Automatic neural classification of ocean colour reflectance spectra at the top of the atmosphere with introduction of expert knowledge

Awa Niang\textsuperscript{a}, Lidwine Gross\textsuperscript{b}, Sylvie Thiria\textsuperscript{b,\*}, Fouad Badran\textsuperscript{c}, Cyril Moulin\textsuperscript{d}

\textsuperscript{a}Laboratoire de Physique de l’Atmosphère Siméon Fongang, Ecole Supérieure Polytechnique, Université Cheikh Anta Diop de Dakar, Senegal
\textsuperscript{b}Laboratoire d’Océanographie Dynamique et de Climatologie, Université Pierre et Marie Curie, Paris, France
\textsuperscript{c}Centre d’Études et de Recherche en Informatique au CNAM, Conservatoire National des Arts et Métiers, Paris, France
\textsuperscript{d}Laboratoire des Sciences du Climat et de l’Environnement, Commissariat de l’Énergie Atomique, Gif-sur-Yvette, France

Received 5 March 2002; received in revised form 5 February 2003; accepted 5 February 2003

Abstract

We propose an automatic neural classification method for ocean colour (OC) reflectance measurements taken at the top of the atmosphere (TOA) by satellite-borne sensors. The goal is to identify aerosol types and cloud contaminated pixels. This information is of importance when selecting appropriate atmospheric correction algorithms for retrieving ocean parameters such as phytoplankton concentrations. The methodology is based on the use of Topological Neural network Algorithms (TNA, so-called Kohonen maps). The pixels of the remotely sensed image are characterised by a vector whose components are the spectral TOA measurement and the standard deviation of a small spatial structure. The method is a three-step method. The first step is an unsupervised classification built from a learning data set; it clusters pixel vectors which are similar into a certain number of groups. Each group is characterised by a specific vector, the so-called reference vector (rv), which summarises the information contained in all the pixels belonging to that group. The second step of the method consists of labeling the reference vectors with the help of an expert in ocean optics. The groups are then clustered into classes corresponding to physical characteristics provided by the expert. The third step consists of analyzing full images and classifying them by using the classifier which has been determined during the first two steps. The method was applied to the Cape Verde region, which exhibits important seasonal variability in terms of aerosols, cloud coverage and ocean chlorophyll-\(a\) concentration. We processed POLDER data to test the algorithm. We considered four classes: pixels contaminated by clouds; two types of pixels containing mineral dusts; and pixels containing maritime aerosols only. The method was able to take into account the information given by the expert and apply it to unlabeled pixels. This methodology could easily be extended to a larger number of classes, the major problem being to find adequate expertise to label the classes.

© 2003 Elsevier Inc. All rights reserved.

Keywords: Ocean colour; Classification; Neural networks; Remote sensing

1. Introduction

During the past few years, several sensors dedicated to ocean colour (OC) observation have been launched on different satellites (POLDER, SEAW-IFS, MOS, MODIS, OCTS). These sensors are multi-spectral. The number of channels of the forthcoming sensors is expected to dramatically increase (MERIS, GLI). We can thus expect to improve the accuracy in determining the physical parameters provided by ocean colour sensors such as atmospheric aerosols or phytoplankton content in the ocean. In fact, aerosol and phytoplankton retrievals are linked since the atmospheric part of the signal is about 80–90% of the top of the atmosphere (TOA) signal. Most of it is due to the scattering and the absorption of solar radiation by aerosols and air molecules (the latter being quite easy to model). A critical issue for OC remote sensing processing is thus to remove the aerosol contribution to the measured signals in order to obtain the actual spectrum of marine reflectance which is directly related to the water constituent concentration. The impact of aerosol on the radiance spectrum depends on the number of particles and on their physicochemical characteristics. Indeed, aerosols made of large particles (e.g. sea-salts) have a near-neutral spectral signature, while small aerosols (e.g. sulfates) scatter more efficiently the solar light at shorter wavelengths.
Current atmospheric correction algorithms use TOA reflectances in the near infrared (NIR) to assess the aerosol optical properties (Gordon & Wang, 1994). At these wavelengths, water absorption is huge and particle scattering is weak, so that the ocean can be considered as “black”. The atmospheric contribution is therefore readily obtained in the NIR and is extrapolated to the visible spectrum to perform the atmospheric correction. These algorithms have been shown to be accurate in most oceanic regions where strongly absorbing aerosols are not present.

However, absorbing aerosols (mineral dust, black carbon, ... ) have a spectral signature that is similar to that of purely scattering aerosols, such as sulfates or sea-salts, in NIR, but that becomes dramatically different at ocean colour wavelengths (i.e. in the blue and green) (Gordon, 1997). If absorbing aerosols are not distinguished from the usual aerosol type and the standard atmospheric correction is applied, it removes too much radiance in the blue part of the spectrum. This in turn causes important derived products that strongly depend on values of radiance in the blue wavelengths, such as chlorophyll concentration to be incorrect. This shows the importance of the knowledge of the atmospheric aerosol composition for performing accurate atmospheric correction.

The present study focuses on the ability of classification techniques to discriminate between the different aerosols. To achieve this, we investigate the use of a specific Topological Neural network Algorithm (TNA), the so-called PRobabilistic Self Organizing Map (PRSOM) (see Section 3), to classify TOA reflectance spectra. TNAs are well adapted for this task (Kohonen, 1984) and have been used with success by Ainsworth and Jones (1999) for classifying the TOA reflectances of the OCTS sensor.

It is expected that the different groups determined by the TNA will correspond to specific aerosol types and concentrations. In the ultimate phase, thanks to the expert knowledge, we expect that the classes provided by the full process will allow us to discriminate between the different aerosols (specifically absorbing or non-absorbing aerosols). This should provide useful background information for inverting the TOA signal and making appropriate atmospheric corrections and also for studying the radiative budget of the earth where aerosol absorption and scattering effects play an important role.

The paper is arranged as follows: in Section 2, we describe the problem and the data set. In Section 3, we briefly describe the methodology we used. Section 4 deals with expert knowledge. In Section 5, we analyze the results. Discussion and conclusion are given in Section 6.

2. The data set

In order to check the feasibility of the method we have developed, we processed images centred on the Cape Verde region. This area presents several advantages for this study. Tropospheric aerosols of different kinds are often present (Moulin, Guillard, Dulac, & Lambert, 1997). There is a natural background of maritime aerosols containing a mixture of biogenic sulfates and sea-salt particles (maritime model (Schwindling & Deschamps, 1998; Shettle & Fenn, 1979)). These aerosols are non-absorbing. Absorbing aerosols having a continental origin are often also present. Mineral dusts from the Sahara and Sahel are (Shettle, 1984) the prevailing aerosols in this region from early spring to autumn, whereas carbonaceous aerosols from biomass burning in Africa are often observed in the southern part of this region during winter (Shettle & Fenn, 1979). They present a high spatial and temporal variability associated with the trade winds (the so-called Harmattan in this region). Both dusts and carbonaceous aerosols strongly absorb in the visible part of the spectrum, and thus require dedicated atmospheric correction (Moulin, Gordon, Chomko, Banzo, & Evans, 2001). It is very important to be able to identify them, since the classical atmospheric correction algorithms no longer operate when they are present. In addition, Sahara dusts can have high optical thicknesses, and thus give a very bright signal. In classical OC processing chains, they are currently confused with water clouds.

The images we used for our study were supplied by the radiometer POLDER on board the ADEOS satellite (Deschamps et al., 1994). They are centred on Cape Verde and have a size of 500 × 500 pixels, i.e. about 3000 × 3500 km² (the resolution of a POLDER pixel is 6 × 7 km² at nadir). We sampled several aerosol events and part of the seasonal variability of the solar illumination by selecting several dates: January 29, 1997 (orbit #5080); March 7, 1997 (orbit #6023); June 8, 1997 (orbit #8180, see Fig. 1a–c).

POLDER acquires measurements in nine spectral bands centred on 443, 490, 565, 670, 763, 765, 865 and 910 nm (among which three offer polarisation measurements: 443, 670 and 865 nm). For each pixel, up to 14 sets of observations with varying viewing geometry angles (θv, θs, φ) area available: respectively, θs, the solar zenith angle; θv, the viewing zenith angle; and φ, the azimuthal difference between the POLDER and Sun plane. For each spectral channel, the POLDER product gives the TOA measurements in reflectance ρ:

\[
ρ(λ, θ_v, φ) = \frac{πL(λ, θ_v, φ)}{E_0(λ)\cos(θ_v)}
\]

where \(L(λ, θ_v, φ)\) is the radiance measured by the POLDER sensor (in W m\(^{-2}\) sr\(^{-1}\) nm\(^{-1}\)), and \(E_0(λ)\) is the extraterrestrial solar irradiance (in W m\(^{-2}\) nm\(^{-1}\), variable with the Sun–Earth distance).

We chose to process five unpolarised reflectances at wavelengths 443, 490, 565, 670 and 865 nm. We selected the spectral reflectances having the minimal view zenith angles (min (θv), this angle ranging between 0° in the centre of POLDER orbits and 70° at the edge) to minimize multi-
scattering effects on the signal. In the following, we neglect the directional aspect of the reflectances.

A TOA reflectance $\rho$ can be divided into several components (Gordon, 1997): $\rho_{\text{path}}$, the reflectance generated along the optical path by scattering in the atmosphere; $\rho_{\text{sp}}$, the contribution arising from specular reflection of direct sunlight from the sea surface (sun glitter); $\rho_{\text{wc}}$, the contribution arising from sunlight and skylight reflected from

Fig. 1. Representation of the three POLDER images. (a–c) RGB composite of the scenes using three channels: $\rho_{443}$ for red, $\rho_{565}$ for green and $\rho_{865}$ for blue. (d–f) Representation of the same images by the 400 rvs. All the pixels that have the same gray level belong to the same rv.
whitecaps on the sea surface; and $\rho_w$, the contribution of the water:

$$\rho(\lambda) = \rho_{\text{path}}(\lambda) + T(\lambda)\rho_g(\lambda) + t(\lambda)\rho_{\text{wc}}(\lambda) + t(\lambda)\rho_w(\lambda)$$

(2)

where $T$ and $t$ are the direct and diffuse transmittance of the atmosphere, respectively. The term $\rho_{\text{path}}$ can be decomposed as follows (Deschamps, Herman, & Tanre, 1983; Gordon & Wang, 1994):

$$\rho_{\text{path}}(\lambda) = \rho_s(\lambda) + \rho_a(\lambda) + \rho_{\text{ra}}(\lambda)$$

(3)

where $\rho_s(\lambda)$ is the reflectance resulting from Rayleigh multiple scattering (air molecules) in the absence of aerosols, $\rho_a(\lambda)$ is the reflectance resulting from multiple scattering of aerosols in the absence of the air and $\rho_{\text{ra}}$ is the interaction term between molecular and aerosol scattering.

We removed pixels contaminated by the sun glitter using a geometrical mask, thus in our data set, the term $T(\lambda)\rho_g(\lambda)$ in Eq. (2) is assumed to be zero. The reflectances were then corrected for Rayleigh scattering $\rho_s(\lambda)$ (Eq. (3)), which can be accurately computed using atmospheric pressure. The reflectances we propose to classify are thus composed of:

- the aerosol contribution, $\rho_a(\lambda) + \rho_{\text{ra}}(\lambda)$ (Antoine & Morel, 1999; Gordon, Du, & Zhang, 1997),
- the whitecap contribution, $\rho_{\text{wc}}(\lambda)$, which is generally small except for high surface wind speed (Frouin, Schwindling, & Deschamps, 1996),
- the water contribution $\rho_w(\lambda)$ (Morel, Voss, & Gentili, 1995; Yang & Gordon, 1997).

The signal we finally used is thus of the form:

$$\rho_{\text{used}} = \rho_s + \rho_{\text{ra}} + t(\rho_{\text{wc}} + \rho_w)$$

(4)

In the above equation, the oceanic term $t(\rho_{\text{wc}} + \rho_w)$ is small in the near infrared, thus $\rho_{\text{used}}$ mainly depends on the atmospheric one $\rho_s + \rho_{\text{ra}}$. It is expected that the atmospheric term in the visible highlights the infrared contribution and allows us to retrieve the aerosols. This hypothesis is supported by the results (see below).

In order to separate liquid water clouds, which are generally optically inhomogeneous on scales of several km$^2$, from aerosol clouds, which are relatively homogeneous on this scale (Moulin et al., 2001), we added a structural information to the multi-spectral information of each pixel. This structural information takes the form of the standard deviation of the reflectances at 670 nm (the wavelength for which Sahara aerosols and the liquid water clouds give a high signal), computed over a region centred on the pixel being considered with a window of $3 \times 3$ pixels (i.e. 18 $\times$ 21 km). This standard deviation (hereafter denoted as $\sigma_{670}$) is used as a measurement of the spatial homogeneity of observed structures to aid the classification.

Each pixel of the POLDER image is now represented by a six-dimensional vector (the five corrected reflectances $\rho(\lambda)$ and the standard deviation at 670 nm, $\sigma_{670}$). In order to create a training data set representative of the studied events and to reduce the dimension of this data set, we sampled the three POLDER images by taking one line over 10. We denote the training set of the TNA (composed by 43739 pixels) by $\mathcal{D}$. The method will be tested on the full images which can be considered as quasi-independent of $\mathcal{D}$, since 9/10 of their pixels were not used in the training of the classifier.

We now present the methodology we used. The method is a three-step method. The first step is an unsupervised classification built from a learning data set; it clusters the pixel vectors that are similar to a large number of groups which condense the information contained in the learning data set. The second step of the method consists of labeling the reference vectors ($\mathcal{R}_i$) with the help of an expert in ocean optics. The groups are then clustered into classes corresponding to physical characteristics provided by the expert. The third step consists of analyzing full images and classifying them.

3. The topological network clustering

In this section, we present the first step which is an unsupervised classification.

The training data set $\mathcal{D}$ consists of a six-component vector associated with each pixel (five channel values of the TOA spectrum plus the spatial standard deviation). The first step in the classification is to summarise the information contained in the training data set $\mathcal{D}$ by producing a set of reference vectors ($\mathcal{R}_i$) that are representative of the data.

![Architecture of the two dimensional self-organizing map and the hierarchical clustering.](image_url)
To do this, we used a neural network-based model, the so-called topological probabilistic map, where each neuron of the map is associated with a particular reference vector (rv) and thus corresponds to a group (in this paper, a set of pixels belonging to D) (see Appendix 1). The different neurons of the topological map C are connected together and determine a topological (neighbourhood) relationship between the different groups (neurons) (See Fig. 2 and Appendix 1). This neighborhood is a smoothing factor of the neuron parameters which facilitate the handling of a large number of groups. In the present study, we used a probabilistic model, the so-called PRobabilistic Self Organizing Map (PRSOM), which is an extension of the well-known Self Organizing Map (SOM) model described by Kohonen (1984) for visualising and clustering high-dimensional data set (see Appendices 2 and 3). We dealt with a two-dimensional topological map with a large number of neurons (20 × 20), providing a highly discriminating representation of the observations. The 43,739 pixels of D are thus clustered into 400 groups.

The clustering of D we obtained by using 400 rvs gives a rational partition of the data set: each reference vector represents 109 pixels on average, at most 560, and at least 13, which means that the observations of D are well distributed over all the reference vectors. Fig. 3 displays the rvs of the PRSOM map, and allows us to visualise the topological order given by the learning algorithm. It is noted that the reference vectors which are close together in the two-dimensional topological map have similar spectra thanks to the neighborhood procedure.

In order to show the ability of the rv to represent data set, we displayed the 15 two-dimensional scatter plots of x_i against x_j for our six-dimensional input space, x for the data set D in Fig. 4 and for the rv set in Fig. 5. When comparing the two figures, we see the good agreement between the 15 scatter plots represented in each figure and conclude that the set of rvs represents the full data set D and compresses the information contained in the initial TOA signal.

Fig. 3. The 400 reference vectors of the topological map: each square represents a neuron of the PRSOM map. In each square, the six-dimensional reference vector is represented (the five corrected reflectances ρ(λ) and σ_670, with a y-scale between 0 and 1).

Fig. 4. The 15 two-dimensional scatter plots with respect to the data space: projection of the D set: (1) ρ_490 vs. ρ_443, (2) ρ_565 vs. ρ_443, (3) ρ_670 vs. ρ_443, (4) ρ_670 vs. ρ_443, (5) σ_670 vs. ρ_443, (6) ρ_565 vs. ρ_443, (7) ρ_670 vs. ρ_443, (8) ρ_670 vs. ρ_443, (9) σ_670 vs. ρ_443, (10) ρ_670 vs. ρ_565, (11) ρ_670 vs. ρ_565, (12) σ_670 vs. ρ_565, (13) ρ_670 vs. ρ_670, (14) σ_670 vs. ρ_670, (15) σ_670 vs. ρ_670. The reflectances ρ(λ) and the σ_670 were normalised between −1 and +1.
At that level, the PRSOM classifier is an unsupervised classifier. We can apply it to the POLDER images to obtain an unsupervised classification. Each pixel of an image is associated with a specific \( rv \) (or neuron of the TNA) among the 400 \( rv \)s. In order to check the physical consistence of this decomposition, we have represented the three POLDER images where each pixel is represented by a gray level corresponding to the value at 865 nm of the spectra associated with its corresponding \( rv \) (Fig. 1d–f).

We recall that a \( rv \) is characterized by the five values of its synthetic spectra (plus the spatial standard deviation). It is noted that these 400 gray values are able to give a coherent partition of the initial maps (Fig. 1a–c) where their major structures are retrieved. This spatial coherence endorses the physical significance of the decomposition of the pixels in 400 groups which constitutes a compression (vector quantization in formal statistic) of the information associated with the image.

**Fig. 5.** The 15 two-dimensional scatter plots with respect the data space: projection of the \( rv \) ensemble.

**Fig. 6.** Reflectance spectra for the four labeled classes: full lines design the mean spectrum and dotted lines the standard deviation.
These above results show that the classification in groups presents a physical consistence and allows us to proceed to the second step of the classification which is a supervised one. It consists of determining classes of reference vectors (the aerosol types) in order to facilitate the interpretation in terms of physical parameters. This is done by identifying the different \( \text{rvs} \) topological map and labeling them with the help of an expert (see Appendix 3 for the algorithmic development).

4. Introducing expert knowledge

The 400 groups are difficult to interpret in terms of physical significance. The aim of this second step, which now is a supervised one, is thus to cluster the 400 \( \text{rvs} \) (reference vectors) into four classes corresponding to different aerosol types using the knowledge of an expert in ocean optics. The expertise is based on simple physical considerations and is decomposed in two phases.

- After a quality control, the expert labels some pixels of the three POLDER images using some knowledge provided by the atmospheric correction of the POLDER data processing. In this case, the knowledge is extracted from the value of the optical thickness of aerosols at 865 nm such as \( \tau_{a}(865)<1.0 \), where the POLDER data processing is confident.
- We build an association between the expert-labeled pixels (from phase 1) and the referent vectors (from Section 3). We use this association to transfer a single label to each \( \text{rv} \), so that each of the 400 \( \text{rvs} \) has one of the four classes. This phase is complex since the expert-labeled pixel set does not represent all the \( \text{rvs} \). The process is described further.

At the present stage of this research, the labeling procedure is a rough approximation and is introduced more for a methodological purpose than for a physical one: the aim is to see how the labeling of some pixels can be dispatched to the pixels of the whole image. More sophisticated labeling procedures are under development and were not available when we accomplished this study.

Four distinct types of ‘aerosols’ were distinguished corresponding to pixels containing maritime aerosols (\( \text{Sea} \) label), to pixels containing Sahara aerosols (\( \text{Aerosols} \) 1 and 2 labels) and to pixels contaminated by clouds (\( \text{Clouds} \) label).

- The type labeled \( \text{Sea} \) is identified as pixels for which \( \tau_{a}(865)<0.1 \) (18,149 pixels). We assume that under this condition, the aerosol contribution to the signal, \( \rho_{a} = \rho_{a}(\lambda) \), is only due to maritime aerosols. The total reflectance spectra of this group (Fig. 6a) are very weak in the NIR and increase while the wavelength decreases. This spectral behavior is mostly due to \( \rho_{a}(\lambda) \) (with respect to the intensity of the reflectance spectra \( \rho \)) which is weak in agreement with relatively homogeneous sea.
- The type \( \text{Aerosols} \) 1 is identified as pixels for which \( 0.5 < \tau_{a}(865) < 0.75 \) (34,777 pixels). This type of data with high optical thickness contains mostly Sahara dusts and may also contain soot. However, as we can see in Fig. 6b, two spectral behaviors are present in this type, showing that it still contains some isolated maritime aerosols. Indeed, we can see reflectance spectra with high values in the blue and weak signal in the NIR, which correspond to the contribution of maritime aerosols, while absorbing aerosols make the blue part of the signal decrease (Gordon et al., 1997). Fortunately, these maritime aerosols represent only 1.63% of this group. It is noticed that \( \sigma_{670} \) has the same behavior as that in the subset \( \text{Sea} \) (not shown).
- The type \( \text{Aerosols} \) 2 is identified as pixels for which \( \tau_{a}(865)>0.75 \) (15,228 pixels), and it represents thick and homogeneous clouds of Sahara dusts (see Fig. 6c).
- The type \( \text{Clouds} \) has been obtained by using a classical test of cloud coverage (threshold on \( \rho_{865} \) and \( \sigma_{865} \)) but with extremely severe limits. Fig. 6d shows the white spectral behavior of clouds, it corresponds to inhomogeneous structures with strong \( \sigma_{670} \). It corresponds to pixels for which the cloud coverage is important (110,782 pixels).

The set of expert-labeled pixels is displayed in the images (Fig. 7a–c). It is noticed that a huge number of pixels (in white on the images) have no label. This set is denoted herein after expert-non-labeled pixels. Notice that the expert-labeled set is different from the training set.

Each reference vector \( \text{rv} \) may correspond to a certain amount of these expert-labeled pixels. In fact, there are three possible outcomes of associating expert-labeled pixels with an \( \text{rv} \). An \( \text{rv} \) may get:

1. One or more pixels, all with the same label,
2. More than one pixel with different labels,
3. No pixel.

For case 1, the \( \text{rv} \) takes the label of the affected pixels. For case 2, an additional step is needed. We apply the majority vote rule to label the \( \text{rv} \) the \( \text{rv} \) takes the most frequent label. For case 3, it is possible that some \( \text{rvs} \) were not associated with expert-labeled pixels. In standard PRSOM, these reference vectors would remain unlabeled. However, by applying the modified eHC-PRSOM method (Yacoub, Frayssinet, Badran, & Thiria, 2000) (see Appendix 3), which is based on a statistical automatic classification (Hierarchical Clustering HC) (Jain & Dubes, 1988), we are able to spread the knowledge of the labeled \( \text{rvs} \) the remaining ones. Hence, at the end of eHC-PRSOM, each reference vector \( \text{rv} \) is identified by its label, denoted \( L_{k} \) \( (k=1 \) stands for \( \text{Sea} \), \( k=2 \) for \( \text{Aerosols} \) 1, \( k=3 \) for \( \text{Aerosols} \) 2 and \( k=4 \) for \( \text{Clouds} \)).

Fig. 8b–e displays on the PRSOM map (which was first presented in Fig. 3) the percentage of expert-labeled pixels having a definite label captured by each reference vector.
Conversely, Fig. 8a shows on the same map the percentage of expert-non-labeled pixels. The left-bottom corner of the PRSOM map is well identified; that is, the percentage of expert-non-labeled pixels is weak. The neurons (hence rvs) of this region have no ambiguity, and Fig. 8e shows that they correspond to Clouds. We also note that three reference vectors are clearly identified as Sea (3rd line, 16th column, and 4th line, 15th and 16th columns of the map in Fig. 8b). The other reference vectors capture a large percentage of unlabeled pixels in addition to a small percentage of labeled ones. After applying the labeling process described above, the total labeled PRSOM map (provided at the end of eHC-PRSOM) is shown in Fig. 8f.

We can see that most of the neurons of the map are dedicated to the label Clouds, since most of the rvs have captured Clouds pixels.
Since PRSOM is a probabilistic model:

- the class of reference vectors with label $L_k$ approximates the probability density function of the reality for that label.
- it allows the computation of the posterior probability $p(\mathbf{rv}/x)$ of each observation $x$ belonging to a given $\mathbf{rv}_j$, and thus the posterior probability $p(L_k/x)$ of an observation $x$ belonging to a group $L_k$.

In our experiment, $k=1,\ldots,4$, so four posterior probabilities $p(L_k/x)$ are computed and an observation (or a pixel) is assigned to the label having the highest probability according to the Bayes’ theorem. At the end of the process, each pixel of the image is assigned to a particular $\mathbf{rv}$ which now belongs to a class and represents a particular aerosol type.

If we plot the reflectance spectra of the labeled $\mathbf{rvs}$ (Fig. 9), in a similar way as in Fig. 6, we observe that their spectral pattern presents the same characteristics as the expert-labeled subsets of pixels. In particular, $Aerosols$ 1 presents two different spectral behaviors, showing some maritime aerosols among Sahara dusts. It is also found that $\sigma_{670}$ of the $\mathbf{rvs}$ are similar to the values of the expert-labeled pixels (not shown).
5. Analyzing the POLDER images

We are now able to proceed to the third step and determine the aerosol types of the three POLDER images. Each pixel of an image is presented to PRSOM which assigns it to a specific \( r_v \) and takes the class of this \( r_v \). The results are displayed in Fig. 7d–f. When comparing Fig. 7a–c and Fig. 7d–f, we see the spreading of expert labeling (which concerns a few pixels of the images) to the whole image. This clearly shows the power of the \( \epsilon \)HC-PRSOM method.

When compared to the RGB images displayed in Fig. 1a–c, it seems that the pixels which have been contaminated by clouds have been correctly classified as Clouds. However, as we can see on orbits #6023 and #8180, some very bright yellow pixels containing mineral dusts have been incorrectly classified as Clouds. We note in Fig. 1d–f that some \( r_v \)s representing these pixels should be separated from Clouds. In Fig. 7b and c, these pixels were not labeled by the expert. Consequently, it is during the procedure transferring the reference vectors to the full images that these pixels were classified as Clouds. The Sea-labeled pixels (maritime aerosols only) in Fig. 7d–f correspond to the dark blue region of the RGB composition (Fig. 1a–c) as expected (at the edges of the RGB pictures, pixels look more blue than in the middle because of multiple scattering effects). The pixels labeled Aerosols 1 and 2 (corresponding to two different optical thickness ranges of mineral dusts) seem very consistent with the RGB composite. Thus, with the exception of pixels which have been labeled Clouds instead of Aerosols 2, the proposed classification method has consistently spread the expert knowledge to the whole set of images.

In order to estimate the amount of information lost during the “spreading” of the expert knowledge, we took each pixel labeled by the expert, and verified to which class they were assigned by our method. The results are summarised by the confusion matrix \( M=\{m_{ij}\} \), where \( m_{ij} \) stands for the number of pixels classified in class \( i \) by the expert and by class \( j \) by TNA. The confusion matrix (Table 1) shows that on all the three images, a mean of 91.54% of the labeled pixels has kept their label. At most, 1% of the Sea and Aerosols 1 and 2 pixels has been classified Clouds, and less than 1% of the Clouds pixels has been badly labeled; thus, we can conclude that the Clouds have been well identified. However, a weak confusion remains on the other classes: some Sea pixels (8.07%) have been classified Aerosols 1, some Aerosols 1 pixels (12.43%) have been classified Aerosols 2 and some Aerosols 2 pixels (8.41%) have been classified Aerosols 1.

![Fig. 9. Reflectance spectra of the \( r_v \)s in four classes using the \( \epsilon \)HC-PRSOM method.](image)

Table 1

<table>
<thead>
<tr>
<th>( \epsilon )HC-PRSOM</th>
<th>Sea</th>
<th>Aerosols 1</th>
<th>Aerosols 2</th>
<th>Clouds</th>
</tr>
</thead>
<tbody>
<tr>
<td>e</td>
<td>0.9177</td>
<td>0.0807</td>
<td>0</td>
<td>0.0015</td>
</tr>
<tr>
<td>x Aerosols 1</td>
<td>0.0283</td>
<td>0.8435</td>
<td>0.1243</td>
<td>0.0039</td>
</tr>
<tr>
<td>p Aerosols 2</td>
<td>0</td>
<td>0.0841</td>
<td>0.9053</td>
<td>0.0106</td>
</tr>
<tr>
<td>Clouds</td>
<td>0.0012</td>
<td>0.0023</td>
<td>0.0015</td>
<td>0.9950</td>
</tr>
</tbody>
</table>

The matrix has been made so that the sum of the amounts on each line is equal to 1.
Fig. 10 shows pixels for which the highest probability \( p(L_k/x) \) (Eq. (20)) is smaller than 0.75. These pixels mostly appear at the edges of the clusters, and correspond to the transition between the sought geophysical classes. For some of them, the largest probability is very close to 0.5 and thus the result of the classifier can be considered undetermined. The probability \( p(L_k/x) \) can be used to control the classification and even to eliminate some unidentified pixels (by providing a reject option).

6. Discussion and conclusion

The classification method we have presented here is a three-step method. The first two concern the determination of the classifier, the third the application. First, we extracted the most pertinent spectral forms from the POLDER images by compressing the information of the images in 400 reference vectors by using PRSOM (unsupervised steep). Second, we labeled these vectors with the help of an expert (supervised step). We clustered these 400 groups into four classes corresponding to the partition given by the expert (Clouds, Aerosols 1 and 2, Maritime Aerosols). For this, we used the majority vote rule provided by the εHC-PRSOM method. This gave coherent results with regards to the initial expert knowledge, and we were able to extend this expertise to new observations.

Most of the reference vectors were assigned to clouds (323) because of the high variability of the spectra of the different cloud pixels (0.02 < \( \rho < 2.10 \)).

Third, we apply the supervised classifier to the three POLDER images. When we consider the application of these 323 reference vectors to the POLDER images, it appears that the resulting “cloud detection” is consistent with both the RGB compositions and the initial expert knowledge. However, a few mineral dust areas with very high optical thickness are still confused with water clouds. This is due to the difficulty of expert classification in such situations and thus it is expected that this could be corrected using more accurate rules for labeling the pixels.

The 77 remaining reference vectors are used for aerosol classification. The PRSOM method reproduced the expert labels 88.88% of the time. The wrong classification for the remaining 11.12% pixels can be explained in different ways: 1—the majority vote rule which may seem too restrictive, 2—the errors in the expert labels. It might be envisaged to use a pondered majority vote rule.

The major advantage of PRSOM with respect to classical algorithms is the ability to provide probabilistic information on the reference vectors (co-variance matrix), which allows us to use statistically relevant criteria (Jain & Dubes, 1988) (Ward Index) to retrieve the expert labels. This statistical information also allows us to estimate the probability of one pixel belonging to a certain class, and thus to obtain a confidence index of the classification of each pixel.

Furthermore, we showed that the PRSOM method can be used for cloud detection procedure which can separate water-clouds from Sahara dusts, even in high optical thickness situation. Due to the large amount of reference vectors dedicated to the “clouds”, this study suggests, however, that it would be better to first build a classifier dedicated to the detection of cloud-contaminated pixels from the others, then to remove these cloud pixels from the image in order to
avoid having to deal with the physical range of cloud spectra. Once clouds are removed, another classifier dedicated to aerosols can be applied. This method also showed that non-absorbing aerosols could be distinguished from two different types of absorbing aerosols. This method could efficiently replace current cloudiness tests in Ocean Colour operational processing, and could be used as the first step of atmospheric correction by selecting the type of algorithm to be used, depending on the abundance of absorbing aerosols.

We saw that the present methodology was able to define a large number of groups (400). This number is somewhat arbitrary. This decomposition is done to condense the information embedded in the learning data set. Since the labeling is done on the group \( \mathbf{r}_v \) set, this set must be large enough to be representative of the learning set \( \mathcal{D} \) and of the large variety of the physical information which can be qualitative (aerosol types) or quantitative (physical parameter, e.g. concentration). If the group set was not large enough, the eHC-PRSOM classification might lead to ambiguous classification. Moreover, a major problem is to correctly interpret this information. We have to face the inability of experts to assign a physical significance to groups.

For an operational purpose, further improvements of the classifier will be the following ones:

- use of the spatial context to improve the coherence of the image. This point is under development.
- enlarge the area under study in order to generalize the validity of the method. The Mediterranean Sea would be a very interesting region to investigate since it experiences very different aerosol conditions, from pollution to mineral dust and fire plumes.
- as this method strongly depends on expert knowledge, increase in the number and the type of the labeled data sets, this might be achieved by using radiative transfer computations for several different aerosol types (small or large as well as absorbing or non-absorbing).
- use of in situ data to validate the classification results. Data collected during intensive field campaigns (INDOEX, ACE-Asia) that include both chemical and optical characterization of the aerosol from the surface to the free troposphere would be suitable to validate our method in regions where different aerosol regimes can be found from 1 day to the other.

Acknowledgements

This work was supported by the Centre National des Etudes Spatiales (convention No. 794/CNES/98/7118), ESCD grants from the Institute de Recherche et de Développement, the European Commission Research Directorate-General (NAOC, contract no. EVG1-CT-2000-00034) and Aérospatiale (Cannes, France). We thank the Polder project at CNES for providing POLDER level 1 and 2 data. We thank Dr. M. Crépon, Dr. Y. Dandonneau and Dr. C. Mejia (Laboratoire d’Océanographie DYnamique et de Climatologie, Université Pierre et Marie Curie, Paris, France) and Dr. D. Cornford of Aston University for stimulating discussions.

Appendix A. Self-Organizing Maps (SOM)

Self-Organizing Maps (SOM) are neural models, which were first introduced by Kohonen (1984), for visualising and clustering \( n \)-dimensional observations. SOM models have two-layers: the input layer and the topological map \( C \) (Fig. 2). The number of neurons of the input layer is equal to the dimension of the observation \( (n = 6 \text{ in the present paper}) \). \( C \) is a discrete lattice of neurons, in the present case, we use a regular two-dimensional grid \( (20 \times 20 \text{ neurons}) \). Each neuron \( i \) of the map is fully connected to the input layer by its weight vector whose dimension is thus the dimension of the observation. So a neuron \( i \) of \( C \) is characterised by its output value and its weight vector, which are respectively a real and an \( n \)-dimensional reference vector denoted \( \mathbf{r}_{v_i} \). For each pair of neurons \( (i,j) \text{ on } C \), we can define a distance \( d(i,j) \) as the length of the shortest path on the lattice between \( i \) and \( j \). For each neuron \( i \), this distance allows us to define a neighbourhood of order \( d \text{ (being a real):} \)

\[
V_i(d) = \{ j \in C | d(i,j) \leq d \} \tag{5}
\]

The aim of the learning phase is to construct a system of reference vectors which represents a partition of \( \mathcal{D} \) in such a way that they satisfy a conservation topological property on \( C \) where two “\( \delta \) similar” neurons \( i \) and \( z \) of \( C \) are associated with two “Euclidean-similar” vectors \( \mathbf{r}_i \) and \( \mathbf{r}_z \) of \( \mathcal{D} \). On the other hand, the reference vectors \( \{ \mathbf{r}_j, j \in C \} \) allow us to define an affectation function \( \chi \) from \( \mathcal{D} \) to \( C \) as follows:

\[
\forall \mathbf{x} \in \mathcal{D} \quad \chi(\mathbf{x}) = \arg \min_j \| \mathbf{x} - \mathbf{r}_j \| \tag{6}
\]

where \( \| \| \) is the Euclidean norm in \( \mathcal{D} \). So this affectation function defines a partition of \( \mathcal{D} \) in subsets \( \{ P_i, i \in C \} \),

where \( P_i = \{ \mathbf{x} \in \mathcal{D}, \chi(\mathbf{x}) = i \} \).

The SOM learning algorithm can be summarised as follows:

1. Initialisation:

   (a) For each neuron \( j \), give random values to the initial reference vectors \( \{ \mathbf{r}_j \in C \} \).
   
   (b) Define the initial neighbourhood order \( d_0 \).
   
   (c) Define the initial learning rate \( h (0) \).

2. At time step \( t \), choose at random an observation \( \mathbf{x} \) from \( \mathcal{D} \).
3. Find the best-matching neuron at time step \( t \) by using the minimum distance Euclidean criterion:

\[
\chi(x(t)) = \arg \min_j \| x(t) - rv_j \| 
\]

with \( j=1,\ldots,400 \) as we use a two-dimensional \( 20 \times 20 \) topological map.

4. Applying a topological constrain to the best matching neuron: adjust the weight vectors of the neurons which belong to that neighborhood by using the updated formula:

\[
rv_j(t+1) = rv_j(t) + h(t)(x(t) - rv_j(t)) \text{ for } j \in V_i(d_i) 
\]

where \( h(t) \) is the learning rate parameter, and let the weight vectors of the other neurons unchanged

\[
rv_j(t+1) = rv_j(t) \text{ for } j \notin V_i(d_i) 
\]

5. Update the neighbourhood order \( d_i+1 \) and the learning parameter \( h(t+1) \). These two parameters decrease for each step.

6. Continue with step 2 until convergence of the parameters.

At the end of the calibration phase, SOM provides a partition of the observations together with a topological order. A given observation \( x \) is assigned to a unique neuron \( j \) (this yielding the smallest \( \chi \)), the topological order on the lattice represents the existing order between the different subsets of the partitions. In SOM formalism, each neuron \( j \) of the map is associated with its reference vector \( rv_j \) and, as mentioned before, represents a subset of observations. As the reference vectors \( rv_j \) and the observations \( x \) have the same dimension, the set of neurons can be viewed as a set of prototype observations, which sums up the observations under study.

### Appendix B. Probabilistic Self-Organizing Maps (PRSOM)

In the present study, we use an advanced SOM model, the Probabilistic Self Organizing Map (PRSOM) (Anouar, Badran, & Thiria, 1997), which deals with a probabilistic formalism. This algorithm approximates the probability distribution of the data with a mixture of normal distributions. In the PRSOM formalism, each output of a neuron of the probabilistic map is associated with a spherical Gaussian density function defined by its mean and its covariance matrix. As in the SOM formalism, each neuron represents a group of observations, but this group is now summarised by the Gaussian function which represents a synthetic reflectance spectrum (the reference vector \( rv_j \)) and a measurement of its dispersion (the variance–covariance matrix).

As the standard Self Organizing Map (SOM) (Kohonen, 1984), PRSOM consists of a discrete set \( \mathcal{C} \) of formal neurons. Its map has a discrete topology defined by an undirected graph; usually, it is a regular grid in one or two dimensions. For each pair of neurons \((c, r)\) on the map, the distance \( \delta(c, r) \) is defined as being the shortest path between \( c \) and \( r \) on the graph. For each neuron \( c \), this discrete distance allows to define a neighborhood of order \( d \):

\[ V_i(d)=\{r \in \mathcal{C} : \delta(c, r) \leq d \} \]

In the following, we introduce a Kernel positive function \( K(\frac{1}{T}K(x)=0) \) and its associated family \( K_T \) parameterized by \( T \):

\[
K_T(d) = \frac{1}{T}K(d/T)
\]

The parameter \( T \) allows to control the neighborhood size.

Let \( \mathcal{D} \) be the data space \( \mathcal{D} \subset \mathbb{R}^n \) and \( \mathcal{D} = \{ \mathbf{z}_i \mid i=1,\ldots,N \} \) the learning data set \( \mathcal{D} \subset \mathcal{D} \). As the standard SOM algorithm, PRSOM defines a mapping from \( \mathcal{C} \) to \( \mathcal{D} \), where a neuron \( c \) is associated with its reference vector \( rv_c \) in \( \mathcal{D} \). At the end of the learning algorithm, two neighbour neurons on the map have close reference vectors in the Euclidean space \( \mathbb{R}^n \).

In contrast to SOM, PRSOM is a probabilistic model which associates to each neuron \( c \) a spherical Gaussian density function \( f_c \). This density function is defined by its mean (reference vector), which is a spherical Gaussian distribution of the data with a mixture of normal densities. In this probabilistic formalism, the classical map \( \mathcal{C} \) is duplicated into two similar maps \( \mathcal{C}_1 \) and \( \mathcal{C}_2 \) provided with the same topology as \( \mathcal{C} \). It is assumed that the model is a folded Markov chain (Luttrel, 1994), so for every input data \( \mathbf{z} \in \mathcal{D} \) and every pair of neurons \((c_1, c_2) \in \mathcal{C}_1 \times \mathcal{C}_2\):

\[
p(c_2|\mathbf{z},c_1) = p(c_2|c_1) \text{ and } p(\mathbf{z}|c_1, c_2) = p(\mathbf{z}|c_1)
\]

It is thus possible to compute the probability of any pattern \( \mathbf{z} \)

\[
p(\mathbf{z}) = \sum_{c_2} p(c_2)p_{c_2}(\mathbf{z})\text{, where }
\]

\[
p_{c_2}(\mathbf{z}) = p(\mathbf{z}|c_2) = \sum_{c_1} p(c_1|c_2)p(\mathbf{z}|c_1)
\]

The probability density \( p_{c_2}(\mathbf{z}) \) is completely defined from the network, giving the conditional probability \( p(c_1|c_2) \) on the map and the conditional probability \( p(\mathbf{z}|c_1) \) on the data. As we mentioned before, \( p(\mathbf{z}|c_1) \) is a normal distribution: \( p(\mathbf{z}|c_1) = f_{\sigma_i}(\mathbf{z}, \mathbf{W}_{c_1}, \sigma_i) \). If we assume that:

\[
p(c_1|c_2) = [1/T_c]K_T(\delta(c_1,c_2)),
\]

where

\[
T_c = \sum_r K_T(\delta(c_2,r)).
\]
The a posteriori density (Eq. (13)) can be expressed with respect to the normal distributions of the neurons:

\[ p_{c_2}(z) = \frac{1}{T_{c_2}} \sum_{r \in C_2} K_T(\delta(c_2, r)) f_r(z, \mathbf{rv}_r, \sigma_r) \]  

Eq. (12) shows that the data density \( p(z) \) is a mixture of local mixtures of normal densities \( f_r(z, \mathbf{rv}_r, \sigma_r) \), whose mean vector and standard deviation are the parameters of the PRSOM model and have to be estimated from the learning data set \( D \) using the learning algorithm of PRSOM. The learning algorithm of PRSOM maximizes the likelihood of the learning set \( D \). It is a dynamic cluster method operating in two steps which are run iteratively until convergence:

- **The assignment step** assigns each observation \( z \) to one cell \( c_2 \) using the assignment function \( \chi \) (Eq. (16)). This step gives rise to a partition of the data space \( \mathbb{D} \), each observation \( z_i \) being assigned to the most likely cell according to the density \( p_{c_2} \):

\[ \chi(z) = \arg \max_{c_2} p_{c_2}(z) \]  

- **The minimization step** assumes that the observations of the learning set \( D \) are independent and maximize the conditional likelihood of the training data set \( D \) with respect to the parameters \((\mathbf{rv}, \sigma)\) and under the hypothesis that each \( z_i \) is generated by the \( p(\mathbf{rv}/x) \) distribution.

\[ p(z_1, z_2, \ldots, z_N/\mathbf{rv}, \sigma, \chi) = \prod_{i=1}^{N} p_{\chi(z_i)}(z_i) \]  

The minimization step minimizes the conditional log-likelihood function \( E(\mathbf{rv}, \sigma/\chi) \) according to \( \mathbf{rv}, \sigma \) and \( \chi \):

\[ E(\mathbf{rv}, \sigma/\chi) = \sum_{i=1}^{N} - \ln \left( \sum_{c \in C} K_T(\delta(\chi(z_i)), r)) f_r(z, \mathbf{rv}_r, \sigma_r) \right) \]  

The parameters \( \mathbf{rv} \) and \( \sigma \) are updated by setting the derivatives of \( E(\mathbf{rv}^k, \sigma^k/\chi^k) \) to zero (\( \chi^k, \mathbf{rv}^k \) and \( \sigma^k \) denote the parameters at iteration \( k \)). To solve this equation, we use, as in (Duda & Hart, 1973), an iterative approximation which assumes that, for the \( k \)th iteration, the initial estimates of the parameters are sufficiently close to the true values. As in the Kohonen algorithm, PR-SOM makes use of a neighborhood system the size of which, controlled by \( T \), decreases as the learning proceeds. At the end of the learning phase, the map provides the topological order; the partition associated with the map is defined by the latest assignment function \( \chi^{\text{new}} \) as defined in Eq. (16) and is based on probability considerations. Therefore, the partitions provided by PRSOM are different from those provided by SOM, which use Euclidean distance. The estimation of the density function gives an extra information which will be of interest later for classification purpose. At the end of the learning phase, \( \mathbb{D} \) is divided in \( M \) subsets: each neuron \( c \) of the map represents a particular subset \( P_c = \{ z/C_{15}^{\text{new}}(z) = c \} \).

**Appendix C**

At that stage, the map is constituted by neurons, some of them being labeled. The eHC-PRSOM method allows to generalize the classification by providing a label for each neuron. The eHC-PRSOM method operates on the partition provided by the topological map (Yacoub et al., 2000). This method is based on a statistical automatic classification (Hierarchical Clustering HC) (Jain & Dubes, 1988). This labeling process can be summarised as follows:

1. Assign each reference vector \( \mathbf{rv} \) to a unique group of observations, some of them having a label provided by the majority vote rule: \( n_{rv} \) is the number of observations of \( D \) which are in the group \( J \) represented by \( \mathbf{rv}_j \).

2. Agglomerate the two closest clusters (groups in the initial phase) according to a distance criteria. When dealing with PRSOM, and since PRSOM gives an estimate of the covariance matrix of the initial groups (see Appendix 2), this estimate is used in order to compute the Ward Index which agglomerates the groups according to their variances.

3. Assign, if possible, a label to the cluster:
   - two groups with the same label agglomerate in a cluster with the same label.
   - a cluster without a label takes the label of a labeled one.
   - two clusters without a label agglomerate in a cluster without label.
   - two clusters with different labels do not agglomerate. Each one keeps its label.

3.1. return to step 2 until all \( \mathbf{rv} \)s have received a label.

At the end of eHC-PRSOM, each \( \mathbf{rv} \) of \( C \) is labeled. The set of reference vector \( \mathbf{rv}_k \) denoted \( C_k \), which have the same label \( L_k \) (i.e. Sea, Aerosols 1, Aerosols 2 or Clouds), corresponds to different density functions which approximate the probability density function of the class \( L_k \). Since PRSOM is a probabilistic model, it allows the computation of the posterior probability \( p(\mathbf{rv}/\mathbf{x}) \) for each observation \( \mathbf{x} \) to belong to a given \( \mathbf{rv} \); and so the posterior probability \( p(L_k/\mathbf{x}) \) of an observation \( \mathbf{x} \) to belong to a class \( L_k \).

If we denote \( N \) the number of observations \( \mathbf{x} \) in the learning set \( D \) and \( n_{rv} \) the number of observations in the cluster assigned to \( \mathbf{rv} \), the a priori probability \( p(\mathbf{rv}) \) of the reference vector \( \mathbf{rv} \) can be estimated as \( n_{rv}/N \). The Bayes rule
allows to compute the a posteriori probability $p(\mathbf{rv}/x)$ which is the probability for a known observation to be assigned to a reference vector $\mathbf{rv}$:

$$p(\mathbf{rv}/x) = \frac{p(\mathbf{rv})p(\mathbf{x}|\mathbf{rv})}{p(\mathbf{x})} = \sum_{j} \frac{n_{\mathbf{rv}j}p_{\mathbf{rv}}(x)}{\sum_{\mathbf{rv}} n_{\mathbf{rv}j}p_{\mathbf{rv}}(x)} \quad (19)$$

where the probability density $p_{\mathbf{rv}}(x)$ is calculated during the PRSOM calibration phase. The sum is taken over by the whole set of reference vectors for the denominator.

As a class is composed by a set of neurons, the posterior probability $p(L_k/x)$ is calculated by a sum over the neurons with label $L_k$.

$$p(L_k/x) = \sum_{\mathbf{rv} \in C_k} p(\mathbf{rv}/x) = \sum_{\mathbf{rv} \in C_k} \frac{n_{\mathbf{rv}j}p_{\mathbf{rv}}(x)}{\sum_{\mathbf{rv}} n_{\mathbf{rv}j}p_{\mathbf{rv}}(x)} \quad (20)$$

In our experiment, $k=1, \ldots, 4$, so four posterior probabilities $p(L_k/x)$ are computed and a pixel is assigned to the label with the highest probability.

References


