_r(\tau,\lambda)$ and $b_r(\tau,\lambda)$ are, respectively, the absorption and scattering coefficients, $p_r(\cos\Theta,\lambda)$ is the phase function for scattering from $\{\mu',\varphi',\lambda'\}$ to $\{\mu,\varphi,\lambda\}$, $S_0(\tau,\lambda)$ is an internal source of radiation and, finally, $G^+(\mu,\varphi,\lambda)$ and $G^-(\mu,\varphi,\lambda)$ are the incident radiation at the boundaries.

We discretize Eqs. (1)–(4), for each region, in the wavelength variable and consider all wavelength-dependent variables as being averages over a wavelength interval (band) $\Delta\lambda_g$. Thus, for a generic variable $F(\lambda)$ we have

$$F_g = F(\lambda_g) = \frac{1}{\Delta\lambda_g}\int\limits_{\Delta\lambda_g} F(\lambda)d\lambda,$$

where λ_g is an average wavelength in the interval g. To further simplify the calculations, we consider neither downscattering nor upscattering, *i.e.*, a particle can only be scattered within the same interval. The scattering and attenuation coefficients also assume average values within each region with respect to the optical variable τ. So, we can rewrite our original RTE, expressed by Eq. (1), as

$$\mu\frac{\partial}{\partial z}L_{r,g}(\tau,\mu,\varphi) + L_{r,g}(\tau,\mu,\varphi) = \frac{b_{r,g}}{c_{r,g}}\int\limits_{-1}^{1}\int\limits_{0}^{2\pi} p_r(\cos\Theta)L_{r,g}(\tau,\mu',\varphi')d\varphi'd\mu' + S_{0,g}(\tau) \qquad (5)$$

subject to the boundary conditions

$$L_{1,g}(\tau_0, \mu, \varphi) = L_{0,g}\delta(\mu - \mu_0)\delta(\varphi - \varphi_0) + D_g^+(\mu) + \rho_s^+ L_{1,g}(\tau_0, -\mu, \varphi)$$

$$+ \frac{\rho_d^+}{\pi} \int_0^{2\pi} \int_0^1 L_{1,g}(\tau_0, -\mu', \varphi')\mu' \, d\mu' d\varphi' \qquad (6)$$

and

$$L_{R,g}(\tau_R, -\mu, \varphi) = D_g^-(\mu) + \rho_s^- L_{R,g}(\tau_R, \mu, \varphi) + \frac{\rho_d^-}{\pi} \int_0^{2\pi} \int_0^1 L_{R,g}(\tau_R, \mu', \varphi')\mu' \, d\mu' d\varphi' \qquad (7)$$

and to the interface conditions, for $r = 1, 2, \ldots, R - 1$,

$$L_{r,g}(\tau_r, \pm\mu, \varphi) = L_{r+1,g}(\tau_r, \pm\mu, \varphi), \qquad (8)$$

for $\mu \in (0, 1]$ and $\varphi \in [0, 2\pi]$. Here, $L_{0,g}$ denotes the incident beam strength, μ_0 and φ_0 are, respectively, the cosine of the polar angle and the azimuthal angle of the incident beam at $\tau = \tau_0$, and $D_g^+(\mu)$ and $D_g^-(\mu)$ represent the incident distributions of radiation at $\tau = \tau_0$, and $\tau = \tau_R$, respectively. Finally, ρ_s^+, ρ_d^+, ρ_s^- and ρ_d^- represent the specular and diffuse reflection at $\tau = \tau_0$, and $\tau = \tau_R$. We also consider that $D_g^+(\mu) = D_g^-(\mu) = \rho_s^+ = \rho_d^+ = \rho_s^- = \rho_d^- = 0$. So, we can write Eqs. (6) and (7) as

$$L_{1,g}(\tau_0, \mu, \varphi) = L_{0,g}\delta(\mu - \mu_0)\delta(\varphi - \varphi_0) \qquad \text{and} \qquad L_{R,g}(\tau_R, -\mu, \varphi) = 0,$$

with the interface conditions given by Eq. (8). In addition, the phase function $p_r(\cos\Theta)$ is represented by a finite Legendre polynomial expansion [7] given in terms of the cosine of the scattering angle Θ,

$$p_r(\cos\Theta) = \frac{1}{4\pi} \sum_{\ell=0}^{\mathcal{L}} \beta_{\ell,r} P_\ell(\cos\Theta),$$

where $\beta_{0,r} = 1$, $|\beta_{\ell,r}| < 2\ell + 1$ for $0 < \ell \leq \mathcal{L}$, $\beta_{\ell,r}$ and P_ℓ are the coefficients and the Legendre polynomials in the $\mathcal{L}^{\text{th}}$-order expansion, respectively. The $\beta_{\ell,r}$ coefficients are determined through the Henyey-Greenstein [22] phase function given by

$$p_r(\cos\Theta) = \frac{1}{4\pi} \frac{1 - \rho^2}{\sqrt{(1 + \rho^2 - 2\rho\cos\Theta)^3}},$$

where ρ, the asymmetry factor, is a parameter that can be adjusted to control the relative amounts of forward and backward scattering.

Therefore, the problem that is solved in this work is defined by Eq. (5) subject to the boundary conditions given by Eq. (6) and to the interface conditions given by Eq. (8). The scattering and attenuation coefficients will be defined in Sec. 3, as well as the adopted wavelengths, number of spatial regions, incident angles and other parameters required to solve the problem. It is important to note that we consider neither internal sources nor refraction of the particles at the surface of the water body.

To solve the problem defined above, we initially perform the Chandrasekhar decomposition [7] of the radiation field into a scattered and unscattered components. So,

$$L_{r,g}(\tau, \mu, \varphi) = L_{u,r,g}(\tau, \mu, \varphi) + L_{s,r,g}(\tau, \mu, \varphi),$$

where the unscattered component $L_{u,r,g}(\tau,\mu,\varphi)$ is the solution of Eq. (5), for the case $b_{r,g} = 0$, subject to the boundary conditions defined by Eq. (6) and interface conditions given by Eq. (8). To obtain the solution for the scattered component $L_{s,r,g}(\tau,\mu,\varphi)$, we initially perform the approximation of the phase function in a finite Legendre polynomials series. Then, we perform the Fourier cosine decomposition [7] in order to eliminate the integration over the azimuthal angle. So, we obtain $\mathcal{L} + 1$ integro-differential equations without φ dependency. The angular integral, which is present in each integro-differential equation, is approximated by the S_N method [7]. This approximation produces a set of N ordinary differential equations of first order for each azimuthal angle. The solution for this set of equations is obtained analytically by the AS_N method [4]. The AS_N method is based on the spectral decomposition of the scattering matrix. For the numerical approximation, we used the PEESNA code developed in [6] which implements the AS_N method.

To solve the problem described above, knowing the inherent optical properties (such as scattering and absorption coefficients and the phase function), the boundary conditions as well as the internal sources (if present), we determine the upwelling radiation at the surface in polar directions selected *a priori*, after the interaction of the incident light beam with the water body. The details about the analytical solution of the RTE can be found in Chandrasekhar [7], Chalhoub [4], Barichello et al. [3] and Chalhoub [5].

3. ARTIFICIAL NEURAL NETWORKS

McCulloch and Pitts [29] proposed the first mathematical model to represents the activity of biological neuron. This model, which is a simplified representation of the complicated activity of a biological neuron, preserves the nature of biological neuron, *i.e.*, it can learn and make associations. On the other hand, artificial neural networks (ANNs) are computational systems that group a set of artificial neurons and they are inspired by biological neural networks. Therefore, these systems can also learn and make associations.

The algorithms used to teach ANNs are based on a set of mathematical rules that allow the network to learn and perform associations between the input and output patterns. This is possible because during the training process, the network update its free variables (weights and biases and, in our methodology, the slope parameter). Thus, at the end of the training process we expect that the ANN method is able to produce results for patterns that were not present in the training process.

There are many types of ANNs, as well as many types of learning algorithms. One of the most common network topology is the multilayer perceptron (MLP) and the most common algorithm for training this network is the backpropagation algorithm, which is based on the delta rule or generalized delta rule. For our purposes we use the MLP network; however, we do not use the backpropagation algorithm due to the fact that it has some problems associated with its convergence [21]. So, in this work we use a learning algorithm based on quasi-Newton method [14]. Furthermore, the learning algorithms based on quasi-Newton methods have shown better results than the backpropagation algorithms, for solving problems in hydrologic optics [11, 12].

In the following paragraphs we describe briefly the learning algorithm based on the quasi-Newton method and more details can be found in Dall Cortivo et al. [11, 12] and Dall Cortivo [10]. In our methodology the learning process is considered as an optimization problem, for which we define a functional $\mathcal{J}(\cdot)$ of squared differences between each output pattern and the answers obtained by the network, and for which we seek a minimum value, i.e.,

$$\mathcal{J}(\mathbf{E}_t, \vec{X}) = \min \frac{1}{2} \sum_{p=1}^{N_p} \sum_{k=1}^{N_o} \left\| d_{kp}(\mathbf{E}_t) - y_{kp}(\mathbf{E}_t, \vec{X}) \right\|_2^2, \tag{9}$$

where $\mathbf{E}_t$ is a matrix that contains the input patterns for the training, and $\vec{X} = [\mathbf{W}\ \mathbf{B}\ \mathbf{\Gamma}]^T$ where $\mathbf{W}$, $\mathbf{B}$ and $\mathbf{\Gamma}$ are matrices that contain, respectively, the synaptic weights, biases and slope parameters.

In addition, N_p are the total patterns and N_o the total neurons in the output layer, $d_{kp}(\cdot)$ and $y_{kp}(\cdot, \vec{X})$ are, respectively, each entry of the matrices that contain the output patterns and the outputs calculated by the network. Note that in our formulation we include slope parameters in the set of the "original" variables (weights and biases).

As the training process of ANN can be considered to be nonlinear optimization problem, there is no guarantee that the global minimum will be found. Thus, an idea that seems attractive is to reduce the search space, and then search for a local minimum within it, anticipating that good results will be achieved. In addition, the reduction of the search space can prevent the saturation of neurons. We reduce the search space by applying constraints to free variables to be optimized, *i.e.*, to weights, biases and slope parameters. Thus, during the training process, the values for the entries of the matrices $\mathbf{W}$, $\mathbf{B}$ and $\mathbf{\Gamma}$ are sought, so that they must satisfy the following constraints:

$$w_{\min}^i \leq w^i \leq w_{\max}^i,$$
$$b_{\min}^i \leq b^i \leq b_{\max}^i, \tag{10}$$
$$\gamma_{\min}^i \leq \gamma^i \leq \gamma_{\max}^i.$$

In general, $w_{\{\min,\max\}} \in \mathcal{R}$ and $b_{\{\min,\max\}} \in \mathcal{R}$, however, the constraints on the γ parameter can change according to the activation functions which are used for neurons. Details about these constraints can be found in Dall Cortivo et al. [11, 12].

For the training process, the variables to be optimized are organized in a vector $\vec{X} = [\mathbf{W}\ \mathbf{B}\ \mathbf{\Gamma}]^T$, thus E04UCF subroutine [35] searches for a value $\vec{X}^*$ for $\vec{X}$, so that

$$\nabla \mathcal{J}(\cdot, \vec{X}^*) = \vec{0}, \tag{11}$$

and also considers that $\vec{X}^*$ must satisfy the constraint conditions given by Eq. (10), see NAG [35] report for details about how the calculations are performed in order to determine the $\vec{X}^*$ point.

The gradient defined in Eq. (11) is obtained from Eq. (9) and is given by

$$\nabla \mathcal{J}(\cdot, \vec{X}) = -\sum_{p=1}^{N_p} \sum_{k=1}^{N_s} \left(d_{kp}(\cdot) - y_{kp}(\cdot, \vec{X}) \right) \frac{\partial}{\partial \vec{X}} y_{kp}(\cdot, \vec{X}), \tag{12}$$

where $y_{kp}(\cdot, \vec{X})$ is defined according to the activation functions and all free variables of the network. The details about the calculations for the derivative present on the right-hand side of the above equation can be found in Dall Cortivo [10].

4. PROBLEMS

In Sec. 2 we described the mathematical model used in this work, but some specific properties about the water body and the absorption and scattering coefficients were not defined. To represent these coefficients we adopt the mathematical bio-optical models for Case 1 waters suggested in the literature. For the absorption coefficient we adopt the model suggested by Morel [32] and for the scattering coefficient we adopt the model suggested by Gordon and Morel [18]. For the profiles of the chlorophyll concentration we adopt the shifted-Gaussian model proposed by Platt et al. [39].

As the absorption and the scattering coefficients depend on the wavelength and the chlorophyll concentration, and in Sec. 2 we considered that the wavelength assumes an average value for each $\Delta\lambda_g$ interval, those coefficients also assumed average values with respect to wavelengths. So, we define the absorption and scattering models, respectively, as

$$a_{r,\lambda_g} = \left[a_{\lambda_g}^w + 0.06\, a_{\lambda_g}^c\, \mathrm{chl}_r^{0.65} \right] \left[1 + 0.2 \exp\{-0.014(\lambda_g - 440)\} \right], \tag{13}$$

and

$$b_{r,\lambda_g} = b_{\lambda_g}^w + \frac{550}{\lambda_g}\, 0.3\, \text{chl}_r^{0.62},$$ (14)

where chl_r is the average chlorophyll concentration in each spatial region, and it is obtained from the adopted model expressed as

$$\text{chl}_r = \text{chl}(\overline{\tau}_m) = \text{chl}_0 + \frac{h}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(\overline{\tau}_m - \tau_m)^2}{2\sigma^2}\right],$$ (15)

where $\overline{\tau}_m$ is the midpoint of the interval $[\tau_{r-1}, \tau_r]$, see Fig. 2.
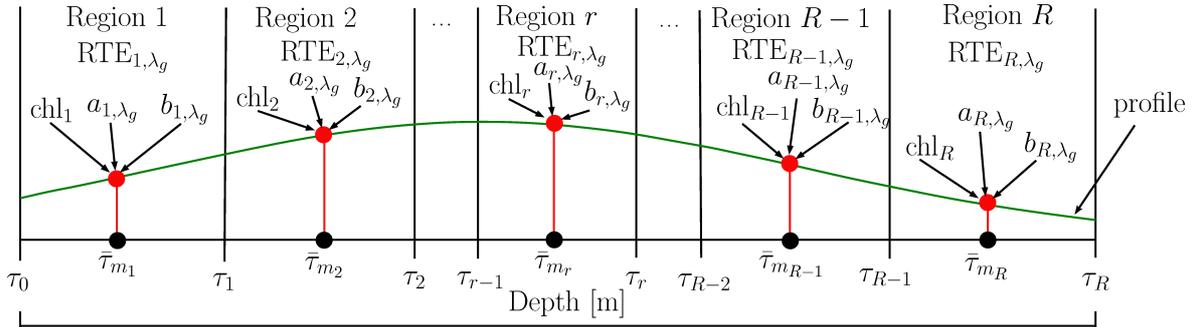

Fig. 2. Illustration of each of the average values adopted for each spatial region.

4.1. Specific details

In our simulations, we adopt two values for the h parameter. The first one is $h_{144} = 144$ mg·m^{-2} and the second one is $h_{30} = 30$ mg·m^{-2}. We adopt these values in order to consider oceanic regions with high and low chlorophyll concentrations, respectively. For the background chlorophyll concentration chl_0 we adopt two values as well. The first one is $\text{chl}_0 = 0.20$ mg · m^{-3} and is associated with h_{144} and the second one is $\text{chl}_0 = 0.01$ mg · m^{-3} which is associated with h_{30}. So, for each of those h values we define two groups of problems. The first one, labeled $\mathbf{G}_{h_{144}}$, is associated with h_{144} and the second one, labeled $\mathbf{G}_{h_{30}}$, is associated with h_{30}. The adopted value for the incident beam strength $L_0 = \pi$ cd, the polar angle $\theta_0 = \pi/3$ rad, the azimuthal angle $\varphi_0 = 0$ rad, the geometric depth of the water body $\zeta = 40$ m, the number of spatial regions $R = 20$ and the asymmetry coefficient $\rho = 0.924$ for the phase function. These values for the parameters are adapted for both groups of problems.

In the next step we define the locations of the maximum values of the chlorophyll concentration (peaks), *i.e.*, the values τ_m present in Eq. (15). These values follow a homogeneous distribution along the depth with a step size of 1 meter. So, we have 40 values equally spaced given by $\tau_{m_i} = [1, 2, \ldots, 40]_{i=1,2,\ldots,40}$. For the σ values, we adopt eight values given by $\sigma_i = [5, 6, 7, 9, 11, 13, 15, 17]_{i=1,2,\ldots,8}$, which are combined with each of the τ_{m_i} values. Therefore, as we have eight chlorophyll profiles for each τ_{m_i}, then we have a total of 320 chlorophyll profiles for each h value.

For the wavelengths we select two sets of values. The first set contains discrete values that belong to the interval $\lambda_g \in [412, 678]$ nm and these values are associated with the wavelengths of the MODIS (moderate-resolution imaging spectroradiometer) sensor in the visible interval. So, according to ocean color [36] those values are given by

$$\lambda_{g_i}^M = [412, 443, 469, 488, 531, 547, 555, 645, 667, 678],$$ (16)

with $i = 1, 2, \ldots, 10$. For the second one the values belong to the interval $\lambda_g \in [560, 610]$ nm. For this set we select only six equally spaced values with a step size of 10 nanometers. Thus,

$$\lambda_{g_i} = [560, 570, 580, 590, 600, 610], \qquad i = 1, 2, \ldots, 6. \tag{17}$$

The last values to be defined for the parameters are the polar directions, where the measurements of the upwelling radiation at the surface is calculated. For the wavelengths defined in Eq. (17), we adopt seven values that belong to the interval $\theta \in [0°, 60°]$ and they are equally spaced with a step size of 10 degrees. So, we have

$$\theta_i = [0°, 10°, 20°, 30°, 40°, 50°, 60°], \qquad i = 1, 2, \ldots, 7. \tag{18}$$

As mentioned before, the two groups of problems were labeled as $\mathbf{G}_{h_{144}}$ and $\mathbf{G}_{h_{30}}$, so for each of those groups we consider two classes of problems: the first one is labeled as $1\mathrm{D}_{(6/10)\lambda}$; and the second one is labeled as $7\mathrm{D}_{1\lambda}$. In the first class of problems, we consider three types of problems, while in the second class we consider only one type of problems. The three problems in class $1\mathrm{D}_{(6/10)\lambda}$ are: problems M_{h_*}; problems P_{1h_*}; and problems P_{2h_*}. The character $*$ used as an index means that we are referring to the values 30 or 144. Figure 3 shows each problem in its respective class for both $\mathbf{G}_{h_{30}}$ and $\mathbf{G}_{h_{144}}$ groups.
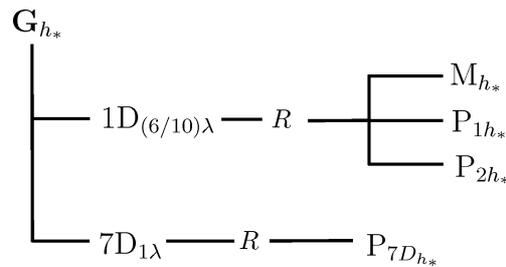

Fig. 3. Graphical representation of the problems in each class for each group.

The characteristics of each problem, for each one of the classes, for both groups are:

1. In the class of problems $1\mathrm{D}_{(6/10)\lambda}$, the measurements (calculations) of the upwelled radiation at the surface are made only in one polar direction, yet in six or ten distinct wavelengths, as follows:

   (a) in problems M_{h_*} the measurements of the upwelling radiation at the surface are made only in direction $\theta = 0°$, however at the wavelengths defined in Eq. (16);

   (b) in problems P_{1h_*} the measurements of the upwelling radiation at the surface are also made only in direction $\theta = 0°$, however at the wavelengths defined in Eq. (17);

   (c) in problems P_{2h_*} the measurements of the upwelling radiation at the surface are made only in direction $\theta = 60°$ and also at the wavelengths defined in Eq. (17);

2. In the class of problems $7\mathrm{D}_{1\lambda}$, the measurements of the upwelling radiation at the surface are made in directions defined in Eq. (18), however for only one wavelength. The selected wavelength was $\lambda_g = 560$ nm.

4.2. Sensitivity analysis

The choice of the values shown in Eq. (16) was based on the MODIS sensor wavelengths. However, the values shown in Eq. (17) were selected in order to highlight the good features associated with the problem. In general, the upwelling radiation at the surface, for Case 1 waters, is more intense in

the blue band (of course depending on the chlorophyll concentration) and at the beginning of the green band. However, the upwelling radiation presents small oscillations related to the variations of chlorophyll concentration (mainly in the green band). This means that different chlorophyll concentrations can generate very close values for the upwelling radiation at the surface [13]. On the other hand, for wavelengths in the red band, different chlorophyll concentrations can produce significant variations in the upwelling radiation at the surface. However, in that spectral band (red) the upwelling radiation at the surface is very small (*i.e.*, it has a low order of magnitude) and the high absorption of water in that band does not allow the incident radiation to reach more than a few meters [31].

Figure 4 shows the single scattering albedo for the sets of wavelengths defined in Eqs. (16) and (17). Equations (13) and (14) for absorption and scattering coefficients and a chlorophyll profile with the following parameters: $chl_0 = 0.2$ mg·m^{-3}, $h = 144$ mg·m^{-2}, $\tau_m = 17$ m and $\sigma = 9$ m were considered. The values adopted for the chlorophyll profile were obtained from Platt and Sathyendranath [40] and are suggested for the Celtic Sea in May.
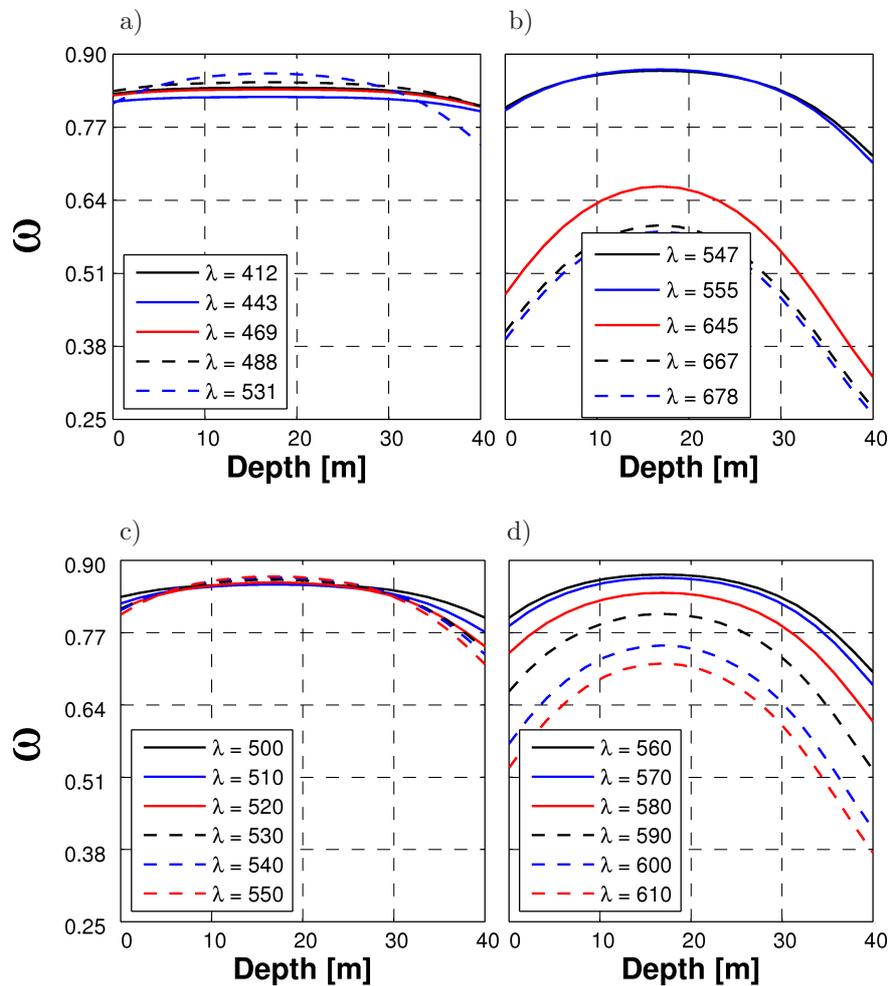
Fig. 4. Single scattering albedo curves for the Celtic Sea chlorophyll profile: a) and b) for the wavelengths of the MODIS sensor, c) and d) for the wavelengths defined in Eq. (17).

From Fig. 4a it is possible to note that the albedo curves have almost no variations along the depth and we can also observe a similar behavior in Fig. 4c. On the other hand, the most significant variations in the curves of the single scattering albedo occur in Figs. 4b and 4d. Thus, the choice of the wavelengths defined in Eq. (17) was associated with the sensitivity of the single scattering albedo for the wavelengths in the interval $\lambda \in [560, 610]$ nm.

An analysis on the selected polar directions is also performed. Figures 5a and 5b show the upwelling radiation at the surface (from simulation of the RTE) in the polar directions defined in Eq. (18) and considering the albedo curves shown in Figs. 4c and 4d. Note that when the albedo curves are close to each other the upwelling radiation curves at the surface are close to each other as well. On the other hand, Figs. 5c and 5d show the upwelling radiation curves at the surface considering different wavelengths; however, they were measured (calculated) in the directions $\theta = 0°$ and $\theta = 60°$. In these figures, we can observe that the shape of the upwelling radiation in the directions $\theta = 0°$ and $\theta = 60°$ is very similar. The difference in those figures is in the order of the radiance magnitude, and the interval, where the upwelling radiation presents the most variation, with respect to the wavelengths, is close to the interval defined in Eq. (17).
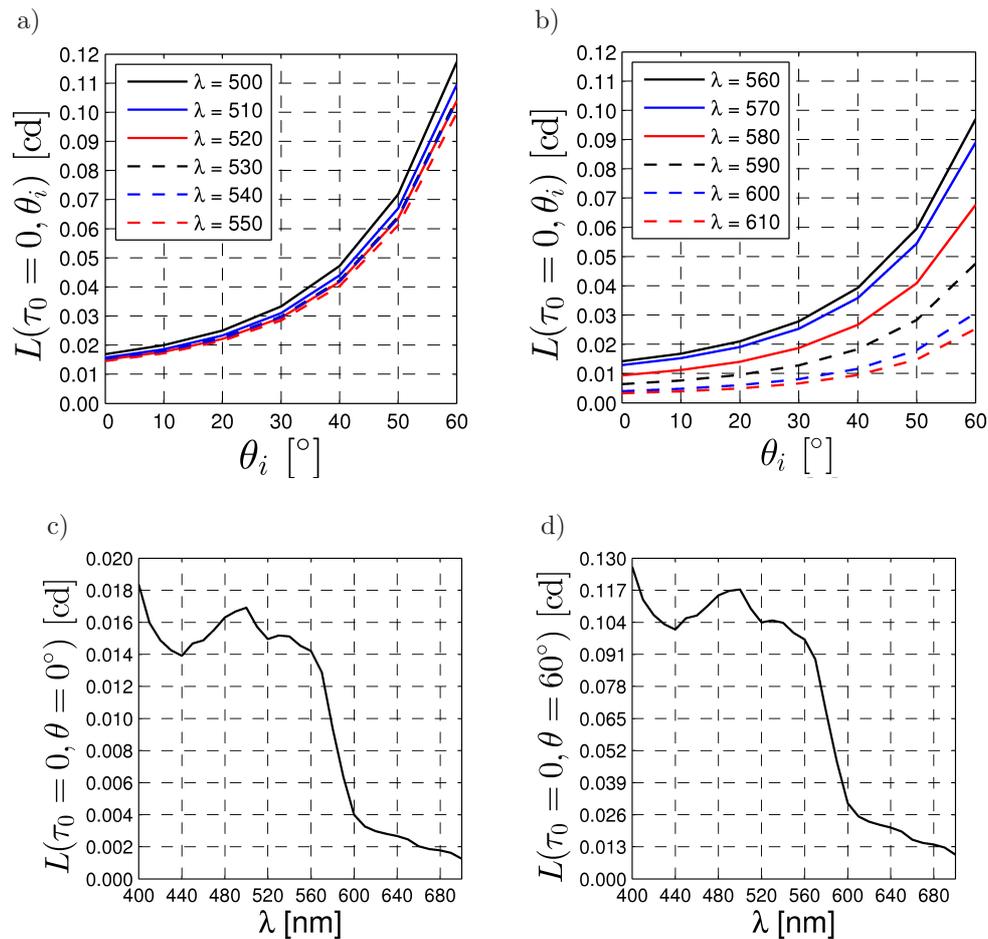
Fig. 5. Upwelled radiation at the surface considering the Celtic Sea chlorophyll profile. Figures 5a and 5b show the upwelling radiation in the polar directions defined in Eq. (18) for the wavelengths within the interval $\lambda \in [500, 550]$ nm and $\lambda \in [560, 610]$ nm, respectively. Figures 5c and 5d show the upwelling radiation at the surface in the directions $\theta = 0°$ and $\theta = 60°$, respectively.

Finally, we select the wavelength $\lambda = 560$ nm because it is the first value of the wavelengths defined in Eq. (17) and it is the value in that interval that produces the highest upwelling radiation at the surface.

4.3. Input and output patterns

The equations and the values which were defined before are the input parameters for the model explained in Sec. 2, *i.e.*, knowing those equations and the values for it parameters, we are able

to solve the RTE and then determinate the upwelling radiation $L_{1,\lambda_g}(\tau_0, \mu)$ at the surface. So, we can consider that the equations and parameters defined before (especially the standard deviation σ and the location of the peak of chlorophyll concentration ζ_m) are the "inputs", and the upwelling radiation at the surface is the "outputs". On the other hand, when we are estimating the profile of the chlorophyll concentration, the standard deviation and the location of the peak of the chlorophyll concentration become the outputs and the upwelling radiation at the surface becomes the inputs, see Fig. 6. Of course, we can reconstruct the chlorophyll profile from the standard deviation and the location of the peak of the chlorophyll concentration because all other parameters are considered as constant values.
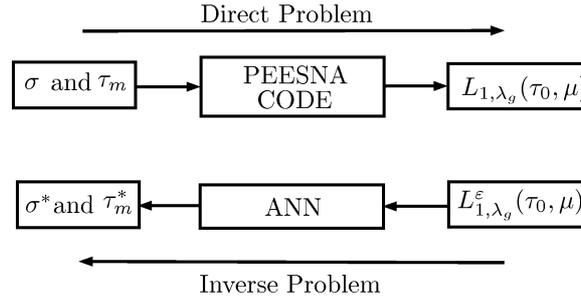

Fig. 6. Schematic illustration of a direct and an inverse problem.

In order to make possible the estimation of the standard deviation and the location of the peak of the chlorophyll profile we must consider, besides the radiation at the surface, the chlorophyll concentration at the surface, labeled as chl_0^S. Therefore, the input patterns for the networks are composed of the upwelling radiation and the chlorophyll concentration at the surface.

In order to calculate chl_0^S we consider $\overline{\tau}_m = 0$ in Eq. (15). In practical applications, we will not be able to calculate this value from the referred equation because we do not know the standard deviation and the location of the peak. However, we can get this value from the ocean color algorithms or from *in situ* measurements.

As mentioned before, three networks were used: $\mathrm{ANN_{Cte}}$, ANN_σ and ANN_{τ_m}. The network $\mathrm{ANN_{Cte}}$ performs a pre-classification of the problems and the other two solve the problem. The network $\mathrm{ANN_{Cte}}$ has, as input patterns, only the upwelling radiation at the surface and the output is the average value of the profile of chlorophyll concentration. So, from the radiance $L_{1,\lambda_g}(\tau_0, \cdot)$ the network $\mathrm{ANN_{Cte}}$ performs the estimation of the average value for the profile of chlorophyll concentration. The network ANN_σ performs the estimation of the standard deviation from the input patterns composed of $L_{1,\lambda_g}(\tau_0, \mu)$ and chl_0^S. Finally, the network ANN_{τ_m}, performs the estimation of the location of the peak from the same input patterns used in the network ANN_σ adding the standard deviation value estimated by it, see Fig. 7, where we do not present the network $\mathrm{ANN_{Cte}}$. This strategy is used in order to help the networks in their assigning the input patterns to the output patterns, and to provide better answers instead of using only one network with two outputs.
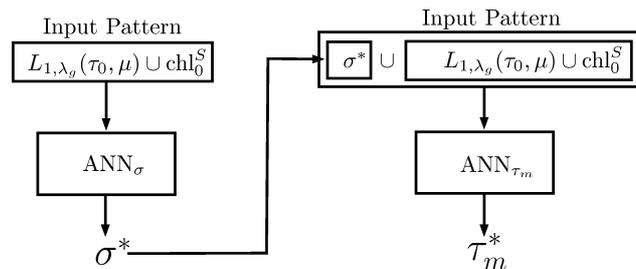

Fig. 7. Graphical representation of the input patterns for each of the networks used to solve the problem. Note that the estimation made by ANN_σ produces an additional entry in the input patterns for the second network.

Thus, the entries for each of the input patterns that compose the training set, for problems M_{h_*} are composed of chl_0^S and $L_{1,\lambda_g}(\tau_0, 0°)$ for the wavelengths defined in Eq. (16). Therefore, we have 11 entries for network ANN_σ and 12 entries for network ANN_{τ_m}. For P_{1h_*} problems, the entries of the input patterns are also composed of chl_0^S and $L_{1,\lambda_g}(\tau_0, 0°)$. However, in this case, $L_{1,\lambda_g}(\tau_0, 0°)$ is calculated using the wavelengths given in Eq. (17). For P_{1h_*} problems, ANN_σ has seven entries while ANN_{τ_m} has eight entries. The difference between problems P_{1h_*} and P_{2h_*} is that $L_{1,\lambda_g}(\tau_0, \cdot)$ is calculated in the direction $\theta = 60°$ for P_{2h_*} problems. Finally, for problems $7D_{1\lambda}$, the entries of each input pattern are composed of chl_0^S, and $L_{1,\lambda_g}(\tau_0, \mu)$ is calculated for the polar directions defined in Eq. (18) for the wavelength $\lambda_g = 560$ nm. So, for the last problems we have eight entries for ANN_σ and nine entries for ANN_{τ_m}.

4.4. Networks classifiers and network system

As mentioned before we need to classify the problems into two classes. The first one contains those problems for which it is possible to determine the profile of the chlorophyll concentration and the second one contains those problems for which it is difficult to determine the profile, but it is possible to estimate an average value for it instead. This classification is necessary due to some of the chlorophyll profiles that possess a deepest peak and a smallest standard deviations, are difficult to estimate from the input patterns defined before. So, that classification is performed by ANNs specially trained for this process, *i.e.*, the ones that identify the problems that have those features, and so facilitate the work of the other two networks.

The difficulty in the estimation of those parameters, from the input patterns, is due to the fact that the upwelling radiation at the surface, which is obtained from the chlorophyll profiles with deeper peaks, has a small order of magnitude in comparison with the upwelling radiation obtained from the chlorophyll profiles with the peaks close to the surface. Figures 8a and 8b show three profiles at 5 meters and 35 meters, and the corresponding single scattering albedo curves are shown in Figs. 8c and 8d, respectively. The upwelling radiation for these profiles is shown in Fig. 9. As we can see, these profiles that have the peak of the chlorophyll concentration close to the surface have the upwelling radiation that is more intense than the other ones.

The profiles shown in Figs. 8a and 8b are symmetric with respect to depth of 20 meters. Thus, the curves of the single scattering albedo are also symmetric. However, the attenuation that occurs in the albedo values, starting from 20 meters for the profile with $\sigma = 5$ m, and for the last 5 meters of the profile with $\sigma = 9$ m (Fig. 8c), does not influence the behavior of the upwelling radiation at the surface (see Fig. 9a). On the other hand, the attenuation that occurs in the first meters, for the albedo values shown in Fig. 8d (see the profiles obtained with $\sigma = 5$ m and $\sigma = 9$ m for $\tau_m = 35$ m), infers directly on the upwelling radiation at the surface, see for reference the black and dot-dashed lines in Fig. 9b. Therefore, the radiation that would be emitted at the surface, carrying the features/information about the distribution of the profile of the chlorophyll concentration at the bottom, is attenuated due to the reduction of the albedo value in the first half of the depth. However, the profile that considers $\sigma = 17$ m prevents the radiation to be attenuated, similarly to the other two cases.

Two or more upwelling radiation curves which are too close to each other can make difficult the correct classification by the networks, and the addition of noise to the upwelling radiation curves can make it more difficult as well. To circumvent this problem we train special networks, in order to recognize those types of problems and then "eliminate" them from the original set of problems. So, we consider another set of problems, similar to the ones described before; however, in this new set we adopt a single spatial region, where the chlorophyll concentration assumes an average value along the depth and we repeat the same features of the problems that were considered in a multiregion problem. As in this new set we have a constant chlorophyll profile (for each problem), it is necessary to define an interval of variation for these values. Therefore, for the problems of group $\mathbf{G}_{h_{144}}$ we adopt the interval $\overline{chl}_{G_{h_{144}}} \in [0.01, 2.0]$ mg $\cdot$ m^{-3}, and for the problems of group
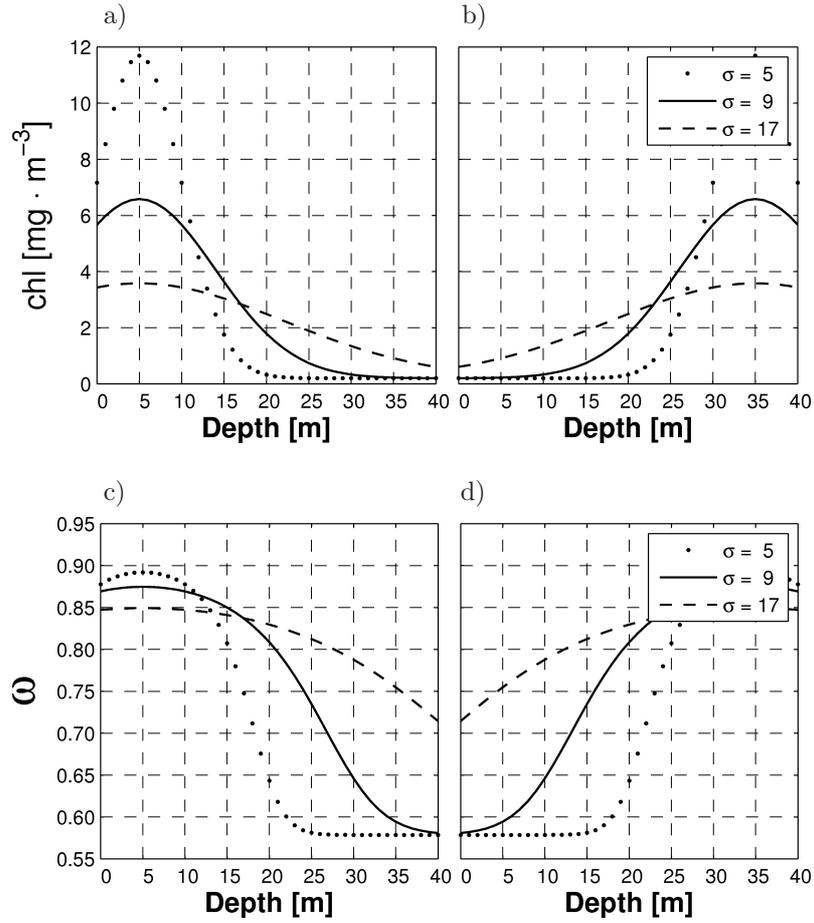
Fig. 8. Profiles of the chlorophyll concentration and its respective single scattering albedo: a) chlorophyll profiles obtained with $\tau_m = 5$ m and $\sigma = 5$ m, 9 m and 17 m, b) chlorophyll profiles obtained with $\tau_m = 35$ m and $\sigma = 5$ m, 9 m and 17 m, c) albedo curve of the profiles shown in (a), d) albedo curve of the profiles shown in (b).
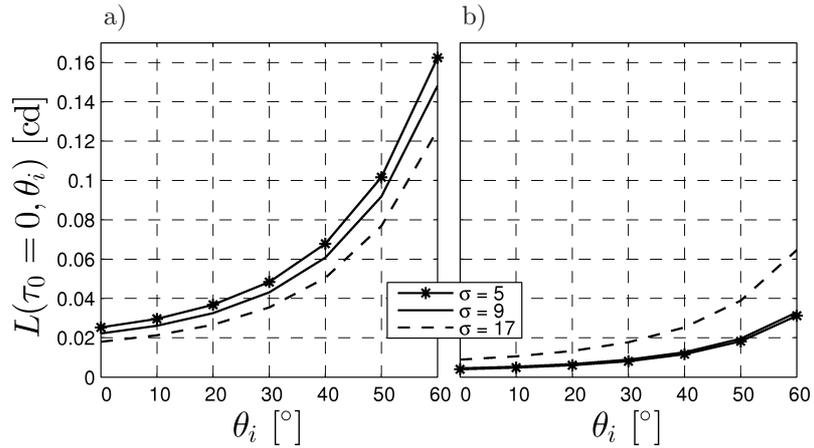
Fig. 9. Upwelling radiation at the surface measured in the polar directions defined in Eq. (18): a) considering the profiles shown in Fig. 8a, b) considering the profiles shown in Fig. 8b.

$\mathbf{G}_{h_{30}}$ we adopt the interval $\overline{\mathrm{chl}}_{\mathrm{G}_{h_{30}}} \in [0.01,\ 1.0]$ mg $\cdot$ m^{-3}. In both intervals we consider a step size given by $\Delta_{\overline{\mathrm{chl}}} = 0.01$ mg $\cdot$ m^{-3}. Thus, we have 200 discrete values for the problems in group $\mathbf{G}_{h_{144}}$, and 100 discrete values for the problems in group $\mathbf{G}_{h_{30}}$.

These networks are trained in order to determine, from the upwelling radiation at the surface, an average value for the chlorophyll concentration. The entries for each input pattern for the networks classifiers are only composed of the upwelling radiation at the surface. The classification which is performed by those networks is based on a threshold Υ value of the chlorophyll concentration. The threshold was selected in order to eliminate, from the whole set of problems, these problems which present the characteristics explained before (upwelling radiation curves very close to each other) and do not allow the correct estimation of the chlorophyll profiles by other networks.

Figure 10 shows how the system of networks works. The solution of a problem starts when an input pattern, which is obtained from those problems that consider a shifted-Gaussian profile and multiregions, is presented to the network classifier ANN_{Cte}. If the estimation given by the network classifier is less than the threshold, then we consider that the standard deviation σ^* and the peak τ_m^* of the chlorophyll concentration cannot be estimated. Thus, the solution for the problem is taken as the average value estimated by the network. On the other hand, if the estimation obtained by the network classifier is greater than the threshold, the next step consists of the estimation of the standard deviation. To perform the estimation of σ^* we use the same input pattern presented to the network classifier; in addition we also add the chlorophyll concentration at the surface corresponding to the problem as an additional entry for the input pattern. Finally, the σ^* estimated by ANN_σ is added as an additional entry for the input pattern of the network ANN_{τ_m}, which performs the estimation of the location τ_m^* of the peak of the chlorophyll concentration.
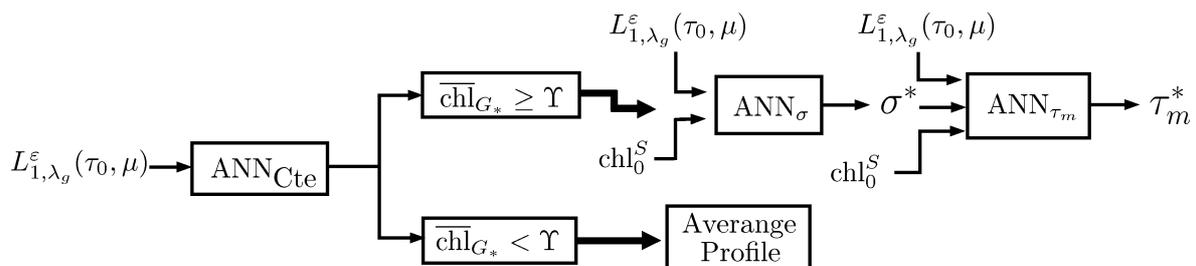
Fig. 10. Graphical representation of the joint operation of the networks classifiers and the networks that determine the value of the standard deviation and the depth of the peak of the chlorophyll concentration.

4.5. Networks training – theory

In this subsection, we present a brief description about the steps which are used to train all networks. They are based on the details given in Sec. 3. Other practical details are presented in Subsec. 5.1.

Initially, we consider that each input pattern is represented by a vector $\vec{L}_p$. Now we define a matrix E where each column is composed by each one of the input patterns. So, we have $E = [\vec{L}_1, \vec{L}_2, \ldots, \vec{L}_{N_p}]$, where N_p are the total patterns in the training set. We also consider that each output pattern is represented by a vector $\vec{d}$. So, $\vec{d} = [\nu_1, \nu_2, \ldots, \nu_{N_p}]$, where ν_p represents each output of the training set. For the outputs calculated by the networks, for each input pattern $\vec{L}_p$, we represent them as a vector $\vec{y}$, so

$$\vec{y} = \left[\nu_1^*, \nu_2^*, \ldots, \nu_{N_p}^*\right]. \tag{19}$$

It is important to note that each entry of the vector $\vec{d}$ depends on each training vector $\vec{L}_p$, and each entry of the vector $\vec{y}$ depends on each training vector $\vec{L}_p$, the synaptic weights, biases and slope parameters. So, considering a matrix W that contains all synaptic weights, a matrix/vector B that contains all biases and a matrix Γ that contains all slope parameters, we can write Eq. (19) as

$$\vec{y}\left(\vec{L}_p, W, B, \Gamma\right) = \left[\nu_1^*, \nu_2^*, \ldots, \nu_{N_p}^*\right], \qquad p = 1, 2, \ldots, N_p.$$

The optimization problem is defined as a functional given by the squared differences between the output patterns and the outputs calculated by the network for each input pattern and for which we seek a minimum value. So,

$$
\mathcal{J}(E, W, B, \Gamma) = \min \frac{1}{2} \sum_{p=1}^{N_p} \left\| d\left(\vec{L}_p\right) - y\left(\vec{L}_p, W, B, \Gamma\right) \right\|_2^2,
\tag{20}
$$

subject to the constraints $w_{\min} \leq w \leq w_{\max}$ for weights, and $b_{\min} \leq b \leq b_{\max}$ for biases. Note that we have only one sum in Eq. (20) because we have only one neuron in the output layer. The constraints to be applied to the slope parameter must be based on the type of the activation functions used for each neuron. For neurons that have the sigmoid function, as an activation function, the constraints in the slopes are given by $0 < \gamma \leq \gamma_{\max}$, and for those neurons that have the linear functions as an activation function, the constraints in the slopes are given by $\gamma_{\min} \leq \gamma \leq \gamma_{\max}$, where $\gamma \neq 0$.

All networks used in this work have an input layer, a hidden layer and an output layer. The activation functions, which are used in all neurons in the hidden layer, are the sigmoid function, given by $\phi(\gamma, v) = 1/(1 + \exp\{-\gamma v\})$, and in the output layer the linear function given by $\phi(\gamma, v) = \gamma v$, where v is the local field given by $v = \sum_{i=1}^n x_i w_i + b$ is used.

Finally, the training of the networks is performed in a batch mode, *i.e.*, the updates on the free variables of the networks are made after all training patterns are presented to the network. The search for the minimum value, for the functional defined in Eq. (20), is performed by E04UCF subroutine implemented in NAG library.

5. NETWORKS TRAINING, RESULTS AND DISCUSSION

Before explaining the training of the networks and discussing the obtained results, it is important to present some explanations considered in the description that follows. Table 1 presents: the problems considered in this work, the number of networks used in each problem, and the threshold Υ used to

Table 1. Networks used in each problem of each class, number of profiles for multi and single regions, threshold value adopted in order to classify the problems in groups $\mathbf{G}_{h_{144}}$ and $\mathbf{G}_{h_{30}}$.

Group	Classes	Regions	Problems	Networks	Profiles	Υ [mg·m^{-3}]
$\mathbf{G}_{h_{144}}$	$1\mathrm{D}_{(6/10)\lambda}$	20	$\mathrm{M}_{h_{144}}$	2		
			$\mathrm{P}_{1h_{144}}$	2	320	1.2
			$\mathrm{P}_{2h_{144}}$	2		
		1	$\mathrm{M}_{h_{144}}$	1		
			$\mathrm{P}_{1h_{144}}$	1	200	–
			$\mathrm{P}_{2h_{144}}$	1		
	$7\mathrm{D}_{1\lambda}$	20	$\mathrm{P}_{7\mathrm{D}h_{144}}$	2	320	1.2
		1	$\mathrm{P}_{7\mathrm{D}h_{144}}$	1	200	–
$\mathbf{G}_{h_{30}}$	$1\mathrm{D}_{(6/10)\lambda}$	20	$\mathrm{M}_{h_{30}}$	2		
			$\mathrm{P}_{1h_{30}}$	2	320	0.11
			$\mathrm{P}_{2h_{30}}$	2		
		1	$\mathrm{M}_{h_{30}}$	1		
			$\mathrm{P}_{1h_{30}}$	1	100	–
			$\mathrm{P}_{2h_{30}}$	1		
	$7\mathrm{D}_{1\lambda}$	20	$\mathrm{P}_{7\mathrm{D}h_{30}}$	2	320	0.11
		1	$\mathrm{P}_{7\mathrm{D}h_{30}}$	1	100	–

classify the problems into $\mathbf{G}_{h_{144}}$ and $\mathbf{G}_{h_{30}}$ groups. The following points show common information for the training of the networks and/or for each problem in each group:

1. For each problem, that considers 20 regions, two networks are trained, *i.e.*, the first one determines the σ^* value and the second one the τ_m^* value; the networks which are trained to solve the problems in $\mathbf{G}_{h_{144}}$ group are not able to solve the problems in $\mathbf{G}_{h_{30}}$ group and vice-versa.

2. Description given for the training of the networks that act in both groups is the same; so, the noise levels, number of patterns in the training and testing and validation sets are the same.

3. All networks are trained by the quasi-Newton method with gradients (for each variable) calculated by means of analytical expressions obtained from Eq. (12); the stopping criteria is based on the cross-validation process.

4. In all networks the slope parameters are considered as variables; thus, during the training process the algorithm searches for better values for them.

5. All networks use the sigmoid function as an activation function for the neurons in the hidden layer, and the linear function for the neuron in the output layer; also, all networks have only one neuron in the output layer and the number of neurons in the hidden layer is determined by *ad hoc* criteria.

6. The input patterns for the networks classifiers are not normalized, while the input patterns for the other networks are; the output patterns for all networks are not normalized.

7. The number of patterns in each set (training, testing and validation), noise levels and the quantity of patterns in each noise level are also determined by *ad hoc* criteria.

   The input patterns are normalized to the interval [0, 1] according to the following equation:

$$p_{ij}^{\mathrm{norm}} = \frac{p_{ij} - \min(P)}{\max(P) - \min(P)}, \tag{21}$$

where P is a matrix that contains the values to be normalized, p_{ij} are the entries, $\min(P)$ and $\max(P)$ are the minimum and maximum values of the matrix P, respectively, and p_{ij}^{norm} are the normalized values.

5.1. Networks training – practice

The upwelling radiation at the surface calculated by our mathematical model was corrupted with four noise levels, which are shown in Table 2. As we can see from this table, three noise levels were used for the training and testing of the networks and one of them was used only in the validation process. This strategy ensures a better generalization for the networks.

Table 2. Noise levels adopted for the training, testing and validation sets.

	ε_1	ε_2	ε_3	ε_4
Noise	1%	2%	3%	5%
Training	×	×	–	×
Testing	×	×	–	×
Validation	–	–	×	–

5.1.1. Training of the networks classifiers

During the training of the networks classifiers all patterns obtained from radiative transfer problems with a single spatial region are used. In this case, the training, testing and validation sets are equal, however the noise level in each set is different. The patterns in the training set were separated into three parts, and in each of them they were corrupted with one of the noise levels, as shown in Table 3. To perform the validation of the networks, the patterns obtained from radiative transfer problems with a single spatial region, were corrupted with the noise level ε_3, which is not present in the training and in the cross-validation process.

Table 3. Distribution of the patterns in each noise level considered in order to train the networks classifiers.

Sets	Noise Levels			
	ε_1	ε_2	ε_3	ε_4
Training	30%	30%	–	40%
Testing	100%	100%	–	100%
Validation	–	–	100%	–

The number of neurons in the input layer is defined according to the characteristics of each problem, and the number of neurons in the hidden layer is defined by *ad hoc* criteria. The threshold values adopted are $\Upsilon = 1.2$ mg·m^{-3} and $\Upsilon = 0.11$ mg·m^{-3} for groups $\mathbf{G}_{h_{144}}$ and $\mathbf{G}_{h_{30}}$, respectively. See Table 1 and, for the topology of each network, see in Table 4.

Table 4. Topology of the classifiers networks.

Group	Classes	Problems	NI	NHN
$\mathbf{G}_{h_{144}}$	$1D_{(6/10)\lambda}$	M_{h_1}	10	25
		$P_{1h_{144}}$	6	
		$P_{2h_{144}}$	6	
	$7D_{1\lambda}$	$P_{7Dh_{144}}$	7	25
$\mathbf{G}_{h_{30}}$	$1D_{(6/10)\lambda}$	$M_{h_{30}}$	10	20
		$P_{1h_{30}}$	6	
		$P_{2h_{30}}$	6	25
	$7D_{1\lambda}$	$P_{7Dh_{30}}$	7	25

NI = Number of Inputs; NHN = Number of Hidden Neurons.

When presented to the networks classifiers the patterns that were obtained from our radiative transfer problems with 20 spatial regions and corrupted with the noise level ε_3, we get that for the problem in group $\mathbf{G}_{h_{144}}$ approximately 50% of all problems were selected, and approximately 60% of all problems in group $\mathbf{G}_{h_{30}}$ were selected. It is important to remember that the networks classifiers were trained with those patterns which were obtained from the radiative transfer problems that consider only one spatial region.

5.1.2. Training of the networks ANN$_\sigma$ and ANN$_{\tau_m}$

To form the training sets to train the ANN$_\sigma$ and ANN$_{\tau_m}$, the selected patterns are divided into two parts. The first one is composed of 70% of the selected patterns and forms the training set. The second one is composed of the last 30% and forms the testing and validation sets. The patterns in the testing and validation sets are not present in the training set, and the difference between them are the noise levels.

As we have a small quantity of patterns in each set and as we are considering three noise levels, the patterns contained in each noise level may not be sufficient to represent the physical problem. To circumvent that problem we replicate the patterns in each set. For the training set, the patterns are replicated six times and for the testing and validation sets, the patterns are replicated ten times. Then, the total patterns in the training set are distributed (and corrupted) with three noise levels according to Table 3. For the testing set, all patterns are corrupted with the noise levels indicated in that table, and the patterns in the validation set are corrupted with the noise level ε_3. Remember that the noise level ε_3 is not present in the training and (cross-validation) testing processes.

Finally, to train the ANN_{τ_m}, the inputs that correspond to the standard deviation (which was estimated by ANN_σ) are not corrupted with any noise level, they are, however, normalized to the interval [0, 1] according to Eq. (21). Both ANN_σ and ANN_{τ_m} networks are trained according to the methodology described in Subsec. 4.5.

5.2. The obtained results and discussion

The main results obtained in this work are shown in Table 5. Those results are obtained by the networks from the testing and validation sets for groups $\mathbf{G}_{h_{144}}$ and $\mathbf{G}_{h_{30}}$. In order to count the hits obtained for each network we considered that a correct answer given by any network must be within the interval given by

$$\nu - 1 \leq \nu^* < \nu + 1, \tag{22}$$

where ν is the correct answer (σ or τ_m) and ν^* is the answer given by the network. We adopted this rule to count the hits because we considered, especially, in the discretization of the τ_m variable, a step size of 1 meter.

Table 5. Results obtained by the networks on the validation set for the problems in groups $\mathbf{G}_{h_{144}}$ and $\mathbf{G}_{h_{30}}$.

Group	Classes	Problem	Networks	I	HN	TPT	HT [%]	TPV	HV [%]	HS [%]	TV [s]	MSE
$\mathbf{G}_{h_{144}}$	$1\text{D}_{(6/10)\lambda}$	$\text{Aqua}_{h_{144}}$	RNA_σ	11	76	1470	62.59	490	59.59	47.55	19.65	1.89
			$\text{RNA}_{\overline{\tau}_m}$	12	111		63.95		65.10		37.19	1.60
		$\text{P}1_{h_{144}}$	RNA_σ	7	80	1470	70.41	490	70.00	55.92	15.29	1.30
			$\text{RNA}_{\overline{\tau}_m}$	8	78		75.31		72.45		12.63	0.92
		$\text{P}2_{h_{144}}$	RNA_σ	7	63	1440	56.35	480	58.96	37.29	7.18	1.94
			$\text{RNA}_{\overline{\tau}_m}$	8	96		58.13		56.46		12.09	2.30
	$7\text{D}_{1\lambda}$	$\text{P}7\text{D}_{h_{144}}$	RNA_σ	8	92	1560	62.63	520	63.08	42.69	27.24	1.93
			$\text{RNA}_{\overline{\tau}_m}$	9			66.47		63.85		27.47	1.87
$\mathbf{G}_{h_{30}}$	$1\text{D}_{(6/10)\lambda}$	$\text{Aqua}_{h_{30}}$	RNA_σ	11	59	1710	73.80	570	71.75	55.61	23.94	1.06
			$\text{RNA}_{\overline{\tau}_m}$	12			72.22		72.46		25.93	1.25
		$\text{P}1_{h_{30}}$	RNA_σ	7	59	1800	74.00	600	74.17	63.33	6.52	0.95
			$\text{RNA}_{\overline{\tau}_m}$	8			80.11		78.83		7.26	1.06
		$\text{P}2_{h_{30}}$	RNA_σ	7	48	1770	57.01	590	52.71	40.00	12.80	2.23
			$\text{RNA}_{\overline{\tau}_m}$	8	78		56.95		57.29		23.94	3.12
	$7\text{D}_{1\lambda}$	$\text{P}7\text{D}_{h_{30}}$	RNA_σ	8	78	1980	79.19	660	79.85	63.03	39.38	0.80
			$\text{RNA}_{\overline{\tau}_m}$	9	79		73.08		75.91		37.53	1.32

In Table 5, we present the groups, classes, problems and networks. We also present the number of inputs in the network (I), the number of neurons in the hidden layer (HN), the total patterns in the

testing set (TPT) and corresponding hits (HT), the total patterns in the validation set (TPV) and corresponding hits (HV), the hits obtained simultaneously (*i.e.,* both networks ought to produce a correct answer) in the validation process (HS), the required time spent to reach the optimal point of the training (TV), and the mean squared error (MSE), given by

$$\text{MSE} = \frac{1}{N_r}\frac{1}{N_P}\sum_{p=1}^{N_P}|d_p - y_p|^2,$$

where N_r is the number of noise levels, N_P are the total patterns in the validation set, d_p is the desired answer and y_p is the answer calculated by the network.

Our discussion will be about the results shown in Table 5. The hits obtained from the testing set (column HT), in the cross-validation process, indicate that the characteristics of problem $\text{P}_{1h_{144}}$ are the most adequate. This can be assumed because in that problem we obtained the greater hit rate when compared with other problems of group $\mathbf{G}_{h_{144}}$. On the other hand, for problems in group $\mathbf{G}_{h_{30}}$ we obtained hit rates very close to each other except for problem $\text{P}_{2h_{30}}$. In this problem, we can observe that the hit rates were less than 60% and a similar behavior occurred for the same problem in group $\mathbf{G}_{h_{144}}$. The low hit rate can be associated with two factors. The first one is that the upwelling radiation at the surface at $\theta = 60°$ is greater than the upwelling radiation at $\theta = 0°$. The second one (which is a consequence of the first) is that it is related to the sensitivity of the problem with noise. Although the inserted noise is proportional to the upwelling radiation, the greater order of magnitude of the upwelling radiation at $\theta = 60°$, the noise tends to become more significant in this case.

The hit rates shown in columns HT and HV do not guarantee that the estimations of σ^* and τ_m^* are correct simultaneously, according to Eq. (22), because these rates were calculated independently in the output of each network. On the other hand, the rates shown in column HS mean that both networks have produced a correct answer according to Eq. (22). As we can see, in this case we have a lower hit rate than in the other two cases (HT and HV). As before, in problems P_{2h_*} we get a lower hit rate, and the hit rates in group $\mathbf{G}_{h_{30}}$ greater than the ones obtained in group $\mathbf{G}_{h_{144}}$.

Finally, the computational time spent for the training (including the cross-validation process) is shown in the column TV. It is the time necessary to reach the best point during the training process, and as we can see it was very short. This is a consequence of good features of the proposed algorithm to train the networks, and the analytical expressions for the gradients obtained from Eq. (12).

As we have a lot of problems to analyze, in Tables 6 and 7 we show the estimations that were considered as simultaneous correct answers, *i.e.,* both networks produced a correct answer. In this case, only the best estimation of the σ^* (Table 6) and τ_m^* (Table 7) is considered and the errors e_σ and e_τ are calculated, respectively, by the expressions $e_\sigma = \sigma - \sigma^*$ and $e_\tau = \tau_m - \tau_m^*$.

Table 6. The best estimations for the standard deviation and the respective values for the peak depth for problems in groups $\mathbf{G}_{h_{144}}$ and $\mathbf{G}_{h_{30}}$.

Group	Class	Problem	σ [m]	σ^* [m]	τ_m [m]	τ_m^* [m]	e_τ [m]
$\mathbf{G}_{h_{144}}$	$1\text{D}_{(6/10)\lambda}$	$\text{M}_{h_{144}}$	7.00	7.00	10.00	9.46	0.54
		$\text{P}_{1h_{144}}$	11.00	11.00	15.00	14.50	0.50
		$\text{P}_{2h_{144}}$	9.00	8.99	6.00	5.15	0.85
	7D_λ	$\text{P}_{7\text{D}h_{144}}$	15.00	15.00	1.00	0.10	0.90
$\mathbf{G}_{h_{30}}$	$1\text{D}_{(6/10)\lambda}$	$\text{M}_{h_{30}}$	15.00	14.99	13.00	12.54	0.46
		$\text{P}_{1h_{30}}$	9.00	9.00	14.00	13.61	0.39
		$\text{P}_{2h_{30}}$	13.00	12.99	28.00	27.81	0.19
	7D_λ	$\text{P}_{7\text{D}h_{30}}$	11.00	11.00	4.00	4.26	-0.26

Table 7. Best estimations of the peak depth and the respective values for the standard deviation for problems in group $\mathbf{G}_{h_{144}}$ and $\mathbf{G}_{h_{30}}$.

Group	Class	Problem	σ [m]	σ^* [m]	e_σ [m]	τ_m [m]	τ_m^* [m]
$\mathbf{G}_{h_{144}}$	$1\mathrm{D}_{(6/10)\lambda}$	$\mathrm{M}_{h_{144}}$	6.00	6.66	-0.66	7.00	7.00
		$\mathrm{P}_{1h_{144}}$	15.00	14.24	0.76	6.00	6.00
		$\mathrm{P}_{2h_{144}}$	15.00	14.90	0.10	14.00	14.00
	$7\mathrm{D}_\lambda$	$\mathrm{P}_{7\mathrm{D}h_{144}}$	13.00	13.54	-0.54	27.00	26.99
$\mathbf{G}_{h_{30}}$	$1\mathrm{D}_{(6/10)\lambda}$	$\mathrm{M}_{h_{30}}$	6.00	6.05	-0.05	8.00	7.99
		$\mathrm{P}_{1h_{30}}$	11.00	10.25	0.75	6.00	6.00
		$\mathrm{P}_{2h_{30}}$	9.00	8.82	0.18	15.00	14.99
	$7\mathrm{D}_\lambda$	$\mathrm{P}_{7\mathrm{D}h_{30}}$	7.00	7.18	-0.18	11.00	11.00

We observe in Table 6 that a good estimation for the standard deviation does not imply a good estimation for the depth of the peak. We notice especially in problem $\mathrm{P}_{7\mathrm{D}h_{144}}$, in Table 6, that the error in the estimation of the standard deviation is null, however we get a "big" error in the estimation of the depth of the peak. A similar case has occurred in problem $\mathrm{P}_{2h_{144}}$ in the same table. Figures 11a and 11b show the profile obtained from the best estimation of the standard deviation for the cases where we get a lower error in the estimation of the peak depth. Figures 11c and 11d show the profile obtained from the best estimation of the peak depth for these cases where
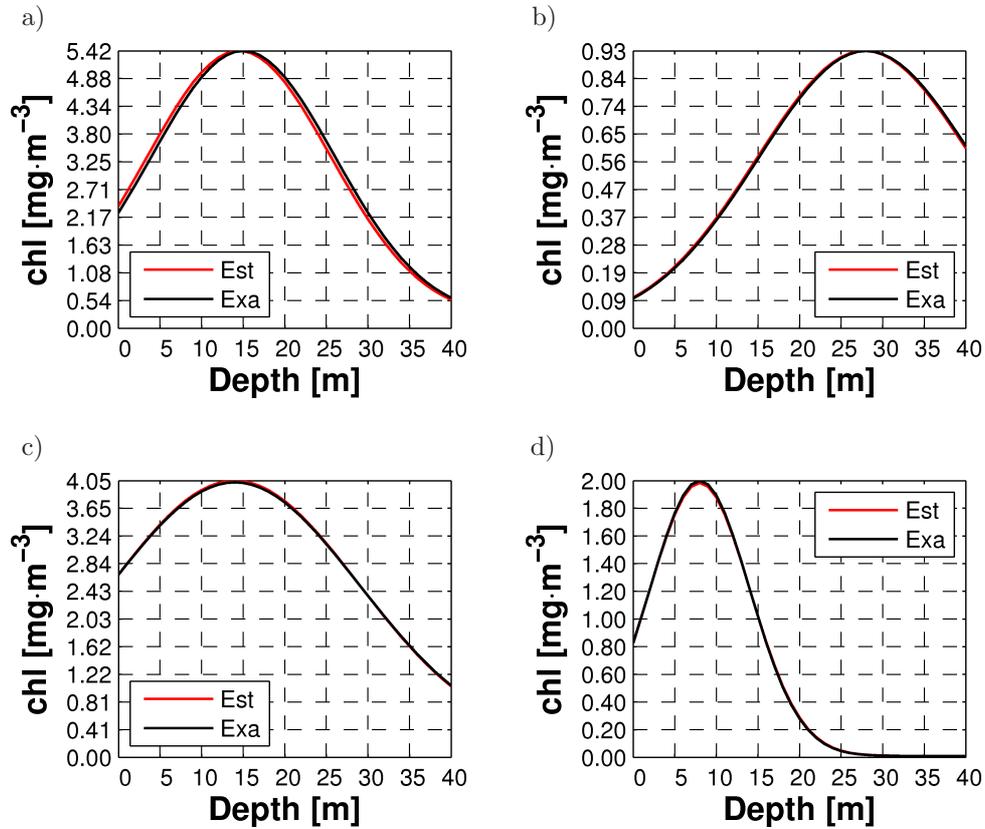

Fig. 11. a) and b) show the profiles obtained from the best estimations of the standard deviation for problems $\mathrm{P}_{1h_{144}}$ and $\mathrm{P}_{2h_{30}}$, respectively, for these cases where we get a lower error in the estimation of the peak depth which are shown in Table 6, c) and d) show the profiles obtained from the best estimations of the peak depth for problems $\mathrm{P}_{2h_{144}}$ and problem $\mathrm{M}_{h_{30}}$, respectively, for the cases where we get a lower error in the estimation of the standard deviation which are shown in Table 7.

we get a lower error in the estimation of the standard deviation, for the problems in groups $\mathbf{G}_{h_{144}}$ and $\mathbf{G}_{h_{30}}$.

On the other hand, a poor estimation of the standard deviation does not imply a poor estimation of the peak depth, as we can see in Table 7 for problems $P_{1h_{144}}$ and $P_{1h_{30}}$. Note that in these cases we get a "big" error in the estimation of the standard deviation, however the error in the estimation of the peak depth is almost insignificant.

Finally, we present the worst estimations of the standard deviation and the peak depth. These estimations are shown in Tables 8 and 9. In this case, we do not take into account the cases where the networks produced a correct or a wrong answer. We simply search for the worst answers obtained on the validation set.

Table 8. The worst estimations for the standard deviation and the respective values for the peak depth for the problems in groups $\mathbf{G}_{h_{144}}$ and $\mathbf{G}_{h_{30}}$.

Group	Class	Problem	σ [m]	σ^* [m]	e_σ [m]	τ_m [m]	τ_m^* [m]	e_τ [m]
$\mathbf{G}_{h_{144}}$	$1D_{(6/10)\lambda}$	$M_{h_{144}}$	17.00	11.38	5.62	9.00	11.79	-2.79
		$P_{1h_{144}}$	11.00	15.36	-4.36	15.00	16.25	-1.25
		$P_{2h_{144}}$	9.00	13.75	-4.75	14.00	16.87	-2.87
	$7D_\lambda$	$P_{7Dh_{144}}$	9.00	14.74	-5.74	21.00	23.51	-2.51
$\mathbf{G}_{h_{30}}$	$1D_{(6/10)\lambda}$	$M_{h_{30}}$	17.00	12.36	4.64	13.00	14.47	-1.47
		$P_{1h_{30}}$	15.00	18.70	-3.70	28.00	29.52	-1.52
		$P_{2h_{30}}$	9.00	14.11	-5.15	14.00	16.16	-2.16
	$7D_\lambda$	$P_{7Dh_{30}}$	13.00	16.32	-3.32	16.00	15.94	0.06

Table 9. The worst estimations of the peak depth and the respective values for the standard deviation for problems in groups $\mathbf{G}_{h_{144}}$ and $\mathbf{G}_{h_{30}}$.

Group	Class	Problem	σ [m]	σ^* [m]	e_σ [m]	τ_m [m]	τ_m^* [m]	e_τ [m]
$\mathbf{G}_{h_{144}}$	$1D_{(6/10)\lambda}$	$M_{h_{144}}$	13.00	16.82	-3.82	11.00	7.03	3.97
		$P_{1h_{144}}$	11.00	14.35	-3.35	20.00	24.10	-4.10
		$P_{2h_{144}}$	11.00	14.67	-3.67	21.00	26.81	-5.81
	$7D_\lambda$	$P_{7Dh_{144}}$	17.00	15.83	1.17	11.00	4.40	6.60
$\mathbf{G}_{h_{30}}$	$1D_{(6/10)\lambda}$	$M_{h_{30}}$	17.00	15.68	1.32	2.00	7.15	-5.15
		$P_{1h_{30}}$	13.00	16.47	-3.47	28.00	32.67	-4.67
		$P_{2h_{30}}$	17.00	17.82	-0.82	10.00	3.79	6.21
	$7D_\lambda$	$P_{7Dh_{30}}$	11.00	12.21	-1.21	6.00	-0.23	6.23

In that situation, we will analyze the worst estimation for the standard deviation and peak depth, which occurred in the groups $\mathbf{G}_{h_{144}}$ and $\mathbf{G}_{h_{30}}$. In Table 8 we can observe that we get the worst estimation for the standard deviation in problem $P_{7Dh_{144}}$, in group $\mathbf{G}_{h_{144}}$. In group $\mathbf{G}_{h_{30}}$ the worst estimation has occurred in problem $P_{2h_{30}}$. The respective profiles are shown in Figs. 12a and 12b. In Table 9 we can observe that we get the worst estimation for the peak depth in problem $P_{2h_{144}}$ in group $\mathbf{G}_{h_{144}}$. In group $\mathbf{G}_{h_{30}}$ the worst estimation has occurred in problem $P_{7Dh_{30}}$. The respective profiles are shown in Figs. 12c and 12d.

To conclude our analysis, it is important to mention two situations that have occurred. One of them is in Table 8 and the other one is in Table 9. The first one is for problem $P_{7Dh_{30}}$ in Table 8, where we get a "big" error in the estimation of the standard deviation, but the estimation of the peak depth was good. The second one is for problem $P_{2h_{30}}$ in Table 9, where we get a good
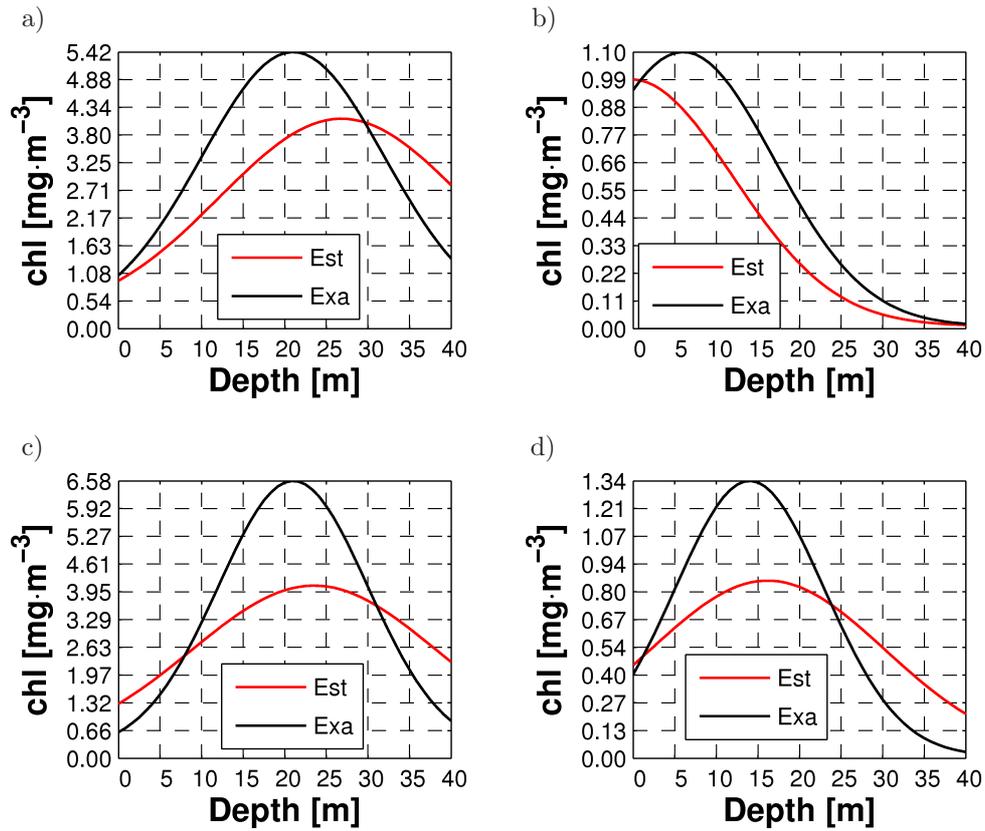
Fig. 12. Profiles obtained from: a) and b) the worst estimation of the standard deviation for problems $P_{7Dh_{144}}$ and $P_{2h_{30}}$, respectively, c) and d) the worst estimation of the peak depth for problems $P_{2h_{144}}$ and $P_{7Dh_{30}}$, respectively.

estimation for the standard deviation, however the estimation of the peak depth was poor. The profiles for those two cases are shown in Fig. 13. These situations emphasize our assertion that, although the networks are "connected", the answers given by them are independent.
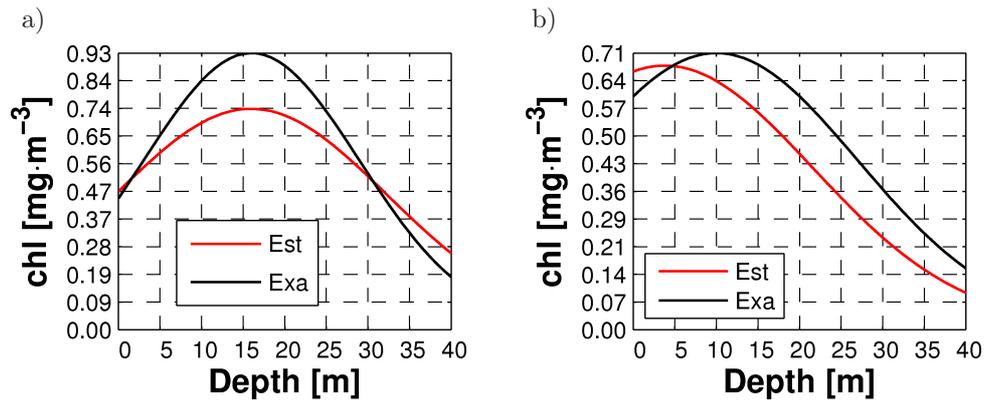
Fig. 13. Profile obtained from: a) a poor estimation of the standard deviation and a good estimation of the peak depth, b) a good estimation of the standard deviation and a poor estimation of the peak depth.

6. CONCLUSION

Although the problems studied in this paper are sensitive to noise and have physical limitations (two different chlorophyll profiles can generate the same upwelling radiation at the surface), the

obtained results show that our methodology is quite promising to estimate the concentration of the chlorophyll profile for ocean waters of Case 1. We are able to make such an assertion because we get good hit rates on the validation set and, when we get a wrong answer the committed error is "controlled", i.e., the error on the estimation remained within "acceptable limits". Furthermore, the problems P_{1h_*}, which considered measurements only in the direction $\theta = 0°$ and in the wavelengths defined in Eq. (17), showed the best results.

We can also attribute the good results obtained to the strategy used to train the networks. The inclusion of the slope parameter into the set of free-variables increases the dimension of the search space, however it becomes possible to get different shapes for the activation functions, and so neurons can become a specialist in a particular "region" of the input patterns. Also, the constraints adopted on the free-variable avoid reaching very high orders of magnitude, and we can thus prevent the saturation of neurons, and a possible overflow in the exponential term of the activation functions.

Another advantage is associated with the computational time spent to solve the problem. As we mentioned before, the training of the networks needs a low computational time, due to the efficiency of the quasi-Newton method and the analytical expressions used to approximate the gradients. This feature encourages the use of laptop computers to solve the problems presented in this work.

The cascade of networks (Fig. 10) has proved to be an efficient strategy to solve the problem. The classification performed by the first network makes easier the work of the last two networks. Also, for the problems for which it is not possible to estimate the chlorophyll profile an average profile along the depth from the network classifier is obtained. Thus, we can assert that we have answer for all problems.

The physical feature of the problem which can produce the "same" upwelling radiation at the surface for two different profiles can be circumvented by adopting the chlorophyll concentration at the surface as an input for the networks. Nevertheless, if the profile is very deep and has a small standard deviation this strategy can become inefficient because we can get the "same" upwelling radiation and the "same" chlorophyll concentration at the surface for two distinct problems. In this work, the approximation of the chlorophyll concentration at the surface was performed by an analytical expression. In practical application, we cannot obtain this value from the same equation, we can however obtain the chlorophyll concentration at the surface from the algorithms which were mentioned at the introduction of this work.

Finally, in view of the obtained results, the use of artificial neural networks to solve inverse problems in hydrological optics has shown an efficient and robust methodology. Of course, we must use an efficient algorithm to train the networks.

References

[1] J.G. Acker. *The heritage of SeaWiFS: a retrospective on the CZCS NIMBUS experiment team (NET) program.* In: NASA Tech. Memo. 104566, Vol. 21, S.B. Hooker, E.R. Firestone [Eds.], NASA Goddard Space Flight Center, Greenbelt, Maryland, pp. 44, 1994.

[2] J. Aiken, G.F. Moore, P.M. Holligan. Remote sensing of oceanic biology in relation to global climate change. *Journal of Phycology*, **28**: 579–590, 1992, doi: 101111/j00223646199200579x.

[3] L.B. Barichello, R.D.M. Garcia, C.E. Siewert. Particular solutions for the discreteordinates method. *Journal of Quantitative Spectroscopy & Radiative Transfer*, **64**: 219–226, 2000.

[4] E.S. Chalhoub, *O método das ordenadas discretas na solução da equação de transporte em geometria plana com dependência azimutal.* PhD thesis, Universidade de São Paulo (USP), São Paulo, 1997. In Portuguese with abstract in English.

[5] E.S. Chalhoub. Discrete-ordinates solution for uncoupled multi-wavelength radiative transfer problems. *Journal of Quantitative Spectroscopy & Radiative Transfer*, **92**: 335–349 2005.

[6] E.S. Chalhoub, H.F. Campos Velho, R.D.M. Garcia, M.T. Vilhena. A comparison of radiances generated by selected methods of solving the radiative-transfer equation. *Transport Theory and Statistical Physics*, **32**(5–7): 473–503, 2003.

[7] S. Chandrasekhar. *Radiative transfer*. Dover Publications, New York, 1950.

[8] E. Chassot, S. Bonhommeau, N.K. Dulvy, F. Mélin, R. Watson, D. Gascuel, O. Le Pape. Global marine primary production constrains fisheries catches. *Ecology Letters*, **13**(4): 495–505, 2010.

[9] D.K. Clark, J.W. Sherman. Nimbus – 7 coastal zone color scanner: ocean color applications. *Marine Technology Society Journal*, **20**(2): 43–56, 1986.

[10] F. Dall Cortivo. *Estimativa do perfil da concentração de clorofila em águas naturais com o uso de redes neurais artificiais*. Ph.D. thesis in Applied Computing, National Institute for Space Research (INPE), São José dos Campos, 2013. In Portuguese with abstract in English.

[11] F. Dall Cortivo, E.S. Chalhoub, H.F. Campos Velho. Comparison of two learning strategies for a supervised neural network. In: *Proceedings of 1st International Symposium on Uncertainty Quantification and Stochastic Modelling*, pp. 366–380, São Paulo, Feb. 2012. CD-ROM.

[12] F. Dall Cortivo, E.S. Chalhoub, H.F. Campos Velho. A committee of MLP with adaptive slope parameter trained by the quasi-Newton method to solve problems in hydrologic optics. In: *Proceedings of International Joint Conference on Neural Networks*, pp. 1–8, New York, Jul. 2012. IEEE Press. ISBN 978-1-4673-1409-9, doi: 101109/IJCNN20126252665.

[13] F. Dall Cortivo, E.S. Chalhoub, H.F. Campos Velho. Estimativa do perfil da concentração de clorofila em águas naturais através de um perceptron de múltiplas camadas. *Tendências em Matemática Aplicada e Computacional*, **13**(3): 233–246, 2012, doi: 105540/tema2013.013.030233. In Portuguese with abstract in English.

[14] J.E. Dennis, J.J. Moré. Quasi-Newton methods motivation and theory. *Society for Industrial and Applied Mathematics Review*, **19**(1): 46–89, 1977.

[15] R.H. Evans Howard, R. Gordon. Coastal zone color scanner "system calibration": a retrospective examination. *Journal of Geophysical Research: Oceans*, **99**(C4): 7293–7307, 1994. ISSN 2156-2202, doi: 101029/93JC02151.

[16] G.C. Feldman, N.A. Kuring, C. Ng, W.E. Esaias, C.R. McClain, J.A. Elrod, N. Maynard, D. Endres, R. Evans, J. Brown, S. Walsh, M. Carle, G. Podesta. Ocean color: availability of the global data set. *EOS*, **70**: 634–641, 1989.

[17] C.B. Field, M.J. Behrenfeld, J.T. Randerson, P. Falkowski. Primary production of the biosphere: integrating terrestrial and oceanic components. *Science*, **281**: 237–240, 1998.

[18] H.R. Gordon, A.Y. Morel. Remote assessment of ocean color for interpretation of satellite visible imagery. A review. In: *Lecture notes on coastal and estuarine studies*, Springer-Verlag, New York, 1983. ISBN 9783540909231.

[19] H.R. Gordon, D.K. Clark, J.W. Brown, O.B. Brown, R.H. Evans, W.W. Broenkow. Phytoplankton pigment concentrations in the Middle Atlantic Bight: comparisons between ship determinations and CZCS estimates. *Applied Optics*, **22**(1): 20–36, 1983.

[20] L. Gross, S. Thiria, R. Frouin, B.G. Mitchell. Artificial neural networks for modeling the transfer function between marine reflectance and phytoplankton pigment concentration. *Journal of Geophysical Research*, **105**(C2): 3483–3495, 2000.

[21] S. Haykin. *Neural networks*. Prentice Hall, 2nd edition, Jul. 1999. ISBN 0132733501.

[22] L.C. Henyey, J.L. Greenstein. Diffuse radiation in the galaxy. *Astrophysical Journal*, **93**: 70–83, 1941.

[23] IOCCG Remote sensing of ocean colour in coastal, and other optically-complex, waters. In: *Reports of the International Ocean Colour Coordinating Group*. IOCCG Report Number 3, Dartmouth Canada 2000. Retrieved from http://www.ioccg.org/reports/report3.pdf.

[24] T. Kameda, S. Matsumura. Chlorophyll biomass off Sanriku Northwestern Pacific estimated by ocean color and temperature scanner (OCTS) and vertical distribution model. *Journal of Oceanography*, **54**: 509–516, 1998.

[25] C.M. Lalli, T.R. Parsons. *Biological oceanography*. Oxford: Pergamon Press Oxford, 1993.

[26] M.R. Lewis, J.J. Cullen, T. Platt. Phytoplankton and thermal structure in the upper ocean: consequences of nonuniformity in chlorophyll profile. *Journal of Geophysical Research*, **88**(C4): 2565–2570, 1983.

[27] S. Maritorena, J.E. O'Reilly. OC2v2: Update on the initial operational SeaWiFS chlorophyll a algorithm. In: S.B. Hooker, E.R. Firestone [Eds.], O'Reilly J.E., and 24 Coauthors, 2000: *SeaWiFS postlaunch calibration and validation analyses, part 3*. NASA Tech. Memo. 2000–206892, Vol. 11, NASA Goddard Space Flight Center, Greenbelt, Maryland, 3–8, 2000.

[28] C.R. McClain. Review of major CZCS applications: U.S case studies In: V. Barale and P.M. Schlittenhardt, Eds. Ocean Colour, volume 3 of Eurocourses: Remote Sensing, pp. 167–188. Springer Netherlands, Ispra, Italy, 1993. ISBN 978-94-010-4788-3, doi: 101007/978-94-011-1791-3-7.

[29] W. McCulloch, W. Pitts. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, **5**: 115–133, 1943.

[30] B. Greg Mitchell. Coastal zone color scanner retrospective. *Journal of Geophysical Research: Oceans*, **99**(C4): 7291–7292, 1994. ISSN 2156-2202, doi: 101029/93JC03259.

[31] C.D. Mobley. *Light and water*. Academic Press, California, 1994.

[32] A.Y. Morel. Light and marine photosynthesis: a spectral model with geochemical and climatological implications. *Progress in Oceanography*, **26**(3): 263–306, 1991.

[33] A.Y. Morel, L. Prieur. Analysis of variations in ocean color. *Limnology and Oceanography*, **22**(4): 709–722, 1977.

[34] I. Murtugudde, R.J. Beauchamo, C.R. McClain, M.R. Lewis, A. Busalacchi. Effects of penetrative radiation on the upper tropical ocean circulation. *Journal of Climate*, **15**(5): 470–486, 2002.

[35] NAG. Fortran library manual. The Numerical Algorithms Group (NAG), Oxford, UK, 1995.

[36] *Project ocean color*. NASA OceanColor Web, 2013. Retrieved from http://oceancolor.gsfcnasa.gov/.

[37] J.E. O'Reilly, S. Maritorena, B.G. Mitchell, D.A. Siegel, K.L. Carder, S.A. Garver, M. Kahru, C. McClain. Ocean color chlorophyll algorithms for SeaWiFS. *Journal of Geophysical Research*, **103**(C11), 1998.

[38] J.E. O'Reilly, S. Maritorena, D.A. Siegel, M.C. O'Brien, D. Toole, B.G. Mitchell, M. Kahru, F.P. Chavez, P. Strutton, G.F. Cota, S.B. Hooker, C.R. McClain, K.L. Carder, F. Muller-Karger, L. Harding, A. Magnuson, D. Phinney, G.F. Moore, J. Aiken, K.R. Arrigo, R. Letelier, M. Culver. Nasa Technical Memorandum 2000–206892. Technical Report, Maryland, 2000.

[39] T. Platt, S. Sathyendranath, G.N. White, P. Ravindran. Attenuation of visible light by phytoplankton in a vertically structured ocean: solutions and applications. *Journal of Plankton Research*, **16**(11): 1461–1487, 1994. doi: 101093/plankt/16111461.

[40] T. Platt, S. Sathyendranath. Oceanic primary production: estimation by remote sensing at local and regional scales. *Science*, **241**(4873): 1613–1620, 1988. doi: 101126/science24148731613.

[41] A.J. Richardson, M.C. Pfaff, J.G. Field, N.F. Silulwane, F.A. Shillington. Identifying characteristic chlorophyll a profiles in the coastal domain using an artificial neural network. *Journal of Plankton Research*, **24**(12): 1289–1303, 2002. doi: 101093/plankt/24121289.

[42] D. Roemmich, J. McGowan. Climatic warming and the decline of zooplankton in California current. *Science*, **267**(5202): 1324–1326, 1995.

[43] C.L. Sabine, R.A. Feely, N. Gruber, R.M. Key, K. Lee, J.L. Bullister, R. Wanninkhof, C.S. Wong, D.W.R. Wallace, B. Tilbrook, F.J. Millero, T.H. Peng, A. Kozyr, T. Ono, A.F. Rios. The oceanic sink for anthropogenic $CO2$. *Science*, **305**(5682): 367–371, 2004.

[44] H. Schiller, R. Doerffer. Neural network for emulation of an inverse model operational derivation of Case II water properties from MERIS data. *International Journal of Remote Sensing*, **20**(9): 1735–1746, 1999. doi: 101080/014311699212443.

[45] U. Sommer, N. Aberle, K. Lengfellner, A. Lewandowska. The Baltic Sea spring phytoplankton bloom in a changing climate: an experimental approach. *Marine Biology*, **159**: 2479–2490, 2012.

[46] R.P. Souto. Recuperaçco de perfis verticais de propriedade óticas inerentes a partir da radiação emergente da água Doutorado em computação aplicada, Instituto Nacional de Pesquisas Espaciais São José dos Campos, 2006. Retrieved from http://mtc-m17.sid.inpe.br/col/sid.inpe.br/MTC-m13@80/2006/06.12.14.30/doc/publicacao.pdf (INPE14195-TDI/1097). In Portugese with abstract in English.

[47] R.P. Souto, V.C.F. Barbosa, H.F. de Campos Velho, S. Stephany. Determining chlorophyll concentration in offshore sea water from multispectral radiances by using second derivative criterion and ant colony metaheuristic. In: *Design and Optimization Symposium IPDO-2007*. Miami, USA. *Inverse Problems*, **1**: 341–348, 2007.

[48] R.P. Souto, H.F. Campos Velho, S. Stephany, M. Kampel. Chlorophyll concentration profiles from in situ radiances by ant colony optimization. *Journal of Physics: Conference Series*, **124**(1): 2008. doi:10.1088/1742-6596/124/1/012047.

[49] R.P. Souto, H.F. Campos Velho, F.F. Paes, S. Stephany, P.O.A. Navaux, A.S. Charao, J.K. Vizzotto. Grid computing for multispectral tomographic reconstruction of chlorophyll concentration in ocean water. In: C. Constanda, M.E. P'erez [Eds.], *Integral Methods in Science and Engineering*, **2**: 327–337. Birkhäuser Boston, Boston, 2010. ISBN 978-0-8176-48961, doi: 101007/978-081764897831.

[50] J. Uitz, H. Claustre, A. Morel, S.B. Hooker. Vertical distribution of phytoplankton communities in open ocean: an assessment based on surface chlorophyll. *Journal of Geophysical Research*, **111**(C8005): 1–23, 2006. doi: 101029/2005JC00327.

[51] J.A. Yoder, W.E. Esaias, G.C. Feldman, C.R. McClain. Satellite ocean colorstatus report. *Oceanography*, **1**(1): 18–20, 1988.

[52] J.A. Yoder, C.R. McClain, G.C. Feldman, W.E. Esaias. Annual cycles of phytoplankton chlorophyll concentrations in the global ocean: a satellite view. *Global Biogeochemical Cycles*, **7**(1): 181–193, 1993. ISSN 1944-9224, doi: 101029/93GB02358.