Retrieval of aerosol type and optical thickness over the Mediterranean from SeaWiFS images using an automatic neural classification method

A. Niang\textsuperscript{a,b}, F. Badran\textsuperscript{a,d}, C. Moulin\textsuperscript{c}, M. Crépon\textsuperscript{a,*}, S. Thiria\textsuperscript{a}

\textsuperscript{a} IPSL/LOCEAN, Université. Paris 6, 75252 Paris, France
\textsuperscript{b} Ecole Supérieure Polytechnique, Université Cheikh Anta Diop de Dakar, BP 5085 Dakar Fann., Senegal
\textsuperscript{c} IPSL/LSCE, 91191 Gif-sur-Yvette, France
\textsuperscript{d} Laboratoire CEDRIC, Conservatoire National des Arts et Métiers, 292 rue Saint Martin, 75003 Paris, France

Received 11 January 2005; received in revised form 4 October 2005; accepted 8 October 2005

Abstract

We present an automatic classification method based on topological neural network algorithms to retrieve aerosol optical properties from multi-spectral ocean-color satellite imagery. The first step of the method consisted in an unsupervised classification of a large set of clear-sky top of the atmosphere reflectance spectra measured by the sensor. We used the so-called Kohonen map which aggregates similar spectra into a reduced set of pertinent groups. The second step consisted in labeling these groups by clustering them with synthetic TOA reflectance spectra whose optical properties (i.e., aerosol type or optical thickness) are known. These synthetic spectra have been computed using a radiative transfer model. In the present study, we dealt with five aerosol types (maritime, coastal, tropospheric, oceanic and mineral) and several aerosol optical thickness values ranging from 0.05 to 0.8. These simulated spectra were then projected onto the Kohonen map to label each group of the map. The last step consisted in applying this method to the SeaWiFS imagery of the Mediterranean region for the years 1999 and 2000. The Kohonen map was “educated” from pixels randomly extracted during the year 1999 in this region. We accounted for the viewing geometry of the sensor by clustering the simulated spectra into ten groups of similar geometries, as defined by both scattering and sun zenith angles. The analysis of SeaWiFS images was performed pixel-by-pixel by selecting the suitable labeling (in terms of viewing geometry), then by identifying the closest spectrum in the Kohonen map, which finally gives the aerosol optical properties. This method led to accurate and coherent results, as shown by the comparison with in situ aerosol measurements provided by the AERONET station at Lampedusa and by the study of two aerosol events over the Mediterranean. One of the major advantages of this method is that it enables us to automatically identify the aerosol type and to retrieve the aerosol optical properties with a better accuracy than classical methods such as those used by SeaWifs. It gives accurate results for optical thickness values larger than 0.35 and is able to retrieve dust aerosols such as African dust aerosol (absorbing aerosol). These should ensure a more precise inversion of ocean-color imagery where the knowledge of atmospheric optical parameters is essential. Moreover the method is able to give probabilities for the estimate values of aerosol properties.

© 2005 Elsevier Inc. All rights reserved.

Keywords: Remote sensing; Aerosols; Classification; Neural networks; Mediterranean

1. Introduction

Atmospheric aerosols are small particles of various origins present in the atmosphere. They modify the radiation balance of the earth by scattering and absorbing solar and long-wave radiative transmission leading to two opposite effects: they cool the atmosphere by backscattering the solar radiation to space and they warm it by absorbing the terrestrial radiation in the lower atmosphere (Brooks & Legrand, 2000; Lelieveld et al., 2002). The balance of these two effects is still controversial (IPCC, 2001). The largest radiative contribution comes from aerosols with radii in the range 0.1–1 µm. Aerosols have different scattering and absorption properties depending on their origin, and it is important to identify them in order to better quantify their radiative impact. Among the most important aerosol types, which affect the radiative budget of the Earth we can mention sea salt, sulfates, soot and mineral dust. Aerosols have other environmental impacts also; they
serve as cloud condensation nuclei and they fertilize the ocean enhancing phytoplankton blooms (Jickells et al., 2005; Ridame & Gieux, 2002).

Aerosols can be measured in the solar spectrum from ground stations or from space using passive sensors. One of the most natural remote sensing techniques is to use multispectral radiometers dedicated to ocean-color remote sensing (i.e., the measurement of a recognized proxy for phytoplankton content in surface waters through the estimate of the Chlorophyll-a concentration, Chl-a). In fact, aerosol and phytoplankton retrievals are linked. The light scattered within the atmosphere by molecules and aerosols contributes to the Top-Of-Atmosphere (TOA) reflectance more than 90% in the blue-green part of the visible spectrum, where the spectral signature of phytoplankton is observable. The critical issue for ocean color is therefore the atmospheric correction. Such a procedure consists in estimating the aerosol contribution with a good accuracy in order to remove it from the measured TOA signals to get the actual spectrum of marine reflectance. The impact of aerosols on the reflectance spectrum depends firstly on the particle concentration in the atmosphere, but also on their specific optical characteristics of light-scattering and absorption, which are controlled primarily by the particle size distribution and complex refractive index.

In the present study we investigated the possibility of retrieving these aerosol optical properties by using an advanced mathematical method that allows an immediate identification and inversion of the TOA reflectance spectra measured by a multi-spectral sensor such as SeaWiFS. Several techniques have been proposed to retrieve the aerosol characteristics from such measurements. A first category has been widely used for atmospheric correction and is based only on the red and near-infrared measurements to retrieve the aerosol optical thickness and the Ångström coefficient, which represent the particle concentration and the particle size distribution, respectively (e.g., Gordon & Wang, 1994; Jamet et al., 2004; Jamet et al., 2005). The major drawback of such techniques is that they are not capable of discriminating between absorbing and non-absorbing aerosols (Gordon, 1997). Another category consists of algorithms developed for the study of a single aerosol type, e.g., mineral dust (Moulin et al., 2001a) using both visible and near-infrared measurements to quantify the absorption effect. The main limitation of this approach is that the aerosol type has to be known a-priori, which is usually not the case. Some recent methods have been proposed to detect the aerosol type, for example to discriminate between absorbing mineral dust and non-absorbing aerosols (Nobileau & Antoine, 2005) by taking into account one wavelength in the visible (510 nm) and wavelengths in the NIR. A classical “atmospheric correction” technique can then be used considering the suitable aerosol type.

We developed an approach that makes use of the full spectrum of measurements to perform the aerosol identification. This method aims at increasing the accuracy and the flexibility of the previous methods, as well as processing satellite imagery at a higher speed. To achieve this, we used a specific Topological Neural network Algorithm (TNA), the so-called PRobabilistic Self Organizing Map (PR SOM) (see Section 3) to classify TOA reflectance spectra. TNAs are well adapted for this task and have been used with success by Ainsworth and Jones (1999) and Niang et al. (2003) for classifying TOA reflectances. The present work continues the latter study by providing a more refined labeling procedure where the expertise is provided by a synthetic database comprising large sets of radiative-transfer computations. The development of the method is carried out in two steps. The first step, which is an unsupervised procedure, classifies the measured TOA spectra from their statistical properties. In the second step, the classes found in the first step are labeled, i.e., are assigned to aerosol optical parameters. In order this to be accomplished, the knowledge contained in the theoretical equations in which all parameters are known is compared with the observations to determine the aerosol type, the optical thickness and the Ångström coefficient. Moreover, configurations where the observed spectrum is not always due to a unique aerosol type, but represents a mixture of different particles can be handled by the proposed methodology, since it computes the probability of a spectrum to belong to each type of aerosols.

In the experiments presented here, we considered five aerosol types: the four non-absorbing aerosols (coastal, maritime, tropospheric, oceanic) used operationally to process SeaWiFS data and an African dust aerosol (absorbing aerosol). It is important to note that the SeaWiFS atmospheric correction algorithm fails when African dust aerosols predominate (Moulin et al., 2001b).

The paper is arranged as follows: Section 2 describes the data set; the methodology is presented in Section 3 (a brief description of the classification method is given in Section 3.1; Section 3.2, which represents the original part of this paper, deals with the technique used to mix theory and observation); analysis and validation of the results for SeaWiFS images are presented in Section 4; and discussion and conclusion are given in Section 5.

2. The data sets

2.1. The SeaWiFS data set

We used SeaWiFS images of the Mediterranean Sea to develop and validate our classification technique. This region is of interest for aerosol studies (e.g., Jamet et al., 2004; Sciare et al., 2003) because it is bounded on its northern side by industrial countries, which generate large amounts of sulfate aerosols, and on its southern side by the Sahara desert, which is the largest source of mineral dust in the world. Besides, the Mediterranean Sea is a source of maritime aerosols containing a mixture of sulfate particles of biological origin and sea-salt particles (Schwindling & Deschamps, 1998). Subsequently we analyzed several aerosol events occurring in 1999 and 2000.

The SeaWiFS sensor on board the SeaStar satellite is an optical radiometer used to observe color variations of the ocean
in eight spectral bands in the visible and near-infrared (412 nm, 443 nm, 490 nm, 510 nm, 555 nm, 670 nm, 765 nm and 865 nm). SeaWiFS level-1 GAC data available from the NASA/GSFC/DAAC web-site (http://daac.gsfc.nasa.gov/data/dataset/SEAWiFS/) consist of raw radiances, \( L \), which are measured at the TOA and whose dimension is expressed in W m\(^{-2}\) sr\(^{-1}\) nm\(^{-1}\). For each wavelength \( \lambda \), the TOA reflectance \( \rho \) is computed as:

\[
\rho(\lambda, \theta_s, \Phi) = \frac{\pi L(\lambda, \theta_s, \Phi)}{E_0(\lambda) \cos(\theta_s)}
\]

where \( E_0(\lambda) \) is the extraterrestrial solar irradiance (in W m\(^{-2}\) nm\(^{-1}\)), varying with the Sun–Earth distance), \( \theta_s, \theta_v \) are the sun and satellite viewing zenith angles, respectively, and \( \Phi \) is the azimuth angle.

According to Gordon (1997) the TOA reflectance \( \rho \) is the sum of several components that can be treated separately: the Rayleigh multiple scattering (air molecules) in the absence of aerosols, can be accurately computed by using the atmospheric pressure, and the whitecap contribution by taking into account the wind speed. We removed pixels contaminated by the sun glister, using a geometrical mask. The signal that was finally used in our classification method was therefore:

\[
\rho_{\text{used}} = \rho_{\text{air}} + \rho_{\text{a}} + \rho_{\text{w}} + \rho_{\text{ra}} + \rho_{\text{mt}}
\]

where \( \rho_{\text{a}} \) is the reflectance resulting from multiple scattering of aerosols in the absence of the air, \( \rho_{\text{ra}} \) is the interaction term between molecular and aerosol scattering, \( \rho_{\text{mt}} \) is the contribution of the water and \( t \) is the transmittance of the atmosphere at a given wavelength (\( \lambda \)).

In Eq. (2), \( \rho_{\text{ra}} \) is small in the red and near-infrared, so that \( \rho_{\text{used}} \) mainly depends on the aerosol term \( \rho_{\text{a}} + \rho_{\text{ra}} \) at 670, 765 and 865 nm. For the other bands in the visible, it is expected that the aerosol term remains large enough in most situations to allow us to retrieve pertinent information (in particular absorption capability) about the particles at these wavelengths. This hypothesis is certainly wrong for aerosol optical thickness values below 0.1, because the marine contribution prevails in \( \rho_{\text{used}} \).

We used two distinct satellite data sets comprising eight dimensional vectors to develop our method. Each vector, whose components correspond to the SeaWiFS wavelengths, represents a \( \rho_{\text{used}} \) spectrum.

1. The first data set consists of observed \( \rho_{\text{obs}}^{\text{used}} \) extracted from pixels of SeaWiFS images of the Mediterranean basin during the year 1999. All the available daily SeaWiFS images were homogeneously sampled (one pixel-line over ten) providing 2,346,147 clear-sky spectra of \( \rho_{\text{obs}}^{\text{used}} \) which constitute Data\(^\text{obs}\). For each spectrum we also have the geometry of the measurement (i.e., the sun and satellite viewing zenith angles \( (\theta_s, \theta_v) \), and the difference in azimuth angle \( (\Delta \Phi) \) between the Sun and the Satellite). Data\(^\text{obs}\) is used below to extract the most pertinent information embedded in the SeaWiFS observations. Moreover full SeaWiFS images taken over the Mediterranean basin during the years 1999 and 2000 will be used for estimating the accuracy of the neural approach and for comparison with the SeaWiFS product.

2. The second data set, Data\(^\text{expert}\), consists of the \( \rho_{\text{expert}}^{\text{used}} \) computed with a 2-layer radiative transfer model (Gordon & Wang, 1994) for various optical thickness values and Ångström exponent values, chlorophyll content and geometry of the measurement. Since the present study was devoted to the Mediterranean, the geometrical parameters were restricted to those of this basin. Ten aerosol optical thickness values between 0.05 and 0.8 were considered in simulations, and 13 Ångström exponent values between –0.1 and 1.5. Variations in Ångström exponent values were obtained by considering the twelve non-absorbing aerosol models defined by Gordon and Wang (1994) and which were based on the four aerosol types (i.e., costal, maritime, tropospheric, oceanic) of Shettle and Fenn (1979). An additional absorbing model was considered to account for African dust (Moulin et al., 2001a). Data\(^\text{expert}\), which comprises 9,278,362 simulated spectra, was used in order to introduce the expertise and to retrieve the aerosol types and the optical thickness values.

2.2. The AERONET data set

Direct measurements of the optical thickness at 865 nm \( \tau(865) \) constitute our third data set (Data\(^\text{photometer}\)) used for validation. We chose the measurements performed at the Lampedusa (35°31N, 12°37E) station in 2000. Lampedusa Island is in the Strait of Sicily in the Mediterranean Sea and is well suited to the measurement of atmospheric parameters. These measurements were made in the framework of the AERONET program (Aerosol Robotic NETwork), which is a federated international network of sun/sky radiometers (Holben et al., 1998). Data\(^\text{photometer}\) comprises 46 level-2.0 sun photometer measurements (cloud-screened and quality-assured), collocated with the satellite measurements \( \rho_{\text{used}}^{\text{obs}} \). A validation can thus be made by comparing the optical thickness values retrieved by the neural network algorithm with (a) those retrieved by the SeaWiFS standard methods and (b) those measured at the coastal station by the photometer. The optical thicknesses \( \tau(865) \) retrieved by the satellite were computed over the closest 3 x 3 pixel marine area from Lampedusa, but at a distance of at least 5 km from the coast, so as to avoid any turbid-water contamination. Among these 46 collocated measurements, only 34 ones could be processed by the SeaWiFS standard methods because of a restrictive cloud screening whereas all the 46 \( \tau(865) \) measurements were processed by the neural network algorithm and used for the validation. The AERONET optical thickness values used for this validation were actually the mean of all measurements made between 10:00 UT and 13:00 UT, because SeaWiFS images over Lampedusa were taken around 11:30 UT.

3. The methodology

As mentioned in the introduction, we used a two-step method. The first step, which was non-supervised, classified
the TOA spectra according to their statistical properties. In the second step, the classes found in the first step were labeled; i.e., were assigned to physical parameters of aerosols.

3.1. First step: the topological network clustering

This first step was an unsupervised neural classification similar to that done by Niang et al. (2003) who broadly described the methodology. We used the PRSOM (PRobabilistic Self Organizing Map), which is an extension of the well-known SOM model (Self Organizing Map) described by Kohonen (2001) for visualizing and clustering a high-dimensional data set. The aim is to summarize the information contained in Data_{obs} by decomposing it into a number of groups based on similar data according to some statistical criterion. In the present work, each group was characterized by a typical spectrum, the so-called reference vector (rv-R). The set of (rv-R) constitutes a compression of the information embedded in the data set. The pertinence of the method depends on the number of groups (we must have a sufficient number of groups to represent the complexity of the data set adequately, but not too large, so as to facilitate its handling) and on the method of clustering into groups. In the light of the results obtained by Niang et al. (2003), we chose the same architecture for the PRSOM map: a two-dimensional array with a large number of neurons (20 \times 20 = 400), providing a highly discriminative representation of Data_{obs}. Each neuron of the map was associated with a particular reference vector (rv-R) and thus corresponded to a group of data (in this paper, a group of \rho_{\text{obs}} belonging to Data_{obs}). The association was made during a learning phase by processing Data_{obs} according to the methodology described by Niang et al. (2003). The different neurons of the topological map C were connected together; they determined a topological (neighborhood) relationship between the different groups (neurons). The 2,346,147 \rho_{\text{obs}} (or pixels) of Data_{obs} were thus clustered into 400 groups. We denoted this topological map PRSOM-R (classification of the Reflectance). The significance of the clustering provided by the map is shown in Fig. 1, for which, four different reflectance ratios, \rho_{443}/\rho_{555}, \rho_{490}/\rho_{555}, \rho_{510}/\rho_{555}, \rho_{765}/\rho_{865}, were computed from the set of (rv-R). Each small square of a map in Fig. 1 represents a ratio computed from the referent of the related neuron. Two reference vectors which are close in the two-dimensional topological map have similar ratios, thanks to the neighborhood procedure. The visualization of the topological order showed that the four ratios provide significant information which can be used for ocean color and for atmospheric correction.

These results showed that the classification into 400 groups presents a physical consistency and allowed us to proceed to the second step of the classification, which is a supervised one. This second step consists in determining classes of reference
vectors in order to decode the SeaWiFS images and to produce maps of aerosol optical thickness and type.

3.2. The labeling procedure

The second data set, Data$_{expert}$, represented the expertise, which was used to decode the SeaWiFS images. The principle of the method is to compare the spectra of Data$_{expert}$ with those of the neurons of PRSOM-R.

3.2.1. Accounting for the viewing geometry

Since the reflectance spectra depend on the geometry of the measurement, we decided to take it into account through the sun zenith angle $\theta_s$ and the scattering angle $\gamma$ defined as:

$$\gamma = \arccos(-\cos(\theta_s)\cos(\phi) + \sin(\theta_s)\sin(\phi)\cos(\Delta \Phi))$$

where $\Delta \Phi = \phi_o - \phi_v$ is the azimuth angle difference between the satellite and the sun.

In the present study we took into account the geometrical effect by decomposing the set ($\theta_s, \gamma$) associated with Data$_{obs}$ into a reduced number of bins (or classes). Since we focused our aerosol typology on the Mediterranean basin, the scattering angle was restricted to the range 110–180° (backscattering) and the sun zenith angle, to the range 7–74° accordingly.

Within this parameter range, it has been found that the geometry set ($\theta_s, \gamma$) can be clustered in ten bins only, denoted bin-$i$ ($i=1, \ldots, 10$) for satisfying the statistical and physical criteria of the problem. This clustering was done statistically, in order to give the same importance to each bin:

1. First we used a dedicated PRSOM map, the so-called PRSOM-Angles, which was calibrated on Data$_{obs}$. PRSOM-Angles is a two-dimensional array with 100 (10*10) neurons approximating the density probability of the geometry set ($\theta_s, \gamma$) with a good accuracy. These 10*10 groups were then “binned” by applying a hierarchical clustering to the set of referent (rv-Angles), which determined ten different bin-$i$ corresponding to the different geometries. These ten bins are shown in Fig. 2. They were determined once but they will also be used in the image processing. The binning is nonlinear; the most frequent geometries were thus accurately binned, whereas the rare ones were clustered into wider areas. As the reflectance varies smoothly with the geometry, taking in account only ten bins allowed us to obtain a more robust determination of the aerosol type. Thanks to its statistical formulation, determining the geometrical characteristics with PRSOM-Angles was much easier and safer than by using a classical method.

2. We then assigned a definite geometry bin-$i$ to each $\rho_{used}^{expert}$ of the Data$_{expert}$. In the following, the set of $\rho_{used}^{expert}$ assigned to the geometry bin-$i$ was denoted Data$_{i}^{expert}$. Knowing the aerosol type and optical thickness associated with each theoretical spectrum $\rho_{used}^{expert}$, we used Data$_{i}^{expert}$ to label PRSOM-R.

3.2.2. Processing Data$_{expert}$ and Neuron labeling

Each spectrum of Data$_{expert}$ was presented to PRSOM-R which assigned it to a specific neuron corresponding to the nearest (rv-R). After the assignment, Data$_{expert}$ was dispatched among the 400 neurons of PRSOM-R. The PRSOM-R map, which included the spectra of Data$_{expert}$ corresponding to a given geometry bin-$i$, was denoted PRSOM-R$_{i}$. Each neuron of PRSOM-R$_{i}$ gathered the $\rho_{used}^{expert}$ that are the most similar to it. A certain discrepancy is possible between the referent vector of a neuron of PRSOM-R$_{i}$ and the $\rho_{used}^{expert}$ captured by

![Fig. 2. Representation of the ten geometry bins, with respect to the scattering angle $\gamma$ (x-axis) and the zenith angle $\theta_s$ (y-axis) determined by the PRSOM-Angles map and a hierarchical clustering of the set of referent vectors (rv-Angles); each grey level corresponds to a particular bin.](image-url)
this neuron, since some measured TOA reflectances are poorly represented by the aerosol models. As the PRSOM algorithm is a probabilistic one, each neuron computed its own variance during the training, which estimates the variability of its corresponding set $\rho_{\text{obs}}^{\text{data}}$. We selected the $\rho_{\text{data}}^{\text{expert}}$ spectra that are within a range of two standard deviations from their corresponding reference vector ($\mathbf{rv-R}_i$), as shown in Fig. 3. So the number of $\rho_{\text{data}}^{\text{expert}}$ embedded in PRSOM-R was somewhat reduced in order to be representative of what the radiometer has really observed. At the end

Fig. 3. Labeling procedure for a neuron associated with a given geometry bin-$i$. The thick central black line represents the mean spectrum of the neuron 62 as determined during the learning phase; the upper and lower thick black lines correspond to spectral values at a distance of two standard deviations from the central line. Only the $\rho_{\text{data}}^{\text{expert}}$ which are included in this two standard deviation region are selected for the labeling; they correspond to the light-gray spectra which are used further to determine the aerosol type and the optical thickness. The dark-gray spectra are rejected as being physically too far from the observation.

Fig. 4. Representation of the referent spectrum (circles) of neuron 66, for a given geometry bin-$i$ (PRSOM-R$_i$): the mean of the captured $\rho_{\text{data}}^{\text{expert}}$ are represented by “plus” for the maritime type ($\rho_{\text{data}}^{\text{expert}}_{\text{maritime}}$), “stars” for the coastal ($\rho_{\text{data}}^{\text{expert}}_{\text{coastal}}$) and “pentagrams” for the oceanic type ($\rho_{\text{data}}^{\text{expert}}_{\text{oceanic}}$).
of the process, we obtained a PRSOM-R \textsubscript{i} map that can be labeled by the physical parameter, associated with the neurons $\rho_{\text{used}}^\text{expert}$, and corresponding to a specific geometry bin-$i$. The procedure was run ten times, one for each bin-$i$ geometry, generating ten distinct PRSOM-R \textsubscript{i} maps. The aerosol type associated with a given neuron might thus be different from one geometry to another.

At the end of the labeling of the ten PRSOM-R \textsubscript{i} maps, each neuron of a map was associated with a set of $\rho_{\text{used}}^\text{expert}$. The neurons for which more than 95% of the captured $\rho_{\text{used}}^\text{expert}$ belonged to the same aerosol type were denoted “pure neuron”. In that case, the neuron simply takes the label of the aerosol type. Note that neurons associated with very low optical thickness values (<0.1) are also classified as “pure neurons” and are labeled “maritime aerosol”. For such low optical thickness values, the aerosol contribution is too small relative to the marine contribution to enable an accurate determination of the aerosol type. The optical thickness value for these “pure neurons” was computed either as the average of the optical thickness value of the spectra belonging to the dominant aerosol type or as the average of the optical thickness values of all the spectra corresponding to a small optical thickness value.

The neurons of PRSOM-R \textsubscript{i} that have not captured any spectrum are called “white neurons”. These “white neurons” (about 15–20% of neurons, depending on the geometry) likely correspond to aerosol types that are not accounted for within the set of theoretical aerosol models used here. But the most problematic neurons are those that have captured spectra corresponding to several aerosol types. They are called “mixed neurons”. They need a specific treatment since we could not choose a specific labeled spectrum type as representative of the neuron from simple physical arguments. For these “mixed neurons”, we computed the mean spectrum $\bar{\rho}_{\text{type}}^\text{expert}$ associated with each aerosol type, as illustrated for a specific neuron in Fig. 4, and the related mean optical thickness. Its definitive type is determined during the processing of a given satellite image (see Appendix A) by taking into account a neighborhood relationship in this image. The aerosol type associated with a given “mixed neuron” might therefore be different from one image to another.

![Fig. 5. Decoding of the SeaWiFS images for the August 6–8, 1999; typology map (right) and optical thickness (left) for August 6 (top), August 7 (middle) and August 8 (bottom).](image-url)
4. Results

We analyzed two different aerosol events by using the PRSOM-R maps. The first event was observed on August 6–8, 1999, the second on June 28–30, 2000. These analyses showed the capability of the PRSOM maps to analyze events that have not been learned and can be considered as a generalization of the PRSOM map processing. The PRSOM-R maps were trained on Data\textsuperscript{obs}, which contained only a few data from the August 1999 event and none at all from the June 2000 event.

4.1. Decoding a satellite image

The satellite image was first processed by PRSOM-Angles in order to assign a specific viewing geometry, bin-\textit{i}, to each pixel. The image was then processed with the ten PRSOM-R. Each bin-\textit{i} pixel of the image was assigned to its nearest neuron of the PRSOM-R map. Pixels associated with “pure neurons” were assigned their aerosol type and optical thickness directly. On the other hand, pixels associated with “white neurons” were of an undetermined type and their optical thickness was computed at the end of the procedure by averaging that of the neighboring neurons. As noted earlier, the “mixed neurons” were ambiguous, since they included several aerosol types. To remove this ambiguity, we associated five registers (one for each aerosol type) with each “mixed neuron” of each PRSOM-R. These registers were initialized to zero. Each pixel assigned to a “mixed neuron” selects the nearest mean spectrum \( \bar{p}_{\text{type}}^{\text{expert}} \) (among the five used herein) by minimizing the Euclidian distance between the pixel spectrum and the “mixed neuron” spectra at the three wavelengths: a visible one (510 nm) as in Nobileau and Antoine (2005) and two NIR ones (670 nm, 865 nm) with respect to which the sea is black and therefore does not influence the TOA measurements. The register of the selected type was then increased by one. At the end of this first image processing, each “mixed neuron” took the type of the most frequently selected aerosol type (i.e., the register having the highest count) and its associated optical thickness. This procedure takes into account the spatial context of the image and the labeling of the “mixed neurons” may differ from one image to another. We obtained a set of ten PRSOM-R maps for the processing of the image, each being denoted “Image-PRSOM-R\textsuperscript{-}\textit{i}.” The typology and optical thickness of the pixels of the image were easily determined by processing a second time the full image with the ten “Image-PRSOM-R\textsuperscript{-}\textit{i}.” The pixels of the SeaWiFS image took the label of the neuron we determined during the first processing. The PRSOM labeling and image-decoding procedure is summarized in Appendix A (Fig. A). The “Image-PRSOM-R\textsuperscript{-}\textit{i}” map was dedicated to a specific image only and must be determined for each image, whereas the PRSOM-R, is very general, being determined from Data\textsuperscript{obs} and Data\textsuperscript{expert}.

4.2. Analysis of aerosol events

Fig. 5 shows a time sequence of three days (August 6–8, 1999) of SeaWiFS images processed by PRSOM-R showing the development of a large Saharian dust event. This event originated in Tunisia and eastern Algeria (Fig. 5, August 6, 1999) was invading the central Mediterranean Sea, reaching Italy (Fig. 5, August 8, 1999) and was driven by a south–west wind blowing during that period (wind map not shown). It is clear that the PRSOM decoding of the three successive SeaWiFS images was temporally and spatially coherent. In particular, the optical thickness values, which were low at the eastern end of the basin during the first 2 days, increased on the August 8, 1999, at which time a heavy aerosol dust cloud covered all the basin.

In order to demonstrate that the temporal variability had been properly sampled by using SeaWiFS images for 1999 only, and that the neural decoding can be used operationally, we considered another large dust event which took place in June 2000 (Fig. 6). It must be emphasized that for year 2000 no pixel was included in the training data set. We focused our analysis on one particular day (June 28, 2000) of this 3-day aerosol event (June 28–30, 2000). For that particular day, the optical thickness values were below 0.35 and, the eastern Mediterranean basin was quite clear (optical thickness values around
0.15), but a dust event was visible on the coast of Morocco (optical thickness values between 0.25 and 0.35). Fig. 6a presents the optical thickness given by the SeaWiFS product and Fig. 6b by the neural decoding for June 28, 2000. These two optical thickness values were in good agreement. It is important to note that the SeaWiFS product restricts its decoding to optical thickness values under 0.35 and flags some pixels that are recognized as dust by the neural decoding. Fig. 6c shows that the probability of dust presence is quite high near the Morocco coast and in agreement with the wind map (not shown). Fig. 7 shows the optical thickness and typology maps obtained when decoding by the neural procedure for the three successive days (June 28–30, 2000). The Saharan dust disappeared on June 29 and 30; although the optical thickness value was low, the typology nevertheless seems to be advected with the wind. The a-posteriori probability for the three successive days showed that the probability of having dust had decreased during these two days, which is in good agreement with the optical thickness maps. This gives a good insight of the quality of the typology and can be used to follow the dust cloud.

4.3. Validation

We finally used in situ AERONET optical thickness measurements to validate the results of our neural decoding. Fig. 8 shows the comparison between measurements taken at Lampedusa station by the AERONET in 2000 and those given by the neural decoding. Each pair of measurements is represented with its error bars. The $r(865)$ (optical thickness value at 865 nm) given by the neural decoding and the AERONET measurements were in good agreement. Moreover, it is evident that the high values were retrieved with the same accuracy as the low optical thickness values.

Besides for the neural decoding and the SeaWiFS products, we compared the Root Mean Square Error (RMSE) and the Mean Relative Error (MRE) which were computed from the in situ measurements and the predicted ones. These statistical estimators are defined as:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_{\text{estimated}} - x_{\text{observed}})^2},$$

$$\text{MRE} = \frac{1}{N} \sum_{i=1}^{N} \frac{|x_{\text{estimated}} - x_{\text{observed}}|}{|x_{\text{observed}}|}.$$

The results are presented in Table 1 for the 34 AERONET measurements collocated with those of SeaWiFS and the Neural classification for which the optical thickness value was less than 0.35 (SeaWiFS critical value) and in Table 2 for the 46 AERONET measurements collocated with those of the Neural Classification available, which include measurements where the optical thickness value was higher than 0.35. The RMSE and MRE were lower by about 6% for the neural
decoding with respect to the data from the standard SeaWiFS algorithm, for the 34 optical thickness values in common. The scatter plot of the 34 computed optical thickness values for the SeaWiFS and the neural classification versus those of AERONET (Fig. 9) shows that the neural method gives slightly better results. But it is also important to note that the neural decoding allows retrieval high optical thickness values (i.e., greater than 0.35, above which the SeaWiFS product does not work) with a good accuracy (Fig. 10).

5. Discussion and conclusion

We have developed an original and efficient method to retrieve optical properties (type and optical thickness) from TOA reflectance measured by satellite borne multi-spectral ocean-color sensors. The method is based on a neural network classification methodology. It makes use of the full spectrum of measurements to perform the aerosol identification. The method is developed in two steps. The first step consists in an unsupervised classification of a large set of clear-sky TOA reflectance spectra $\rho_{\text{obs}}$ measured by a space-borne sensor. For this classification, we used the so-called Kohonen map which aggregates similar spectra into a reduced set of pertinent groups. In the present work, we used a two-dimensional array with a large number of neurons ($20 \times 20 = 400$), providing a highly discriminative representation of $\rho_{\text{obs}}$. Each neuron of the map is associated with a particular reference vector (rv-R), which characterizes a group. The second step consists in labeling these groups by clustering them with synthetic TOA reflectance spectra whose optical properties (i.e., aerosol type or optical thickness) are known. These synthetic spectra were computed by using sophisticated radiative-transfer models. In the present study, we dealt with five aerosol types (maritime, oceanic, coastal, tropospheric and mineral) and several optical thickness values ranging from 0.05 to 0.8. We accounted for the viewing geometry of the sensor by clustering the simulated spectra into ten groups each of similar geometry, as defined by the sun zenith angle and the scattering angle. These simulated spectra were then projected onto the Kohonen map to label each neuron (group). This step is fully automatic and constitutes the focal point of the present paper. Difficulties may arise since some neurons captured several synthetic spectra corresponding to different aerosol types, so we have introduced an original procedure to overcome this ambiguity in the labeling. The ambiguity is removed during the image analysis by minimizing the Euclidian distance between the
synthetic spectra captured by the neuron and the observed pixel spectrum for the 510 nm, 670 nm and 865 nm wavelengths and taking into account the spatial context, as was explained in Section 4.1.

When the labeling is done, the Kohonen map can be used to process new images and estimate aerosol parameters. We applied it to analyze SeaWiFS images for the years 1999 and 2000 in the Mediterranean region. The Kohonen map was educated from pixels randomly extracted during the year 1999 in this region. The Data\textsuperscript{obs} covers all the seasons and the Mediterranean homogeneously. The analysis of SeaWiFS images was performed pixel-by-pixel by selecting the suitable labeling (in terms of viewing geometry), then by identifying the closest spectrum in the Kohonen map, which

![Graph](image1)

Fig. 9. Scatter plot of the optical thickness measurements computed by using neural decoding and the SeaWiFS product with respect to the AERONET measurement for the 34 concomitant measurements smaller than 0.35 at the Lampedusa station (35°31N, 12°37E) in the Mediterranean in 2000.

![Graph](image2)

Fig. 10. Scatter plot of the optical thickness measurements computed by using neural decoding with respect to the AERONET measurement for the 46 concomitant measurements at the Lampedusa station (35°31N, 12°37E) in the Mediterranean in 2000. The neural decoding is able to compute the optical thickness values up to 0.8.
finally gives the aerosol optical properties. This method leads to accurate and coherent results as shown by the comparison with in situ AERONET aerosol measurements done with a sunphotometer based at the Lampedusa Island in the Strait of Sicily in the Mediterranean and by the study of two aerosol events over the Mediterranean. We found that the PRSOM method is better by 6% than the SeaWiFS product. This improvement is statistically significant when this new method is compared to a standard one whose results are considered to be correct.

This new method allows identification of the aerosol type and retrieval of the aerosol optical properties from the statistical properties of the data. Besides, it gives accurate results for optical thickness values greater than 0.35, which is not the case for the SeaWiFS product. The method is also able to detect absorbing aerosols such as Saharian dusts, which is still a challenge as mentioned in the introduction. Moreover, a major advantage of the PRSOM method with respect to classical algorithms is the ability to compute probabilistic information associated with the reference vectors (co-variance matrix), which allows us to determine statistically relevant criteria for the retrieving of the expert labels. This statistical information can be used to estimate the probability of one pixel belonging to a certain class and thus to obtain a confidence of the classification of each pixel and consequently of the retrieved values of the aerosol parameter. This information has been used with success for estimating the aerosol type as shown in Figs. 6c and 7. The significance of these statistical indices is related to the number of data processed during the learning phase versus the number of classes. In the present study, the learning phase comprises 2,346,147 clear-sky spectra versus 400 classes in the PRSOM algorithm; this ratio enabled us to have a significant amount of spectra in each class ensuring an efficient computation of the statistical indices. Besides the power of the PRSOM algorithm would allow us to increase the number of retrieved aerosol types with respect to classical methods. The accuracy of the results is mainly controlled by the quality of the aerosol models introduced into the radiative transfer models used to construct the synthetic database.

The neural methodology presented in this paper and the labeling procedure are very general, on the contrary to physical methods that are dedicated to a specific problem. They can thus be applied for solving a large variety of problems encountered in geophysics.

Acknowledgments

We are grateful for the support we received from the European project NAOC (Neural Algorithm for Ocean Color). We thank Dr. H.R. Gordon for providing the syntheitical database, Dr. S. Pugnaghi, Principal Investigator of Lampedusa AERONET station for maintaining and collecting the ground-based data and Amine Boussarsar for the job he has done training the PRSOM-A map. We also thank Dr. R. Griffiths for stimulating discussions.

Appendix A. PRSOM labeling and image-decoding procedure

A.1. Labeling procedure with Dataexpert (this phase is processed once)

1. Decompose the geometry of Data^{obs}_{i} into pertinent bin-\textit{i} classes (ten in the present study) by building a dedicated Kohonen map (PRSOM-Angles).
2. Assign a geometry bin-\textit{i} to each \rho^{exp}_{q} of the Data^{exp}_{i}.
   Let us denote Data^{exp}_{i}, the set of \rho^{exp}_{q} assigned to the geometry bin-\textit{i}.
3. Process Data^{exp}_{i} with PRSOM-R, to obtain the PRSOM-R_{i} maps in which the neurons have captured some spectra of Data^{exp}_{i}.
4. Label the “pure neurons” (neurons corresponding to synthetic spectra having close optical properties); a “pure neuron” takes the label of the common type or the label “maritime”, for the small optical thickness.
5. Label the “white neurons” (neurons having captured no synthetic spectra) by taking the label of the closest neurons on the map (use of the topology of the map).
6. Label “mixed neurons” during the Image decoding.

A.2. SeaWiFS image decoding and labeling “mixed neurons” (this phase is processed for each image)

1. Process the image with PRSOM-Angles to assign a bin-\textit{i} class to each pixel.
2. Process the image with the ten PRSOM-R_{i}. Begin to label the “mixed neurons”.
   Initialize the aerosol counters to zero. Each time an aerosol type is encountered, increase the corresponding counter by one unit. The aerosol type is the type of the counter having captured the highest number of units. We obtained a dedicated PRSOM-R_{i} map, the so-called “Image-PRSOM-R,” map in which the “mixed neurons” are labeled.
3. Process the image with the ten “Image-PRSOM-R,” maps. The pixels of the SeaWiFS image take the label of the neurons that have been determined during the preceding procedure described in (2).

1. For each pixel \textit{p} of the SeaWiFS image
   1.1 Determine the geometry bin-\textit{i} using the PRSOM-A
   1.2 Choose PRSOM-R_{i} map and Assign pixel \textit{p} to the the nearest neuron \textit{N}
   \textbf{IF} \textit{N} is a “pure neuron” with type \textit{T}
   \textbf{THEN} \textit{p} has label \textit{T}
   \textbf{Go to 1.}
   \textbf{ELSE if} \textit{N} is a “mixed neuron”
   Determine the nearest mean spectrum in the NIR (510 nm, 670 nm, 865 nm); let us denote \textit{T} its type
   Increase by one the counter of the neuron \textit{N} for the selected type \textit{T}
   \textbf{Go to 1.1}
2. Determine the type \textit{T} of each “mixed neuron” using the majority vote
3. Compute the optical thickness of the “mixed neuron” and “white neuron”
4. Assign the aerosol type and the optical thickness of the selected neuron \textit{N} to each neuron of the SeaWiFS image
The flow diagram is as follows (Fig. A).

Fig. A. Schematic representation for the introduction of the expertise.

References


