SVME: an ensemble of support vector machines for detecting oil spills from full resolution MODIS images

Marco Cococcioni · Linda Corucci · Andrea Masini · Fabio Nardelli

Abstract This paper addresses oil spill detection from remotely sensed optical images. In particular, it focuses on the automatic classification of regions of interest (ROIs) in two classes, namely oil spills or look-alikes. Candidate regions and the corresponding boundaries have been manually identified from full resolution Moderate Resolution Imaging Spectroradiometer images, related to the Mediterranean Sea over the years 2008 and 2009. Then, a set of features has been extracted from each ROI, allowing to formulate the oil spill detection problem as a two-class classification task on the provided regions (i.e. using a supervised learning strategy). Since ROI classification is challenging, some desired characteristics for the classification algorithm are first identified, such as accuracy, robustness, etc. Then, a solution (called SVME) is provided: it is based on an ensemble of incremental/decremental cost-oriented Support Vector Machines, aggregated with the Receiving Operating Characteristic (ROC) convex hull method in the ROC space. Such a solution addresses all the desired characteristics. Finally, the results obtained on the collected dataset are shown. The importance of this study is the devising of a powerful classification technique that may have an impact on optical oil spill detection from space, especially if fused with satellite synthetic aperture radar data. Moreover, it is shown how the proposed system can be used as a decision support tool, to help a junior operator in making more reliable detections.

M. Cococcioni (✉) · L. Corucci
Dipartimento di Ingegneria dell'Informazione: Elettronica, Informatica, Telecomunicazioni, University of Pisa,
Largo Lucio Lazzarino 1,
56123 Pisa, Italy
e-mail: m.cococcioni@iet.unipi.it

A. Masini · F. Nardelli
Flyby S.r.l.,
Via Puini 97,
57128 Livorno, Italy

1 Introduction

Today, 90% of oil and refined products are transported by sea. Out of the 1.5 to 1.8 billion tonnes of crude oil transported worldwide yearly, Europe is the main recipient, with nearly 500 million tonnes of crude oil and 250 to 300 million tonnes of refined products per year (Oceana 2003). Moreover, many oil tankers transport their cargo to other destinations through European waters, meaning that the total amount of crude oil passing through European waters could be over 1 billion tonnes. Accidents resulting in massive and devastating oil spills, announced by the mass media, affect public opinion and mobilise policymakers. However, such dramatic accidents occur only occasionally and represent a small fraction of the pollution events at sea. Oil released into European seas as a result of operational discharges greatly exceeds the amount released during accidental spills. As an example, yearly, in the Mediterranean, these operations can add up to nearly 20 times the amount that was spilled by the

'Prestige' off the northern Spanish coasts in 2002 (Kluser et al. 2006).

Due to the lack of adequate waste reception facilities in many ports, added to poor surveillance, every year, at a lower cost, oil tanker crews release million of tonnes of oil from oily ballast waters and tank-washing residues in our seas. Another part is due to machinery space bilge and fuel oil sludge, the latter constituting more than 50% of the total operational discharges.

More precisely, illegal dumping and routine operations of vessels account for between 666,000 and 2.5 million tonnes of hydrocarbons of marine pollution per year. Furthermore, it is estimated that at least 3,000 major illegal hydrocarbon-dumping incidents take place in European waters yearly. The Mediterranean is the sea that is most affected by this type of dumping, amounting to more than 400,000 tonnes of released oil per year (Oceana 2003). Even if the estimates of the Joint Group of Experts on the Scientific Aspects of Marine Environmental Protection (GESAMP) are lower (the GESAMP is an inter-agency advisory body of the United Nations), their latest reports (GESAMP 2007, 2009) confirm the severity of oil pollution at sea due to ships.

It must be observed that not all oil discharges at sea are illegal: the International Convention for the Prevention of Pollution from Ships (MARPOL 73/78) only limits them in 'Special Areas', but the entire Mediterranean Sea is one of them. In special areas like the Mediterranean Sea, the MARPOL 73/78 convention allows for only a small percentage of oil (15 parts per million) to be released in the sea. During the years, however, such constraint has been frequently overlooked. This solicited the adoption of 2005/35/EC Directive (amended by the 2009/123/EC Directive) by the European Parliament on 12 July 2005, to ensure that persons responsible for illicit discharges are subject to adequate penalties.

The European Maritime Safety Agency (EMSA) has been tasked to support the Member States and the European Commission for the implementation of the directive. For this purpose, since 2007, it has started the CleanSeaNet service, which is intended to provide a near real-time satellite-based oil spill and vessel monitoring service. It entered into operation on 16 April 2007. Despite this and other services currently available, which are mainly based on satellite Synthetic Aperture Radar (SAR) data, further research is required for looking at complementary or alternative solutions, in order to strengthen the available ones and/or reduce their costs. In particular, the exploitation of satellite images for oil spill detection, which is the topic of this work, allows monitoring large areas in an economical and easy way, thus offering many advantages in cost and time-saving terms. Using another source for oil spill detection could also help in reducing the number of false alarms (i.e. oil spill alarms that reveal to be natural phenomena after in situ verification) generated by current SAR-based services.

1.1 Organisation of the paper

The paper is organised as follows. Section 2 reviews the state-of-the-art of oil spill detection from SAR and optical satellite data. Section 3 shows the proposed approach to oil spill detection from remotely sensed optical images, the collected dataset and the extracted features. Section 4 describes the classification algorithm, where the classifier is trained over the collected dataset. In particular, a powerful framework for two-class classification [the Receiving Operating Characteristic (ROC) space and the ROC convex hull] is first shown, then some desired characteristics for the classification algorithm are defined, and finally a powerful classifier, based on an ensemble of Support Vector Machines (SVMs), is built in three steps. Section 5 provides the results of two experiments which demonstrate the effectiveness of the proposed approach. Section 6 summarises the whole methodology and discusses the possible alternative uses of the proposed system, while Section 7 is devoted to conclusions and future work.

2 Satellite sensors for oil spill detection

Many different types of sensor are able to detect oil spills at sea, either mounted on satellites or aircrafts, like the synthetic aperture radar, the microwave radiometer, the ultraviolet radiometer, visible and near infrared radiometers, etc. Each sensor exploits different properties of oil, water and their interaction. In Brekke and Solberg (2005), a review of the literature in oil spill detection based on different sensors is provided, along with the associated strengths and weaknesses.

However, few of such sensors are suitable for monitoring large areas, like the entire Mediterranean Sea, in a cost-effective way. In this paper, only sensors onboard satellites have been considered, since collecting data from aircrafts would be far more expensive. Satellite SAR-based studies and services are very popular in oil spill detection and are reviewed in the next subsection. Then, the state-of-the-art using the optical sensors (especially visible and near infrared) is provided as well.

2.1 Oil spill detection using SAR: state-of-the-art

SAR, as any radar-based instrument, is able to detect oil spills due to the difference between oil and water in wave-damping capacity. A radar image of the sea surface is formed by reconstructing the reflection of active radar signals from small gravity and capillary waves. The presence of an oil slick, due to oil viscosity, suppresses capillary ripples, so that the return signal from an oil-covered surface is less intense. Thus, oil slicks appear as dark objects in

SAR images. However, calm sea regions appear as dark objects too and represent a source of error for oil spill detection. Sea state is another limiting factor for SAR oil spill detection. Indeed, when the wind speed is below 1.5 m/s, the wind stress is insufficient to generate ripples, while above 12 m/s, the oil film is disrupted and washed down by breaking waves (Alpers and Huhnerfuss 1989). More precisely, the optimal wind velocity range for oil spill detection has been estimated to be between 1.5 and 6 m/s (Liu et al. 2000). However, this range must be handled with care, since it also depends on the frequency of the SAR sensor used.

Since SAR data have been widely exploited for oil spill detection, many contributions can be found in the literature regarding the study of oil slicks in SAR images, and several efforts have been made in order to develop automatic detection systems. Images from remote sensing instruments, in particular SAR images, are usually combined with information from trained and experienced human observers. Visual observation relies on many of the same physical mechanisms which are used by various remote sensing instruments and which give rise to the perception of colour (wavelength distribution), brightness (signal intensity) differences between oil and water, variations in surface roughness (wave damping), and finally, the interpretation of spatial patterns and the analysis of surrounding elements (e.g. proximity to land, presence of other similar objects in the scene, etc.). Thus, a fully automated or semi-automated system for oil spill detection should actually resemble the expert's decisions based on similar criteria, knowledge and rules, and this has been subject of several studies reported in literature.

Kubat et al. (1998) proposed a neural network for the classification of dark regions detected in SAR images, to be used as a training set of the system. In their paper, the authors also analyse in detail the issues related to the application of machine learning techniques to oil spill detection. Del Frate et al. (2000) also used a neural network for semi-automatic detection of oil spills in SAR images, building an input vector for the network, containing a set of features which characterise oil spill candidates. Later, Topouzelis et al. (2004) compared the performance obtained using multi-layer perceptron and radial basis function neural networks. In particular, the authors used original reconstructed SAR images as input for the classification system, instead of using a vector of features extracted from dark regions in the images. Solberg et al. (1999) developed a semi-automated classifier for oil spill detection, based on statistical modelling combined with a rule-based approach. They identified 11 different object features (Solberg and Solberg 1996) and used them to build a classification procedure based on Bayesian inference, where three different categories of real oil spill probability (low, medium and high) were

recognised. Fiscella et al. (2000) developed a stochastic classifier based on Mahalanobis statistical tests and classical compound probabilities.

Fuzzy logic systems, able to better resemble a human expert's decision, have also been used by Keramitsoglou et al. (2006) to develop a fully automated system for oil spill identification in SAR images. Finally, it must be observed that the interesting conclusions provided in Brekke and Solberg (2005), concerning the use of satellite SAR and the related limitations, are still valid today. As regards to SAR-based services, a recent review of them can be found in Ferraro et al. (2010), in addition to the CleanSeaNet service run by EMSA mentioned above.

2.2 Oil spill detection using optical satellite images

Although SAR-based solutions are really powerful for oil spill detection from space, mainly due to their all-weather and all-day detection capabilities, they have well known limitations. For instance, the difficulty in detecting oil spills for high wind speeds (>6 m/s), which exposes to the possibility of missing true oil spills, or, when the wind speed is low (<1.5 m/s), to that of generating too many false alarms, i.e. alerts indicating supposed oil spills which could reveal to be look-alikes, after in situ verification. Since the latter is very expensive, there is a need for keeping the number of false alarms as small as possible, while trying not to miss any true oil spill case.

Thus, there is a need for alternative or complementary solutions. This suggests the possibility of exploiting optical data, which, up to now, have been little used for oil spill detection applications. The development of a new approach, based on optical data, could be used either on its own or as a support to SAR-based solutions, in order to meet the need of environmental protection authorities for efficient and cost-effective monitoring tools. Indeed, with the exclusion of hyper-spectral data, which are very expensive, optical satellite images are generally cheaper (sometimes freely available) and provide more frequent (daily) information if compared to SAR images, allowing for large area monitoring and remote zone control.

The reason why optical satellite images have been less commonly used in the field of operational oil spill detection is that good weather and daylight are mandatory conditions to perform a correct detection, while these conditions are not required when using SAR data. However, even the use of SAR data implies many other limiting factors, such as cost, low revisit time and sea state (in particular, wind speed).

The use of optical data can improve operational monitoring for many reasons. The key point is that oil spill identification by optical sensors and by SAR sensors relies on different mechanisms that are respectively controlled by differences in oil and water optical properties (Appendix 1),

which give rise to oil–water contrast (Appendix 2), and by differences in oil and water wave-damping capacity. Thus, oil spill detection systems based on optical data are not affected by the same false alarm types (such as very calm sea areas) as SAR-based systems. Moreover, while microwave, infrared and ultraviolet radiation are strongly absorbed by water, visible radiation can penetrate water to a depth of many metres, especially in the blue. Hence, optical sensors operating in the visible range of the spectrum can potentially detect oil not only on the surface, but also dispersed in the water column.

The possibility of detecting oil spills from optical data has also been demonstrated in Hu et al. (2003), where the authors use optical images, captured by NASA's Moderate Resolution Imaging Spectroradiometer (MODIS, see Table 1), to access the magnitude, area covered and duration of a big oil spill which occurred in Lake Maracaibo, Venezuela, caused by several accidents related to the oil industry operating in the lake's area. More recently, empirical relationships based on MODIS higher resolution bands (250 m) have been used to detect oil spills in the Arabian Sea (Alawadi et al. 2008). MODIS images have also been used to detect natural crude oil slicks and consequently estimate the annual seepage rates in the northwestern Gulf of Mexico (Hu et al. 2009).

These studies prove that oil spill identification from optical images, in particular those acquired by MODIS, is possible, and it can be used for operational monitoring. Beside this, the interpretation of the remotely sensed multi-spectral signal produced by an oil slick on the sea surface appears to be quite problematic because many factors contribute to oil–water contrast, such as oil type, oil absorption properties, oil thickness and water constituent concentrations (see Appendix 2). These factors often give conflicting contributions, which vary through the visible and near infrared spectral range. An additional issue is represented by the difficulty in performing an accurate atmospheric correction, in order to reduce path radiance contribution, and by the presence of sunglint. These factors make spectral analysis a difficult task.

The approach to oil spill detection which has been chosen in the present work consists in employing full resolution MODIS optical data and using a powerful and flexible

Table 1 MODIS specifications

Orbit	Altitude 705 km, 10:30 a.m. descending node (TERRA) or 1:30 p.m. ascending node (AQUA), sun-synchronous, near-polar, circular
Scan rate	20.3 rpm, cross track
Swath dimensions	2,330 km (cross track) by 10 km (along track at nadir)
Quantization	12 bits

classification algorithm, avoiding any spectral analysis. The identification of oil spills is then mainly based on geometrical features, together with differences in the contrast. The use of a supervised classification algorithm allows overcoming difficulties related to contrast interpretation and atmospheric correction, which are typical of an analytic approach.

3 The adopted approach

In this paper, oil spill detection is tackled as a classification problem. A numerical vector (called *feature vector*) is extracted from each Regions of Interest (ROI), and it is fed to a classification algorithm. The classification itself is described in the next section, while in the following subsections, the collected dataset and the feature vector are described.

It is important to observe that the rationale behind this way of work (collect a meaningful dataset of oil spills and look-alikes and train a supervised learning algorithm to distinguish them) is to verify whether or not an automatic classification software can mimic the behaviour of a skilled operator. If the performance of an automatic system is satisfactory enough, then, in the future, both an automatic oil spill alert generation system could be developed for optical images and a decision support system for aiding a junior operator in making more reliable decisions (these aspects will be further elaborated in Section 6).

3.1 The collected dataset

The supervised learning procedure that will be described in Section 4 requires a dataset of input and target elements which serves partly for training the classifier, partly for validation and partly for testing the classification performance. Dataset collection represents a relevant issue in supervised classification. Training set size must be appropriate to the number of unknown parameters in the classifier structure. In addition, elements in the dataset must compose a representative set of the data the trained classifier will process.

For the present study, a considerable dataset of regions of interest has been built by collecting a number of optical images taken during the years 2008 and 2009, over the entire area of the Mediterranean Sea. In particular, full resolution level L1B data from MODIS sensor have been used, having a spatial resolution of 250 m. Level L1B data consist in calibrated radiance or reflectance values, which are freely available from NASA (http://ladsweb.nascom.nasa.gov/data/). The MODIS instrument is onboard TERRA and AQUA satellites. TERRA's orbit around the Earth is timed so that it passes from north to south across the equator

in the morning, while AQUA passes south to north over the equator in the afternoon (see Table 1). Using data from both MODIS-TERRA and MODIS-AQUA sensors, up to four images per day are available, each one partially covering the Mediterranean Sea.

As Table 2 shows, only bands 1 and 2, respectively, centred in the visible at 0.65 μm and in the near infrared at 0.85 μm, are available from MODIS at the highest resolution, so that any spectral analysis is precluded. On the other hand, it is important to have a spatial resolution as good as possible, since most slicks produced by illegal discharge from moving ships could not be detected in lower resolution images (e.g. 1 km) due to their small area.

Only images acquired in clear sky conditions have been selected when building the dataset. After georeferencing, a contrast enhancement based on histogram equalisation has been applied in order to improve the visualisation. Then, the boundaries of ROIs have been drawn by hand.

Eleven ROIs were real oil spills verified in situ by the Italian Coast Guard (dataset provided by the Italian Ministry of Environment). Another 11 ROIs have been labelled as oil spill after verifying the presence of a similar dark patch on satellite SAR images (Envisat-ASAR image at 150 m spatial resolution), acquired within a reasonable time frame. The remaining 135 ROIs have been labelled as oil spill events according to photo-interpretation only, on the basis of the contrast between regions and surrounding areas (which have been supposed to be clean waters). Similarly, look-alike

labels have been assigned by photo-interpretation and, when available, by checking a related SAR image. Typical look-alike events are due to algal blooms or effects due to current patterns. Figure 1 shows an interesting case of oil spill event visible both on MODIS and Envisat-ASAR images.

In particular, the photo-interpretation analysis has been conducted examining the scene where the candidate oil spill or look-alike had been detected, that is, considering the context, the location and the possible presence of other elements in the surrounding area. This allowed distinguishing between linear slicks, which might be caused by a moving ship releasing oil, like the one shown in Fig. 2a–c, and effects related to current patterns, usually occurring in particular spatial patterns, such as those shown in Fig. 2d–f.

In the end, the number of detected oil slicks is approximately equal to the number of look-alikes, for a total of 316 ROIs. A search for outliers has also been performed according to the distribution of the features characterising oil spills described in the next subsection. After this analysis, 12 ROIs have been removed, obtaining a dataset composed of 304 elements, 157 oil spills and 147 look-alikes (see Table 3).

The outliers found resulted to be represented by big nonlinear-shaped slicks and dark regions in a particularly complex background, which have been manually discarded from the dataset and not further considered in this application. The reason for this choice is motivated by the fact that the classification system proposed in this work is intended for the identification of illegal oil discharge by moving ships or of oil released during tank cleaning in the sea. Thus, since training elements should resemble data that the classifier will be asked to process in the finished application, only these types of slicks have to be well represented in the dataset. Very big slicks, produced for instance by oil tanker accidents, which can cause environmental disasters to happen, have not been handled, since the appropriate authorities are usually informed about these events.

Table 2 MODIS 1–19 bands and associated spatial resolution

Primary use	Band	Bandwidth (nm)	Spatial resolution (m)
Land/cloud/aerosols boundaries	1	620–670	250
	2	841–876	
Land/cloud/aerosols properties	3	459–479	500
	4	545–565	
	5	1,230–1,250	
	6	1,628–1,652	
	7	2,105–2,155	
Ocean colour/phytoplankton/ biogeochemistry	8	405–420	1,000
	9	438–448	
	10	483–493	
	11	526–536	
	12	546–556	
	13	662–672	
	14	673–683	
	15	743–753	
	16	862–877	
Atmospheric water vapour	17	890–920	
	18	931–941	
	19	915–965	

3.2 The extracted features

In order to discriminate between oil spills and look-alikes, a number of physical and geometrical features characterising the object to be classified have been exploited. Following the results of SAR oil spill detection (Brekke and Solberg 2005; Solberg et al. 1999; Del Frate et al. 2000), a set of grey level features, characterising the differences between the object and the surrounding area, and a set of geometrical features, describing shape and extension, have been used:

1. **Geometrical features**
   - Area of the object (A) expressed in square kilometer
   - Perimeter (P) expressed in kilometer

Fig. 1 An oil spill viewed both by optical (**a**) and SAR (**b**) instruments on 14 July 2008 offshore south-west Peloponnese, Greece (MODIS and Envisat-ASAR, respectively). The two patches slightly differ due to the effect of currents, since they have been acquired at different times
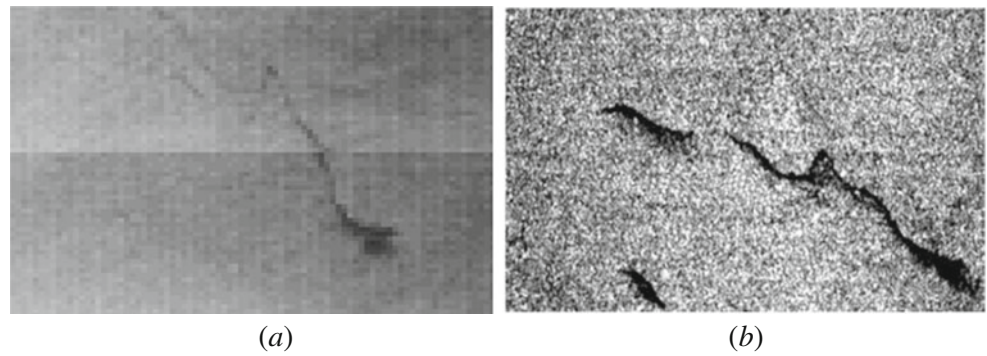
– Complexity (*C*): defined as

$$C = \frac{P}{2\sqrt{\pi A}}.$$

This feature generally assumes small numerical values for regions with simple geometry and larger values for regions with complex geometry.

– *Spreading* (*S*) was computed performing a principal component analysis on the vector whose components

are the coordinates of the pixels belonging to the object:

$$S = \frac{100\lambda_2}{\lambda_1 + \lambda_2},$$

where λ_1 and λ_2 are the two eigenvalues associated with the covariance matrix ($\lambda_1 > \lambda_2$). This feature assumes low values for long and thin objects and high values for objects closer to a circular shape.

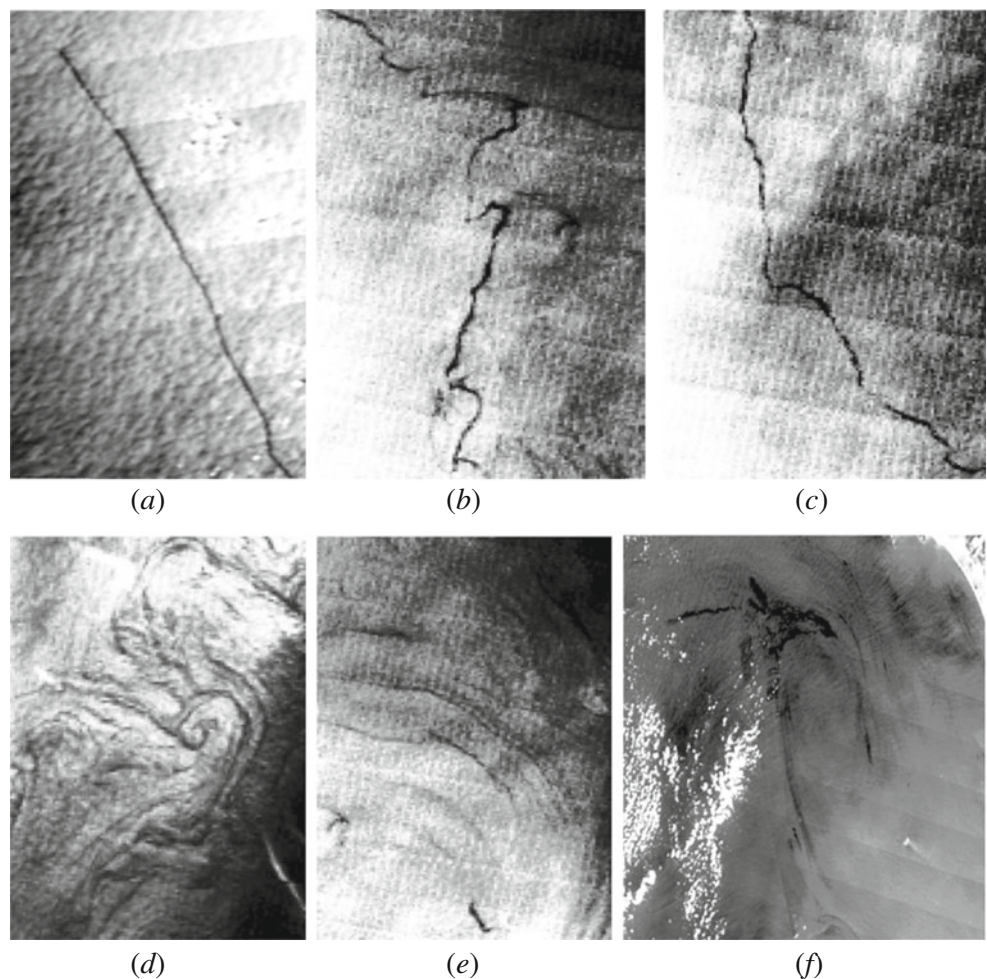**Fig. 2** Examples of oil spills (**a**, **b** and **c**) and look-alikes (**d**, **e**, **f**) from the dataset

Table 3 The collected dataset

MODIS images	46
Regions of interest (ROIs)	304
Oil spills (OSs)	157[a]
Look-alikes (LAs)	147

[a] Among the 157 oil spills, 11 (7% of total) have been verified in situ by the Italian Ministry of Environment and additional 11 were also detected in a related SAR image. The remaining 135 were labelled by the authors by using photo-interpretation only

2. **Grey level features**

– *Object standard deviation* is the standard deviation of the intensity (in radiance or reflectance), values of pixels belonging to the object.
– *Max contrast* is difference between the background mean intensity value and the lowest intensity value inside the object.
– *Mean contrast* is difference between the background mean intensity value and the object mean intensity value.

Spectral features have not been used, since the two bands available from MODIS full resolution images are insufficient for a spectral analysis. Grey level features have been calculated on a two-band ratio (band 2/band 1), since computing the ratio between the two bands enhances the contrast.

Tables 4 and 5 report some statistical parameters of the above-mentioned set of features, computed for the 157 oil spill and 147 look-alike cases in the dataset where outliers had been removed. The tables show that oil spills are generally smaller and have a thinner shape. This is consistent with the choice of considering slicks produced by moving ships.

4 Building a powerful SVM ensemble classifier in three steps

As already said, oil spill detection has been tackled as a classification problem of oil spill candidate regions. Since

Table 4 Statistical parameters of the features calculated for oil spill cases in the dataset

Feature	Min	Max	Mean	Std. dev.
Area (km^2)	0.688	41.7	6.96	6.58
Perimeter (km)	2.91	94.6	20.4	15.9
Complexity	0.229	5.40	2.16	0.770
Std. dev. band ratio	0.00610	0.0519	0.0199	0.00940
Mean contrast band ratio	−0.0588	0.156	0.0505	0.0311
Max contrast band ratio	−0.0255	0.216	0.0933	0.0449
Spreading	0.114	26.8	4.25	5.35

Min minimum, *Max* maximum, *Std. dev.* standard deviation

Table 5 Statistical parameters of the features calculated for look-alike cases in the dataset

Feature	Min	Max	Mean	Std. dev.
Area (km^2)	0.875	73.9	14.4	14.7
Perimeter (km)	1.00	147	26.4	21.2
Complexity	0.0912	5.009	2.016	0.893
Std. dev. band ratio	0.00521	0.0999	0.0283	0.0179
Mean contrast band ratio	−0.00535	0.234	0.0733	0.0519
Max contrast band ratio	0.0198	0.451	0.134	0.0822
Spreading	0.548	44.3	11.9	10.0

Min minimum, *Max* maximum, *Std. dev.* standard deviation

the number of classes is two (*oil spill* and *look-alike*), all the powerful theoretical results and tools specifically devised for two-class problems can be exploited. This allows building a more powerful classifier with a strong theoretical background, with respect to general purpose multi-class classification algorithms. Indeed, the latter classifiers can be used safely only when all the following conditions hold: (a) the classes are balanced, (b) their a priori probabilities are similar and (c) misclassification costs are equal for patterns belonging to any classes. The first condition means that general purpose multi-class classification algorithms, such as multi-layer perceptron and radial basis function neural networks, are unreliable when class distributions are very imbalanced (e.g. one or more of them is overrepresented in the training set). As an example, let us consider the case of a problem having 1,000 instances, 990 of which belonging to the first class (class A) and only ten to the second (class B). Once trained, a neural network classifier will tend to classify all the patterns as belonging to the majority class (class A), apparently reaching very good classification accuracy (99%), while the classifier is not able to detect any class B (minority class) pattern. A similar effect happens when the second condition holds, i.e. when the a priori probabilities are highly imbalanced: in such cases for instance, a Bayesian classifier will tend to classify all the patterns as belonging to the a priori most probable class. The third condition means that classification has to be carefully made when the misclassification costs are class dependent and highly imbalanced. Indeed if misclassifying a pattern belonging to class A as belonging to class B costs several hundred times more than the opposite, the former kind of misclassification should be avoided as much as possible, while the latter can be tolerated much more. When all the three above conditions hold, any general purpose classifier can be used. Otherwise, the classifier needs to be carefully designed in a framework that allows coping with all the conditions. Fortunately, for two-class problems, such a framework already exists.

4.1 A powerful framework for two-class problems

A very important point that differentiates two-class problems from general purpose multi-class problems is that the performance of a classifier can be analysed in the so called ROC space, i.e. a portion of the first quadrant of a two-dimension plane where the false positive rate (*FPR*) is represented on the abscissa and the true positive rate (*TPR*) on the ordinate axes. In the application on hand, the TPR is the ratio between the number of oil spills correctly classified as oil spill and the total number of oil spills, while the FPR is the ratio between the number of look-alikes wrongly classified as oil spills and the total number of look-alikes.

For a given classifier, describing its performance trough the pair (FPR and TPR) is far more meaningful than measuring the accuracy using a mono-dimensional performance metric, such as the percentage of correct classification, as discussed in the previous subsection. When the classifier is a continuous mapping between the F input space (the *feature* space) and the output space (typically, the interval [0,1]), then it is called *continuous*:

$$\Gamma(x) : \Re^F \rightarrow [0, 1].$$

Continuous classifiers are very common: neural networks classifiers, fuzzy classifiers and support vector machines are usually continuous, for instance. To actually perform a binary classification using a continuous classifier, one has to choose a threshold τ on the classifier output. More precisely, the binary classifier can be obtained as follows:

$$\Gamma_\tau(x) : \begin{cases} N & \text{if } \Gamma(x) < \tau \\ Y & \text{if } \Gamma(x) \geq \tau \end{cases}$$

where $0 \leq \tau \leq 1$, N and Y stand for negative and positive classifications, respectively. So doing, the continuous classifier can thus be associated with a set of pairs $(\text{FPR}_{\tau_i}, \text{TPR}_{\tau_i})$, $i=1,\dots, I$, where I is the number of thresholds τ. This allows charting the performance of a continuous classifier in the ROC space by means of a ROC curve, instead of a single point: each point belonging to the curve corresponds to a certain threshold and consequently to a certain binary classifier.

Observe that FPR and TPR can now be more formally defined using probability as (Provost and Fawcett 2001):

$$\text{FPR} = p(Y|\boldsymbol{n})$$

$$\text{TPR} = p(Y|\boldsymbol{p})$$

where $\boldsymbol{p}$ and $\boldsymbol{n}$ are the positive (oil spill) and the negative (look-alike) classes, respectively.

A very important aspect of the ROC analysis is that it allows selecting the optimal threshold, given the prior probabilities and the misclassification costs. This means that the user must provide priors and costs and then the optimal operating point can be automatically determined. This follows from the fact that in two-class problems a mono-dimensional performance index can be defined as:

$$J[\text{FPR}, \text{TPR}] = \text{FPR} \cdot C(Y, \boldsymbol{n}) \cdot p(\boldsymbol{n}) + (1 - \text{TPR}) \cdot C(N, \boldsymbol{p}) \cdot p(\boldsymbol{p}) \tag{1}$$

where $C(Y, \boldsymbol{n})$ and $C(N, \boldsymbol{p})$ are the cost functions associated to the misclassification of, respectively, a negative and a positive pattern, and where $p(\boldsymbol{n})$ and $p(\boldsymbol{p})$ are, respectively, the a priori probabilities for negatives and positives.

Such probabilities are typically estimated through the relative frequencies of positives and negatives in the available dataset. Here, it is worth noting that the cost functions $C(Y, \boldsymbol{n})$ and $C(N, \boldsymbol{p})$ must be provided by the user and that they could change over time.

By using Eq. 1, the optimal threshold $\underline{\tau}$ can be easily found even if priors and misclassification costs are changing over the time:

$$\underline{\tau} = \arg \min_{\tau_i} \{J[\text{FPR}(\tau_i), \text{TPR}(\tau_i)]\}.$$

Doing so, a single optimal binary classifier can be automatically extracted from the associated continuous classifier.

When using the above index, it can be further demonstrated that iso-performance lines are straight lines with slope $[C(Y, \boldsymbol{n}) \cdot p(\boldsymbol{n})]/[C(N, \boldsymbol{p}) \cdot p(\boldsymbol{p})]$ (Provost and Fawcett 2001). From this result, it derives that only the points on the convex hull of the ROC curve can be actually optimal. It is worth noting that the ROC space is insensitive to changes in misclassification costs and a priori probabilities: the ROC curves remain the same even if those parameters change over time.

Another important aspect is that the same theoretical framework can be exploited even when multiple classifiers (i.e. an ensemble of classifiers) are used. There are many reasons for opting for multiple classifiers instead of using just one, like performance and robustness. What is interesting in this scenario is that a single convex hull can be computed for the set of ROC curves, and at each time instant, the optimal classifier (i.e. one of the classifiers within the ensemble and the associated optimal threshold $\underline{\tau}$) can be easily determined (see Fig. 3).

It can be observed now that the (FPR and TPR) pairs constituting the ROC convex hull are all optimal in the Pareto sense because each of them has either a better TPR but a worse FRP or vice versa, with respect to any other pair of the convex hull. This observation allows further appreciating the usefulness of the mono-dimensional index J defined in Eq. 1, which
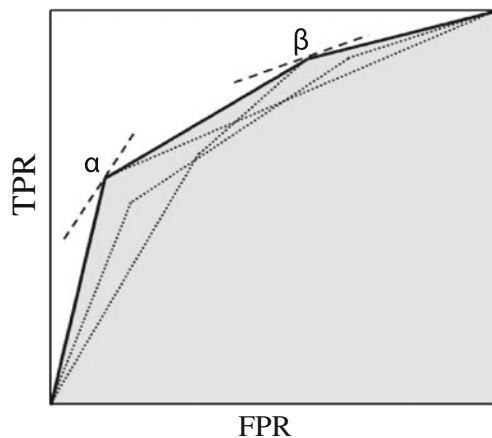
Fig. 3 Convex hull of three ROC curves. *Lines* α and β are two iso-performance lines, both tangent to the convex hull, but with different slopes, thus corresponding to different costs and class distributions

consents the automatic selection of a single optimal pair among the Pareto optimal ones.

Finally, it can be observed that the convex hull of multiple ROC curves associated to an ensemble of continuous classifiers can also be used in an *online* mode, i.e. taking into account the arrival of new ROIs. The following section will now introduce some desired characteristics for building the preferred two-class classifier.

4.2 Defining the desired characteristics for the two-class classifier

Once a dataset of oil spills and look-alikes has been collected, as new satellite images are available and new oil spill candidates are detected, the dataset can be updated, and the classification system can consequently be re-trained, benefiting from a larger dataset. In this framework, an online learning approach allows to easily improve the classification capability of the system. Every time a new satellite image is downloaded and analysed, new candidates detected in the image are included in the dataset and the proposed classification system is able to learn quickly from these new candidates, by means of the online learning process. The described situation well represents an operational scenario where a dataset of oil spills is available, and it is continuously updated (say daily) by the detection of other oil spills in some new images. These considerations suggest the investigation of an online approach to oil spill classification.

Coping with changing a priori probabilities and misclassification costs is crucial in oil spill detection. For instance, if the coast guard verifies that too many look-alikes have been labelled as real oil spills, then the misclassification cost for the look-alikes can be increased, thus improving the classification of future events.

In other words, in order to fulfil the above-mentioned characteristics, a classifier has to:

– Be flexible to changes in a priori probabilities (*d.prob*);
– Be cost-oriented (*d.co*);
– Automatically select the best operating condition (i.e., the optimal threshold) (*d.aut*);
– Be as much robust and stable as possible (*d.rs*);
– Be time varying (*d.tv*);
– Be computationally efficient (*d.eff*); and
– Be as accurate as possible (*d.acc*).

Most of existing classification algorithms cannot be adapted to simultaneously meet all the above desired characteristics, with the exception of SVMs. The SVMs are powerful classifiers that have demonstrated to be able to achieve state-of-the-art performance on widely used classification benchmarks (thus they meet desired characteristic *d.eff*).

However, using SVMs is just one step towards meeting all the listed desired characteristics. In particular, it is necessary to employ them within the powerful framework of two-class classifiers and ROC analysis. The following sub-sections will show how to build the desired classifier in three steps. In the first step, the used SVM formulation, namely, the cost-oriented (CO) and incremental/decremental (ID) formulation, is presented within the ROC space. This allows meeting desired characteristics *d.co*, *d.tv*, *d.prob*, *d. aut*, other than *d.acc*. In step 2, the use of an ensemble of COID-SVMs (in place of a single COID-SVM) is suggested, in order to make the classifier more robust and stable (i.e. to meet desired characteristic *d.rs*). The use of an ensemble of SVMs in the ROC space, combined with the use of the ROC convex hull, also allows improving the classification accuracy (*d.acc*).

Finally, in step 3, a technique called 'concavity repairing' is employed. This technique is able to further increase the classification performance (*d.acc*) by combining the SVMs instead of selecting the best SVM.

4.3 Step 1: cost-oriented, incremental decremental SVMs (COID-SVMs)

The proposed COID-SVM formulation in the ROC space is based on the incremental/decremental SVM formulation described in Cauwenberghs and Poggio (2001). Their formulation of the incremental learning algorithm builds the solution recursively by adding one new point (pattern) at a time. The constraints for the SVM problem are retained on the previously considered patterns, while the new point is added adiabatically to the solution. At the same time, the decremental unlearning algorithm, in an analogous way, allows to remove data from the fully trained solution. Thus, the so built classifier is able to update the solution adapting the classification to time-varying conditions, making this

approach more desirable than a batch one. Moreover, benefits are also obtained in terms of computational complexity.

In this paper, the ID formulation has been extended to make it both cost-oriented and online-oriented. The cost-oriented characteristic has been achieved by following the approach described in Cortes and Vapnik (1995), which consists in maximising the margin between two separating hyperplanes though tolerating some errors, allowing some positive class elements to be on the negative side and some negative class elements to be on the positive side. The upper bound on the maximum distance between an element and the correct hyperplane determines the cost associated with the misclassification of that pattern [either $M(Y, \boldsymbol{n})$ or $M(N, \boldsymbol{p})$, depending on the class of the pattern]. This modification of the original (not cost-oriented) SVM classification problem results in an additional term in the objective function, which involves costs $M(Y, \boldsymbol{n})$ and $M(N, \boldsymbol{p})$, and which makes the optimization process search for a tradeoff between a large margin and a small misclassification error penalty.

Regarding online classification, it is worth pointing out that in the original algorithm by Cauwenberghs and Poggio (2001) data acquisition was a batch process, while pattern evaluation was performed incrementally, following the same order as the acquisition. On the other hand, with the online approach, data acquisition happens incrementally, so the SVM structures can be dynamically modified, instead of being fixed by the dimension of the data acquired in batch mode. This intrinsic dynamicity improves the adaptability of the system to time-varying conditions. The online approach is handled by introducing a sliding window (which moves one step ahead at each online epoch), over which the training is performed. The use of this window arises from the study of some typical problems of real world applications, such as limited storage capability and the continuous incoming data stream, which require the system update in order to improve the classification.

In order to build an online training for the considered SVM, the algorithm has been structured in the following three sub-steps:

1. *SVM initialization*: This is performed by training the SVM over those patterns belonging to the window. In this way, the SVM structure is initialized based on a small data sample in batch mode. Thus, the resulting SVM trained on the window can be used in incremental mode. In this first step, the only difference with respect to the original ID-SVM described in Cauwenberghs and Poggio (2001) is that the considered one is cost-oriented.
2. *SVM incremental learning*: A new pattern enters the window and the cost-oriented incremental learning starts.
3. *SVM decremental unlearning*: The oldest pattern exits the window and the decremental unlearning starts.

Besides the online training process, which is built by introducing the sliding window, performance evaluation in ROC space (which implies drawing ROC curves, computing the convex hull and choosing the optimum threshold $\underline{\tau}$) requires the presence of a validation set. In addition, performance should be evaluated using a third set, the test set.

In order to integrate all these characteristics, the adopted sliding window has been structured as in Fig. 4. The sliding window has a size twice the size of the desired window for training. At each online epoch, two new ROIs enter the window and two are excluded from it. The SVM is trained over the odd elements within the window, while the even patterns compose the validation set, used to evaluate the expected costs, to build the ROC curves and convex hull and to determine the optimal threshold. This procedure allows only one ROI at a time to enter the training set (another one enters the validation set). The test set is instead chosen outside the window, so that the patterns used for testing are not known to the SVM, thus, the settings established during the validation phase can be evaluated in an independent way. In particular, the test set is composed of a number of patterns following the moving window and equal to the size of the number of elements used for training and for validation. Test set size is thus half of the size of the sliding window. The following subsection explains how the implemented model for online COID-SVM has been included in an ensemble of classifiers.

4.4 Step 2: an ensemble of COID-SVMs in the ROC space

Instead of using just one COID-SVM, with dynamically changing costs $M(Y, \boldsymbol{n})$ and $M(N, \boldsymbol{p})$, a set of COID-SVMs has been employed, using static misclassification costs. The advantages associated with this approach are threefold:

1. An ensemble tends to provide a performance that can outperform the single best classifier;
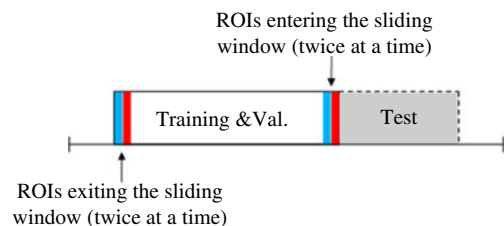

Fig. 4 Sliding window used for the online learning of the SVME. The window is made of two parts: training & validation and test. The training & validation part is made of 2G ROIs [odd ROIs (*blue*) used for training, even ones (*red*) for validation], while the test part is made of the subsequent G ROIs. A value of $G=50$ has been used. The window moves ahead two ROIs at a time, for a total number of 77 online epochs (i.e. 154 ROIs entered the sliding window, two by two, during the phase 2, while the remaining 150 where used for SVME construction during phase 1)

2. An ensemble tends to provide more robust and stable performance; and

3. Static cost functions give better stability to the algorithm used for the SVM online training.

The only drawback is the increase in computational complexity. However, the use of incremental and decremental learning mitigates this problem. Moreover, the computational complexity of the method is negligible if compared to the time needed to download a new satellite image, select, within the image, those regions which could contain possible oil spills, and extract the input features for the classifier. This means that the proposed classifier could work in near real-time conditions, while being able to adapt the classification to the most recent patterns.

All the COID-SVMs composing the ensemble undergo the training process described in step 1; therefore, the validation set is used to draw the ROC curves corresponding to each SVM and to compute the overall convex hull. This means that the information derived by each classifier is automatically selected by means of the convex hull, and the optimum can be chosen according to the iso-performance line, taking into account the entire classifier ensemble.

4.5 Step 3: concavity repairing

In order to improve the convex hull, a technique called *SwapOne*, proposed by Flach and Wu (2003) for repairing concavities in ROC curves has been exploited. It consists of combining together the outputs of two binary classifiers (two points in the ROC curve), in order to obtain a third classifier possibly having a point in the ROC space above the segment connecting the original two points. When this happens, any concavity of the ROC curve between the two classifiers is 'repaired' by the new classifier, which can also extend the ROC convex hull.

By putting all the above described pieces together, the whole classification approach is obtained, consisting of two phases: in phase 1, the SVME is constructed (Fig. 5), while in phase 2, it is used in online mode and it is updated over a sliding window (Fig. 6). In phase 1 and phase 2, 150 ROIs

Fig. 5 Phase 1 (SVME construction)

(first part) and 154 (second part) ROIs of the collected dataset have been used, respectively.

5 Results

The proposed algorithm for building an ensemble of online COID-SVMs, described in the previous section, has been integrated in a software tool, which is structured according to the flowcharts shown in Figs. 5 and 6. The software is provided with a graphic interface which allows the user to set the inputs and to obtain the optimum for the classification, based on the convex hull method. In particular, the user can select the dataset and can set the static misclassification costs for each COID-SVM in the ensemble, the cost functions which define the iso-performance lines and the size of the sliding window. For each data subset entering the sliding window, the software produces a plot showing the ROC curve for each SVM, the convex hull and the optimum for the classification, all computed over the validation set. Moreover, the cost index J[FPR, TPR], defined in Eq. 1, is computed for the ensemble on the validation set and on the test set. The confusion matrix for each SVM can also be computed. Possible concavities in the convex hull can be repaired by enabling the corresponding function. The software employs an ensemble of five SVMs, namely COID-SVM 1, COID-SVM 2, COID-SVM 3, COID-SVM 4 and COID-SVM 5.

The interface allows to choose the values of the static misclassification costs for the negative and positive class [$M(Y, \boldsymbol{n})$ and $M(N, \boldsymbol{p})$, respectively] for each COID-SVM in the ensemble. The cost functions $C(Y, \boldsymbol{n})$ and $C(N, \boldsymbol{p})$ that define the iso-performance lines in the ROC space, used to select the optimum from the convex hull, are also set from the interface. In particular, these cost functions can be either constant or time varying. In the latter case, the symbols $C(Y, \boldsymbol{n})(t)$ and $C(N, \boldsymbol{p})(t)$ will be used.

The size of the moving window can be selected by means of a sliding bar, while concavity repair is enabled by a checkbox. In the following subsection the results obtained by applying the online COID-SVM ensemble to the oil spill dataset will be presented, and the effect of repairing convex hull concavities will be investigated.

5.1 Experiment without concavities repair

Figure 7 shows a typical result for an online classification step. The figure shows the ROC space where the ROC curves for the ensemble of online COID-SVMs are drawn. The violet line represents the overall convex hull and the black line is the iso-performance line tangent to the convex hull. The black square mark represents the optimum, resulting from the algorithm. In order to find a tradeoff between
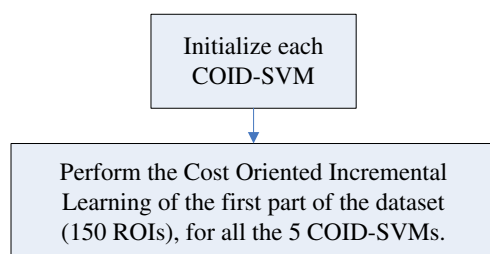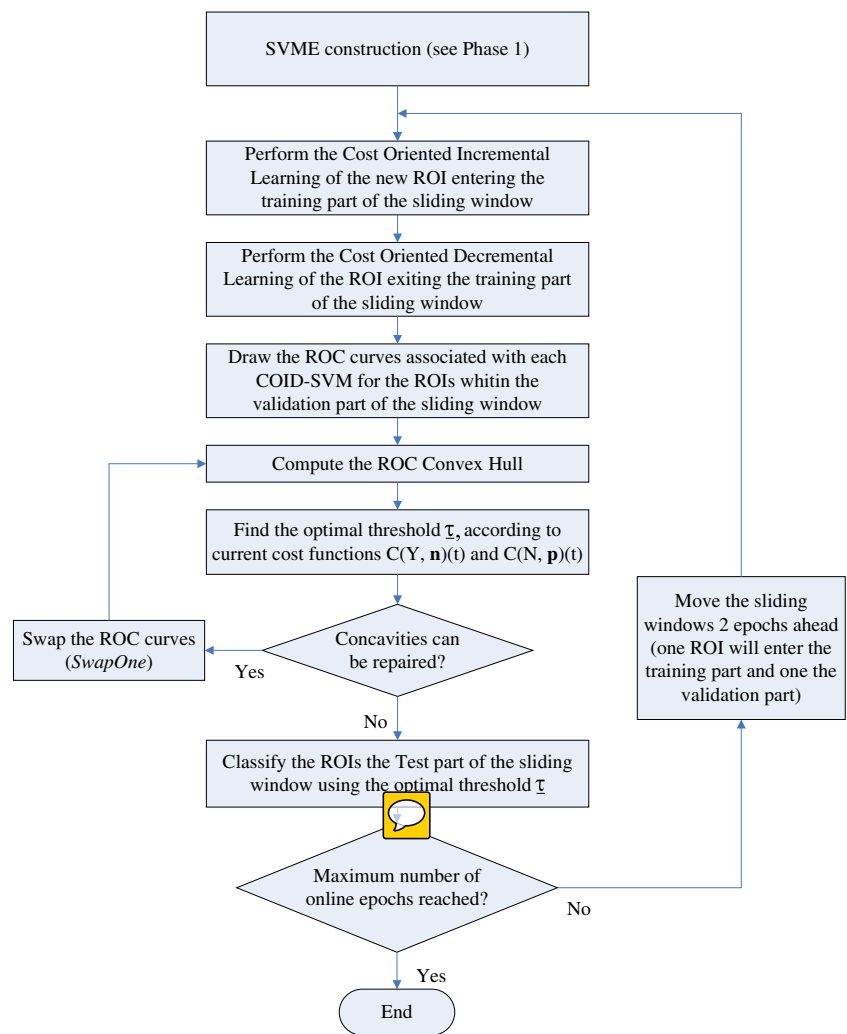
Fig. 6 Phase 2 (SVME online use and update)

the size of the dataset (304 ROIs) and the necessity to test the online learning procedure, a sliding window of 100 ROIs has been chosen, 50 used for training and 50 for validation. According to Fig. 4, each online epoch has been tested on 50 test ROIs.

Time-varying sigmoid cost functions $C(Y, \boldsymbol{n})(t)$ and $C(N, \boldsymbol{p})(t)$ have been used, with values in the interval [1, 2]. As shown in Fig. 8, where the costs are plotted as a function of the online epochs, $C(Y, \boldsymbol{n})(t)$ decreases with time while $C(N, \boldsymbol{p})(t)$ increases. It is important to notice that the absolute values used here for $C(N, \boldsymbol{p})(t)$ and $C(Y, \boldsymbol{n})(t)$ are rather arbitrary: they have been chosen to replicate a situation in which one is greater than the other at the beginning, but the opposite happens at the end. This simulates a real situation with changing needs: after tolerating losing too many real oil spill events, the operator decreases $C(Y, \boldsymbol{n})(t)$ and increases $C(N, \boldsymbol{p})(t)$, in order to detect more true oil spills. This, of course, has the drawback of generating more false alarms near the end

The static misclassification costs associated with each COID-SVM in the ensemble are shown in Table 6. Note

that these misclassification costs could assume any value. In particular, for each SVM, $M(Y, \boldsymbol{n})$ and $M(N, \boldsymbol{p})$ are not necessarily equal. Experimental tests showed that, in this application, using equal misclassification costs for the positive and negative classes, for each COID-SVM, gives better performance. This could be explained with the fact that the oil spill dataset is actually balanced, since the number of elements belonging to the oil spill class is approximately equal to the number of elements belonging to the look-alike class. Indeed, according to the cost-oriented formulation of SVMs (Cortes and Vapnik 1995), the static misclassification costs $M(Y, \boldsymbol{n})$ and $M(N, \boldsymbol{p})$ represent the upper bound on the maximum allowed distance between an element and the separating hyperplane associated (which is defined on the higher dimensional space of the kernel functions used by the SVMs) to the negative class and to the positive class, respectively. Thus, using equal values for the two misclassification static costs does not exactly mean that the cost of misclassifying a positive element is being considered equal to the cost of misclassifying a negative one.
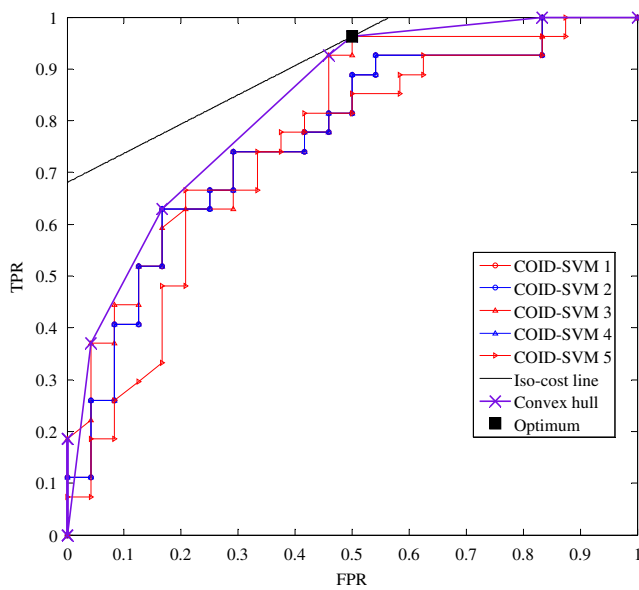
Fig. 7 ROC curves for the ensemble of online COID-SVMs. The *violet curve* represents the convex hull, whereas the *black line* represents the iso-performance line. The *black square mark* is the optimum (it corresponds to the optimal threshold $\underline{\tau}$)

A measure of the classification performance of the system can be achieved by considering the cost index J [FPR, TPR] as defined in Eq. 1, by evaluating it on the test set at the optimal threshold for the ensemble (the optimal threshold being determined on the validation set). However, each online step produces an optimal threshold for the classification of the elements belonging to the current sliding window. This intrinsic dynamicity of the system must be taken into account. In order to obtain a performance index involving the complete dataset, all the cost indexes, each one corresponding to a single online epoch, have been summed over all the steps, so
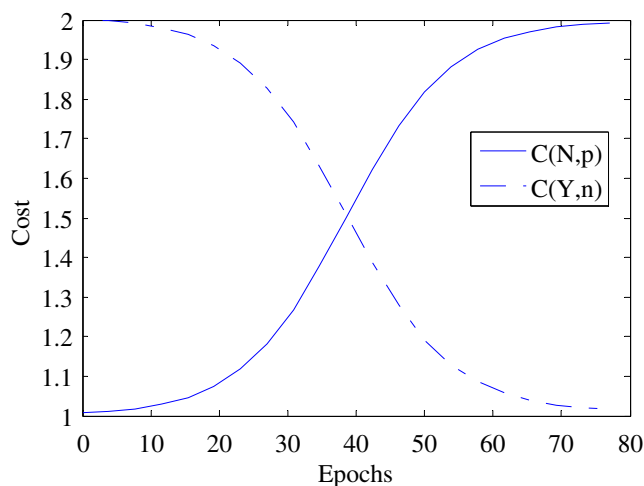

Fig. 8 Time varying cost functions $C(Y, \boldsymbol{n})(t)$ and $C(N, \boldsymbol{p})(t)$, represented as a function of the online epoch number

Table 6 Misclassification costs associated to each COID-SVM in the ensemble

	$M(Y, \boldsymbol{n})$	$M(N, \boldsymbol{p})$
COID-SVM 1	1.0	1.0
COID-SVM 2	1.5	1.5
COID-SVM 3	2.0	2.0
COID-SVM 4	2.5	2.5
COID-SVM 5	3.0	3.0

as to obtain an integrated global cost index for the ensemble applied to the dataset. On the resulting 77 sliding steps, a global cost index of 92.89 has been obtained on the test set and of 77.39 on the validation set. Regarding the mean execution time on the considered dataset, in terms of elapsed CPU seconds, for a single online epoch, this resulted to be 5.52 s, on AMD Athlon X2 2.6 GHz processor, with 2 Gb RAM.

5.2 Experiment with concavities repair

The application of the concavity repairing technique described in previous section (step 3) results in a modification of the convex hull, which brings to an increase in the area below the convex hull. In order to estimate the effect of repairing concavities, the corresponding function has been applied at each online step, and the area under the convex hull has been computed. On the 77 online epochs, the obtained mean area under the convex hull was of 0.70, while repairing concavities at each step resulted in a mean area under the convex hull of 0.72, thus increasing the area of about 3%. Moreover, the increase in the area corresponds to a decrease in the global cost index of the 2% on the test set and of the 5% on the validation set. Figure 9 shows an example of the effect of repairing concavities in the convex hull. In the figure, the variation in the area is highlighted in grey; the optimum computed on the improved convex hull is represented by a red square mark, while the old optimum is represented by a black square mark. From the theoretical point of view, such an improvement is even more significant because it demonstrates that combining classifiers provides better results, in this case, than performing a classifier dynamic selection.

Despite concavity repairing, the SVME misclassifies some ROIs. Figure 10 shows an example of misclassification, on a particularly difficult setting.

6 Summary and discussions

In this study, a dataset made of 157 oil spills and 147 natural phenomena has been collected from MODIS images in the

Fig. 9 Effect of repairing convex hull concavities
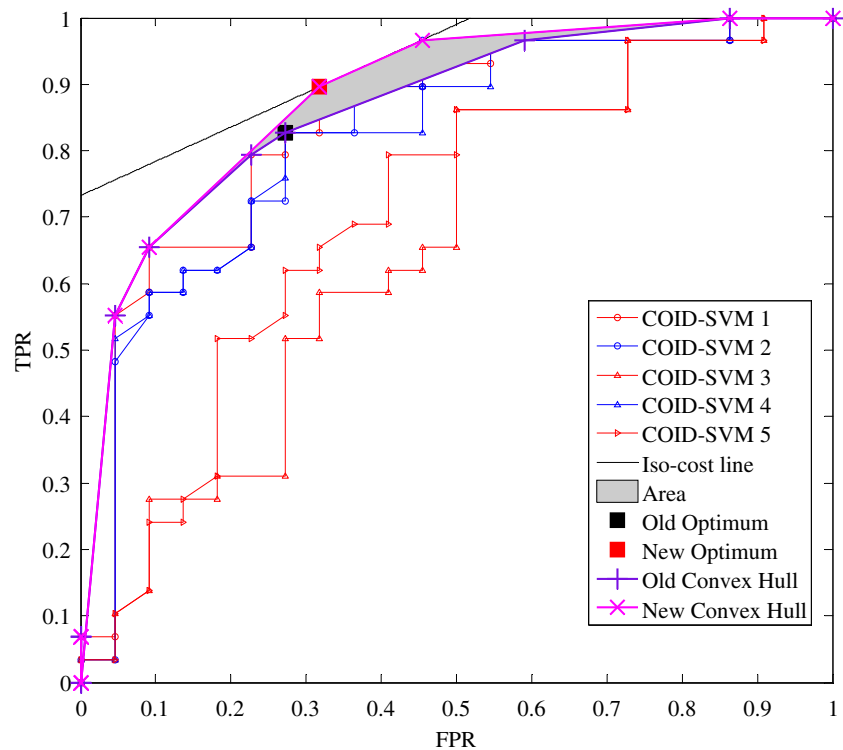
Mediterranean Sea during 2008 and 2009. Then, a powerful supervised classification algorithm, called SVME, has been built to automatically classify them.

It is interesting to notice that the SVME system can be alternatively used in two ways: (a) as a module of an automatic oil spill detection system for MODIS images and (b) as a decision support tool. To use it in the first way, a further module is needed: a ROI generator aimed at making the procedure, described in Section 3, fully automatic. Unfortunately, building a reliable and automatic ROI generator is a challenging task and its implementation is still under study. However, the system can be used in the second way to assist a junior operator (JO), who has some photo-interpretation knowledge, in making more reliable decisions (see Fig. 11).

The role of the JO is to manually select the ROIs from an incoming MODIS image (if the ROI generator does not exist) and to ask the SVME system to classify them. Then, he can decide whether to confirm or discard the classification result. If he is confident enough with his final decision, he can also ask the system to learn the new ROIs, in order to extend the knowledge base, and optionally, to unlearn the eldest ones (to make the system time varying and thus more adaptive to changes). Otherwise, the new ROIs are not learned by the SVME, until a more reliable label can be assigned after an in situ verification by the coast guard or after checking a SAR image or even after consulting a skilled operator. The ultimate goal of the SVME when used as a decision support tool is to bias the decision of the JO towards more correct classifications, ideally trying to reach those of a skilled operator.
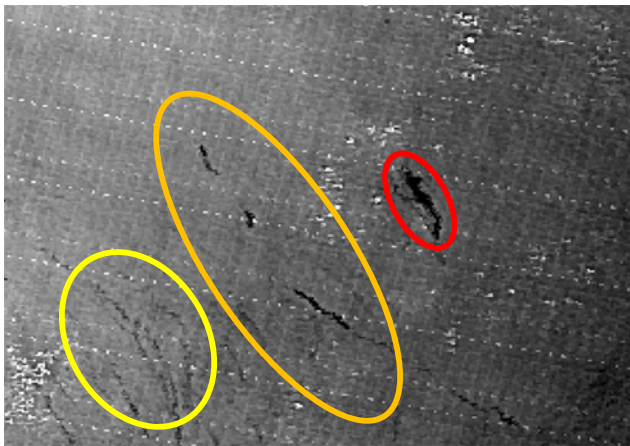

Fig. 10 During phase 2, the system is able to correctly identify the *red region* as an OS and the regions in the *yellow area* as look-alikes, but it misclassifies some of the regions in the *orange area*. This is tolerable, since even for a skilled operator it is difficult to classify the ROIs in the orange area. Only by considering the context, beside geometric and grey level features, the skilled operator can correctly identify the ROIs in the orange area as look-alikes, by recognising the contribution of marine currents

7 Conclusions and future work

The conclusions that can be drawn from this study are: (a) automatic classification of ROIs identified on MODIS
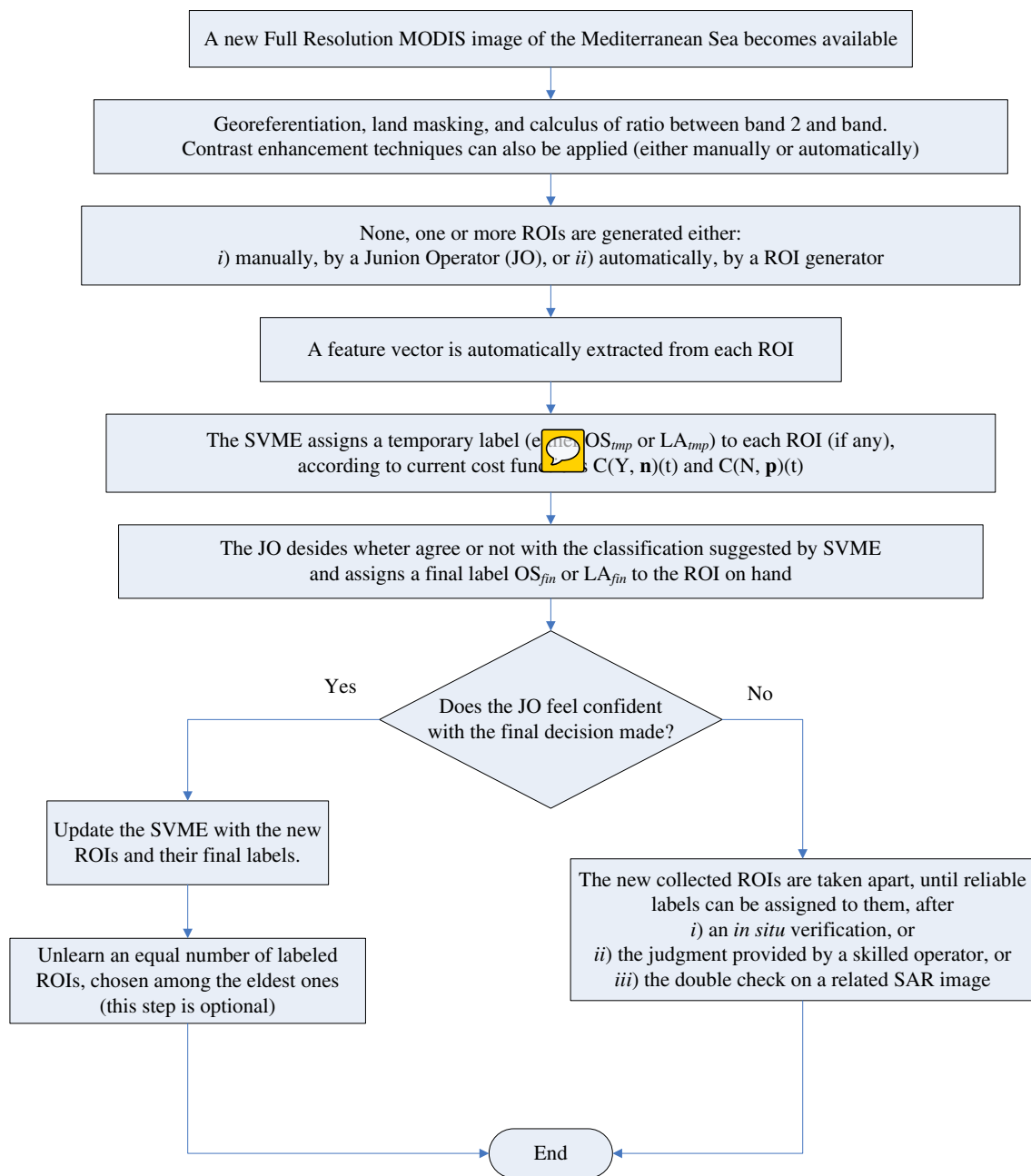
Fig. 11 An example on the use of the SVME on incoming images, as tool for supporting the decision of a junior operator. Observe that only phase 2 is shown (online use and system update)

images is possible, with promising performance, and (b) the proposed classification algorithm, based on an ensemble of SVMs, is flexible and powerful enough to accommodate some important characteristics that have been put as requirements.

In the future, dataset extension is planned, especially aimed at collecting more verified oils spills and more sunglint-contaminated ROIs. The availability of a significant dataset of sunglint contaminated cases, in particular, might allow to develop an independent classifier, tailored to handle such cases. This would increase the reliability of the

overall system, by exploiting the fact that in the presence of sunglint, the smoother oil tends to reflect more than the more corrugated water.

The implementation of an automatic ROI detection software from MODIS images is planned as well: the availability of such module will allow to run a processing chain that will automatically download the latest MODIS image, process it and possibly generate an oil spill alert, at no cost, since MODIS images are free of charge. Investigating how to combine optical and SAR data is another planned activity, to be run in parallel with the above ones.

Appendix 1 Properties of crude and refined oils

Crude oils are complex mixtures of various hydrocarbons (aromatic and aliphatic), resins and asphaltenes. Refined products contain only a subset of these compounds each. The relative proportion of the various constituents determines both physical properties and optical characteristics of any given oil.

1.1 Chemical composition

Crude oils can be classified according to the relative amount of n-alkanes, branched alkanes, cycloalkanes, aromatic hydrocarbons and NSO compounds:

- Paraffinic oils: mainly composed of acyclic alkanes. These oils have a low sulphur content.
- Paraffinic–naphthenic (or mixed-base) oils are composed of both cyclic and acyclic alkanes. These oils have a low sulphur content. The majority of crude oils belongs to this group.
- Aromatic-intermediate oils: These oils contain more than 50% aromatic hydrocarbons and usually have a higher content of NSO compounds, particularly those containing sulphur.

Refined oils are distillation products of crude oils and are classified according to the number of carbon atoms contained in the molecules making up the mixture. In decreasing the order of volatility (increasing density and viscosity), these are:

- Petrol and naphtha with benzene and other volatile oils (C4–C10).
- Kerosene and lamp oils (C11–C13).
- Diesel and light gas oils (C14–C18).
- Heavy gas oils and home heating oils (C19–C25).
- Lubricating oils and light fuel oils (C26–C40).
- Residual and heavy fuel oils (>C40).

1.1.1 Density and specific gravity

Oil density is often measured relative to water, in order to obtain a dimensionless quantity, specific gravity (SG). Most oils are lighter than water, and in particular, crude oils are often divided into light, medium and heavy, on the basis of American Petroleum Institute (API) gravity measurements (Killops and Killops 2005):

$$°API = \left(\frac{141.5}{SG_{60F}}\right) - 131.5 \tag{2}$$

where SG_{60} is the specific gravity of the oil at 60°F (i.e. the ratio between oil density at 60°F and pure water density at the same temperature). The API scale is thus inversely proportional to density. According to API definition, the following classes have been specified:

- Light crudes: API gravity ≥30°API (SG ≤0.87).
- Medium crudes: API gravity between 22 and 30°API (0.87<SG≤0.92).
- Heavy crudes: API gravity between 10 and 22°API (0.92<SG≤1.00).
- Tars or tar sands: API gravity <10°API, (SG ≥1.00).

1.1.2 Viscosity

The viscosity of oil is a measure of the oil's resistance to shear, so it is a measure of its resistance to flow. When viscosity is determined by directly measuring shear stress and shear rate, it is expressed in centipoises (cP) or in Pa s (1 cP=1 mPa s) and is referred to as the absolute or dynamic viscosity. For instance, water at 20°C has a dynamic viscosity of 1 cP. The viscosity of single compound hydrocarbons increases with the number of carbon atoms contained in a molecule. High molecular weight hydrocarbons are generally soluble in low molecular weight alkanes or aromatics, and the viscosity of the mixture depends on the relative content of low and high molecular weight compounds. Viscosity also varies with density. Light oils have viscosities less than 30 mPa s at 20°C, while heavy asphaltic oils have viscosities measured in thousands of mPa s. For natural tars, viscosity is usually greater than 10^4 mPa s (North 1985).

1.2 Oil optical properties

Inherent optical properties that are relevant for oil detection are: absorption, fluorescence and refractive index. Scattering is a small effect at visible wavelengths and it usually has a magnitude and spectral behaviour close to that of pure seawater.

1.2.1 Absorption

Typical crude oil spectra show strong absorption at short wavelengths that range from ultraviolet (UV) or violet for the lightest oils to red or NIR for the heaviest crudes. Absorption coefficients decrease exponentially towards longer wavelengths. Decay widths do not vary much within crude oils. Actually, this behaviour is similar to that of CDOM, so that it may be difficult to distinguish the spectral signature of dispersed oil from that of CDOM. Refined oils have an anomalous behaviour, due to the process of fractionation. In particular, the absorption coefficient of light

refined oils decreases steeply towards visible wavelengths, while heavy refined oils show lower absorption in the UV, but larger decay widths extending to NIR wavelengths.

1.2.2 Fluorescence

Oil fluorescence is generally excited by light at UV to blue wavelengths. The spectral fluorescence properties of oils when excited by laser pulses can be used to distinguish between different oil types. Generally speaking, as excitation wavelength increases, fluorescence is induced in a smaller number of compounds, thus, light oils fluoresce at shorter wavelengths, from UV to green, while heavy oils show broader fluorescence spectra which peak at longer wavelengths (Ryder 2005). Solar radiance could also induce oil fluorescence, but, since it is less intense than excitation by laser, the fluorescence signal will be weaker. However, solar-induced fluorescence may contribute to solar reflectance on particularly bright days. Fluorescence peaks vary in width and wavelength position according to oil type and decay exponentially toward the red and NIR.

1.2.3 Refraction

Oil refractive index is greater than that of seawater ($\approx$1.34) at visible and NIR wavelengths. There is a significant variability between oil types, but usually, heavy oils have a higher refractive index. For crude oils, the refractive index lies between 1.57 and 1.67 in the UV and between 1.48 and 1.52 in the visible (Osadchy et al. 1994). Refractive index of course influences light specular reflection at the air–sea interface for clean and oil-covered sea surfaces.

Appendix 2 Oil–water contrast

The presence of an oil film on the sea surface implies some modifications in the upwelling radiance, and thus, in the remote sensing reflectance (Osadchy et al. 1994) measured over the sea surface.

In particular, the optical properties of crude and refined oils allow distinguishing them from water. More precisely, as mentioned before:

– Oil refractive index is greater than that of seawater.
– Oil absorption coefficients are several orders of magnitude greater than that of water in the blue and decay exponentially with wavelength.
– Crude and refined oils fluoresce when subjected to bright natural light, with fluorescence peaks that vary in width and wavelength position according to oil type, and decay exponentially towards the red and NIR.

These characteristics contribute to the detectable oil–water contrast. This can be described by analysing the contributions to upwelling radiance measured above an oil film (Byfield and Boxall 1999), which is represented in Fig. 12.

– Atmospheric path radiance (a): this contribution should be removed, but for oil spill monitoring applications, this is not essential.
– Specular reflection of sky radiance (b): this contribution is greater from an oil-covered seawater surface due to oil's higher refractive index. Contributions due to reflection at the oil–water interface are two orders of magnitude lower than at the air–oil interface, and may thus be neglected.
– Water-leaving radiance (c): this contribution arises from a portion of incident light that is transmitted through the surface and scattered back up by the seawater and its constituents. The presence of an oil film has a double effect on this contribution: reduced transmittance through the air–oil interface and absorption within the oil layer. When the oil is sufficiently thick, all light is absorbed and further increases in oil thickness are not detectable. This limiting thickness depends on the oil absorption coefficient and the wavelength of the light, from 0.02 mm for heavy crudes at 440 nm to about 2 mm for light crudes at 750 nm (Byfield 1998).
– Oil fluorescence and scattering (d): oil fluorescence excited by sunlight in the UV and scattering from water entrained in the oil also contribute. In particular, in heavier oils with high absorption coefficients throughout the
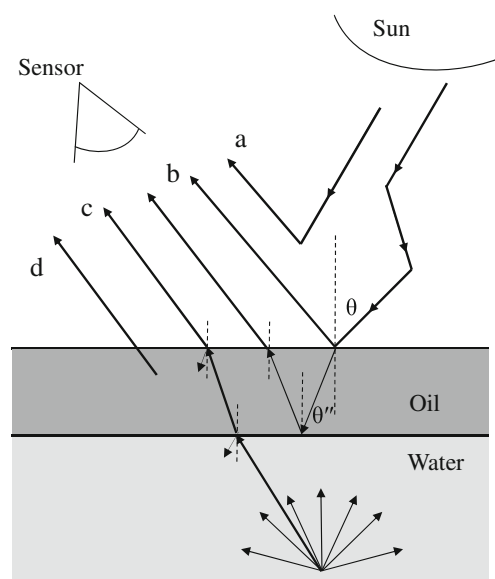

Fig. 12 Contributions to radiance measured above an oil-covered seawater surface. **a** Atmospheric path radiance. **b** Specular reflection of sky radiance. **c** Water leaving radiance. **d** Fluorescence and scattering contributions

optical range, this contribution is low. The absorption coefficient of light crudes decays more rapidly, so fluorescence and scattering may make a significant contribution at longer wavelengths.

Oil–water contrast, C, can be defined as the difference between the remote sensing reflectance, in the presence of an oil film, and remote sensing reflectance of clean water, normalised to remote sensing reflectance of clean water (in the following angular and wavelength dependence are omitted for simplicity):

$$C = \frac{R_{rs}^{(oil)} - R_{rs}^{(clean\ water)}}{R_{rs}^{(clean\ water)}}.$$

==Taking into account== all the above-mentioned contributions, the contrast can be written as (Byfield and Boxall 1999):

$$C = \frac{(r_o - r_w)R_{rs}^{(sky)} - (t_w t_{dw} - t_o t_{do} e^{-[1/\mu_d + 1/\cos\theta'']a_0 z})R_{rs} + F_o + B_o}{r_w R_{rs}^{(sky)} - t_w t_{dw} R_{rs} + P}$$

$$(2)$$

where $R_{rs}^{(sky)}$ represents the sky remote sensing reflectance (that is remote sensing reflectance, where the numerator is only sky radiance), r_o and r_w are the specular reflection (Fresnel) coefficients for oil and water, R_{rs} is the remote sensing reflectance, t_o and t_w are the Fresnel transmittance coefficients for oil and water, t_{do} and t_{dw} are the diffuse transmittance coefficients for incident light, a_o is the oil absorption coefficient, z is oil thickness, μ_d is the average cosine of downwelling irradiance within the oil, F represents oil fluorescence, B represents scattering by water in the oil and P is the atmospheric contribution. It is clear from Eq. 2 that oil–water contrast depends on many factors (oil type, oil thickness and water constituents), which affect scattering and absorption in the seawater and atmosphere contribution. Thus, contrast can be positive or negative depending on water type and on environmental conditions, but it also vary according to wavelength, since absorption and fluorescence have a strong spectral dependence. Thus, for a given oil on a seawater surface, C can be positive in some spectral regions and negative in others. In particular, from Eq. 2, it can be observed that C is positive when:

$$(r_o - r_w)R_{rs}^{(sky)} + F_o + B_o$$
$$> \left(t_w t_{dw} - t_o t_{do} e^{-[1/\mu_d + 1/\cos\theta'']a_0 z}\right)R_{rs}.$$

For thin oils, this occurs at wavelengths where $R_{rs}^{(sky)}$ is high, or where R_{rs} is low, usually the violet and blue (between 400 and 480 nm), and NIR (starting at 700 nm). As the thickness increases, the exponential term approaches 0, and the

contrast becomes negative throughout most of the visible spectrum. In the NIR, R_{rs} is negligible, due to high water absorption, and the contrast is positive, except when there is a high concentration of scattering particles in water. However, near the seawater reflectance peak, between 480 and 570 nm in coastal waters, R_{rs} may be sufficiently high to give low or negative contrast even for thin oils.

The change in contrast between oil spills and seawater, depending on both oil type and water type, for different sun-observer geometries, as also been investigated in Otremba and Piskozub (2001) and Otremba and Piskozub (2003).

References

Alawadi F, Amos C, Byfield V, Petrov P (2008) The application of hyperspectral image techniques on MODIS data for the detection of oil spills in the RSA. Proc SPIE, 7110:71100Q–71100Q12. doi:10.1117/12.799374

Alpers W, Huhnerfuss H (1989) The damping of ocean waves by surface films: a new look at an old problem. J Geophys Res 94 (C5):6251–6265

Brekke C, Solberg AHS (2005) Oil spill detection by satellite remote sensing. Remote Sens Environ 95:1–13

Byfield V (1998) Optical remote sensing of oil in the marine environment, Ph.D. Thesis, University of Southampton

Byfield V, Boxall SR (1999) Thickness estimates and classification of surface oil using passive sensing at visible and near-infrared wavelengths. Proc. of IEEE IGARSS'99, vol. 3, pp. 1475–1477

Cauwenberghs G, Poggio T (2001) Incremental and decremental support vector machine learning, Adv. Neural Information Processing Systems (NIPS*2000), Cambridge, MA: MIT Press, vol. 13

Cortes C, Vapnik V (1995) Support vector networks. Mach Learn 20:273–297

Del Frate F, Petrocchi A, Lichtenegger J, Calabresi G (2000) Neural networks for oil spill detection using ERS-SAR data. IEEE Trans Geosci Rem Sens 38:2282–2287

Ferraro G, Baschek B, de Montpellier G, Njoten O, Perkovic M, Vespe M (2010) On the SAR derived alert in the detection of oil spills according to the analysis of the EGEMP. Mar Pollut Bull 60:91–102. doi:10.1016/j.marpolbul.2009.08.025

Fiscella B, Giancaspro A, Nirchio F, Pavese P, Trivero P (2000) Oil spill detection using marine SAR images. Int J Remote Sens 21:3561–3566

Flach PA, Wu S (2003) Repairing concavities in ROC curves, in Proc. 2003 UK workshop on computational intelligence, pp. 38–44

GESAMP (Joint Group of Experts on the Scientific Aspects of Marine Environmental Protection) (2007) Estimates of oil entering the marine environment from sea-based activities, GESAMP reports and studies 75

GESAMP (Joint Group of Experts on the Scientific Aspects of Marine Environmental Protection) (2009) Pollution in the open oceans: a review of assessments and related studies, GESAMP reports and studies 79

Hu C, Müller-Krager FE, Taylor CJ, Myhre D, Murch B, Odriozola AL et al (2003) MODIS detects oil spills in Lake Maracaibo, Venezuela. EOS Trans Am Geophys Union 84(33):313–319

Hu C, Li X, Pichel WG, Muller-Karger FE (2009) Detection of natural oil slicks in the NW Gulf of Mexico using MODIS imagery. Geophys Res Lett 36:L01604. doi:10.1029/2008GL036119

Keramitsoglou I, Cartalis C, Kiranoudis CT (2006) Automatic identification of oil spills on satellite images. Environ Model Softw 21:640–652

Killops S, Killops V (2005) Introduction to organic geochemistry, 2nd edn. Blackwell, Oxford

Kluser S, Richard JP, Giuliani G, De Bono A, Peduzzi P (2006) Illegal oil discharge in European seas, Coll. Environment Alert Bulletin, UNEP/DEWA–Europe/GRID–Geneva, pp. 4

Kubat M, Holte RC, Matwin S (1998) Machine learning for the detection of oil spills in satellite radar images. Mach Learn 30:195–215

Liu AK, Wu SY, Tseng WY, Pichel WG (2000) Wavelet analysis of SAR images for coastal monitoring. Can J Remote Sens 26 (6):494–500

North FK (1985) Petroleum geology. Allen & Unwin, Boston

Oceana (2003) The dumping of hydrocarbons from ships into the seas and oceans of Europe—the other side of oil slicks. Oceana Europe Report, pp. 26, available at http://eu.oceana.org/sites/default/files/reports/oil-report-english.pdf

Osadchy VU, Shifrin KS, Gurevich IY (1994) Remote sensing and measurement of the thickness of oil films on the sea surface using the reflectivity contrast. Proc. SPIE, Ocean Optics XII, 2258, pp. 747–758

Otremba Z, Piskozub J (2001) Modelling of the optical contrast of an oil film on a sea surface. Opt Express 9(8):411–416

Otremba Z, Piskozub J (2003) Modeling the remotely sensed optical contrast caused by oil suspended in the sea water column. Opt Express 11(1):2–6

Provost F, Fawcett T (2001) Robust classification for imprecise environments. Mach Learn 42:203–231

Ryder AG (2005) Analysis of crude petroleum oils using fluorescence spectroscopy. Reviews in fluorescence. Annual volumes, pp. 169–198

Solberg AH, Solberg R (1996) A large-scale evaluation of features for automatic detection of oil spills in ERS SAR images. In Proc. of IEEE IGARSS'96, pp. 1484–1486

Solberg AH, Storvik G, Solberg R, Volden E (1999) Automatic detection of oil spills in ERS SAR images. IEEE Trans Geosci Rem Sens 37:1916–1924

Topouzelis K, Karathanassi V, Pavlakis P, Rokos D (2004) Oil spill detection using RBF neural networks and SAR data. In Proc. of XXth ISPRS Congress, Istanbul, Turkey