CLASSIFYING OIL SPILLS AND LOOK-ALIKES IN ENVISAT ASAR IMAGES

Camilla Brekke[1,2], **Anne Solberg**[2] **and Geir Storvik**[3]

[1] *Norwegian Defence Research Establishment,*
PO Box 25, NO-2027 Kjeller, Norway, E-mail: Camilla.Brekke@ffi.no

[2] *Department of Informatics, University of Oslo,*
PO Box 1080 Blindern, 0316 Oslo, Norway, E-mail: anne@ifi.uio.no

[3] *Department of Mathematics, University of Oslo,*
PO Box 1053, Blindern, 0316 Oslo, Norway, E-mail: geirs@math.uio.no

ABSTRACT

We propose an improved classification approach for automatic oil spill detection in Synthetic Aperture Radar (SAR) images in the framework of a dark spot detector, a dark spot feature extractor and dark spot classification. New ideas for optimal subclass estimation are discussed. A regularized statistical classifier for oil spill and look-alike classification is applied within each subclass. To allow the user to tune the system with respect to the trade-off between the number of true positive alarms and the number of false positives, an automatic confidence estimator has been developed. The system is trained on 76 ENVISAT ASAR images and performance tested on 27 images.

1. INTRODUCTION

Oil spills and several other ocean features (*look-alikes*) dampen out the small scale waves on the sea surface. This reduces the backscattering back to the SAR antenna and dark slicks appear in the SAR images. A part of the oil spill detection problem is to distinguish oil slicks from the look-alikes. Our goal is to develop an automatic system for oil spill detection, in which objects with a high probability of being oil spills are identified.

The framework of our algorithm is a dark spot detector, a dark spot feature extractor and a dark spot classifier. Dark spots in the images are primarily detected by adaptive thresholding. For each of them a number of features are computed in order to classify the slick as either an oil spill or a look-alike. The classification scheme is based on statistical modelling.

As the behaviours of the features will change with different wind levels, we spilt the feature space into subclasses. A search for an optimal feature space division for the purpose of oil spill classification within each subclass has been performed based on Maximum Likelihood Estimation (MLE) and is the focus of this paper.

Various classifiers have been applied to the oil spill detection problem: A Mahalanobis classifier and a compound probability classifier were applied in [4]. The probabilistic approach was improved in [6]. A neural-network approach is described in [3], and a classifier based on fuzzy logic was developed in [5]. The work presented here is based on a statistical classifier [10, 8, 1, 2].

2. ALGORITHM DESIGN

The algorithm consists of the following main steps: 1) dark spot detection based on segmentation of the SAR image, 2) feature extraction from the segmented image, 3) classification of the detected dark spots as oil spills or look-alikes and finally 4) estimation of confidence levels for each of the slicks classified as oil spills.

Pre-processing of the SAR image, consisting of converting a land mask to the image grid and a normalization of the backscatter with respect to incidence angles, is performed ahead of the segmentation step.

2.1. Segmentation of Dark Spots

In the segmentation step, dark spots are separated from the background. This is the most computationally intensive step of the algorithm.

Proc. 'Envisat Symposium 2007', Montreux, Switzerland
23–27 April 2007 (ESA SP-636, July 2007)

First, an image pyramid is created by averaging pixels in the original image. From the original image, the next level in the pyramid is created with half the pixel size of the original image, and so on. Adaptive thresholding is then applied to each level in the pyramid. The threshold is set adaptively based on estimates from the SAR image of the roughness/texture of the surrounding sea. After segmenting each level in the pyramid, the different levels are merged. More details are given in [8].

To detect thin linear slicks, elongated segments are first located in a coarse segmented version of the image. Then these segments are grown in the direction of their orientation if certain criteria are fulfilled (e.g. the backscatter value of a pixel has to be below a threshold value and it or one of its 8-neighbour pixels has to represents an edge pixel). See [1] for more details.

The 27 segmented test images contains 12245 look-alikes and 41 oil spills (64 segmented regions).

2.2. Dark Spot Feature Extraction

Discrimination between oil spills and look-alikes is based on a number of features computed for each of the segmented dark spots. The features are computed and collected in a feature vector $\mathbf{x}_i$. These features are constructed such that they typically will be different depending on if the dark spot is an oil spill or a look-alike. The following features were selected in previous experiments [1].

- **Shape features:** "Sum of External Angles", "Slick Moment (MOM)", "Slick Area", "Slick Complexity" and "Slick Width"
- **Contrast features:** "Slick Local Contrast", "Slick Border", "Smoothness Contrast"
- **Texture:** "Power-to-mean Ratio (PMR)", "Slick Variance"
- **Surroundings:** "Regions in Small Neighbourhood", "Distance to Ship/Oilrig", "Wind", "Low Wind Area"

2.3. Optimal Subclass Boundary Estimation

When selecting a classifier to solve this problem, a possibility is to use Bayesian estimation techniques to calculate the *posterior* probability for a detected spot being an oil spill. Then, a distribution needs to be selected to model the classes. The simplest choice would be multivariate Gaussian densities.

However, the behaviour of the features will change with different wind levels. In low wind, the backscatter difference between an oil spill and the surrounding sea will be large. As the wind level increases, the backscatter contrast will be lower and the feature values will change. A multimodal conditional density will likely occur. Describing the conditional density by a unimodal density (e.g. Gaussian) is therefore not appropriate. If we split the problem and assume different densities depending on the wind level, the data within each subdivision is more likely to be more homogeneous.

Even within each wind level, both the oil slicks and the look-alikes may vary quite a lot in shape, contrast and other features. In [2], the wind level was first used to divide the samples in two different subclasses and then these were divided into five subclasses based on the shape descriptor. We have applied MOM as a shape descriptor and the wind level is represented by PMR. In low wind with many look-alikes, the PMR value will be high, and high contrast between slicks and their surrounding can be expected. As the wind increases the PMR value will decrease, and the expected contrast between the oil and the surrounding sea will also decrease. Let $\Omega_1, ..., \Omega_K$ represent the subclasses, where K = the number of subclasses. Tab. 1 shows the initial configuration with $K = 10$ as applied in earlier studies [2]. The goal of the current study is to estimate these subclass borders. $\hat{\theta} = \{p, m_1, m_2, m_3, m_4\}$ are the parameters defining the subclasses and their values need to be optimised for a best possible performance. Tab. 2 presents the parameter values applied in [2]. These values are here used as the initial input to an optimiser searching for the optimal $\hat{\theta}$.

Table 1. The initial subclass division.

	Slick	**Surrounding**
Subclass	**MOM**	**PMR**
Ω_1	MOM $< m_1$	PMR $\geq p$
Ω_2	MOM $\in [m_1, m_2 >$	PMR $\geq p$
Ω_3	MOM $\in [m_2, m_3 >$	PMR $\geq p$
Ω_4	MOM $\in [m_3, m_4 >$	PMR $\geq p$
Ω_5	MOM $\geq m_4$	PMR $\geq p$
Ω_6	MOM $< m_1$	PMR $< p$
Ω_7	MOM $\in [m_1, m_2 >$	PMR $< p$
Ω_8	MOM $\in [m_2, m_3 >$	PMR $< p$
Ω_9	MOM $\in [m_3, m_4 >$	PMR $< p$
Ω_{10}	MOM $\geq m_4$	PMR $< p$

Table 2. θ: best guess for $\hat{\theta}$.

p	m_1	m_2	m_3	m_4
0.04	0.3	0.5	0.8	1.2

For the parameter search, the oil spill and the look-alike classes are merged in the training set, and the sum of the negative log-likelihood for each subclass is used as a criteria for optimisation. During the search, Gaussian densities are assumed within each subclass: $f_{\Omega_k} = N(\eta_{\Omega_k}, \mathrm{T}_{\Omega_k})$, and the $\hat{\theta}$ minimizing the sum of all $l(\hat{\theta})_{\Omega_k}$ for $k = 1, ..., K$ is selected, where $l(\hat{\theta})_{\Omega_k}$ is defined as:

$$l(\hat{\theta})_{\Omega_k} = -\sum_{\mathbf{x}_i \in \Omega_k} [-\frac{p}{2}\log(2\pi) - \frac{1}{2}\log|\mathrm{T}_{\Omega_k}| - \frac{1}{2}(\mathbf{x}_i - \eta_{\Omega_k})^T \mathrm{T}_{\Omega_k}^{-1}(\mathbf{x}_i - \eta_{\Omega_k})] \quad (1)$$

where p is the number of features. Because η_{Ω_k} and T_{Ω_k} are unknown, these must be estimated as well. Using maximum likelihood estimates for these, Eq. 1 reduces to

$$l(\hat{\theta})_{\Omega_k} = -(-\frac{1}{2} n_{\Omega_k} \log|\widehat{\mathrm{T}}_{\Omega_k}|) \quad (2)$$

where n_{Ω_k} represents the number of samples within subclass Ω_k and $\widehat{\mathrm{T}}_{\Omega_k}$ is the maximum likelihood estimate for T_{Ω_k} based on these n_{Ω_k} samples.

2.4. *Log*-transform of the data set

If we look at the training data, some of the features appear to have a skewed distribution, therefore we did some experiments with a *Log*-transform ($x' = \log(x - min(x) + 1)$) on the data set. As an example, Fig. 1 shows histograms for the feature "Slick Complexity" before and after the *Log*-transform. A Linear Bayes Normal Classifier (LDC) in Matlab was trained for each subclass to evaluate the effect of the *Log*-transform on the test set. The initial configuration with 10 subclasses (Tab. 1) and θ (Tab. 2) was applied. The results presented in Tab. 3 shows that a normalization with the *Log*-transform does not improve the result. Therefore,

Table 3. Log-*transform. Classification results.*

LDC		
	No transform	*Log*-transform
Oil spills (error rate)	9/64	12/64
Look-alikes (error rate)	1947/12245	1394/12245
Total error rate:	15.0%	15.1%

we have not applied any transformation on the feature vectors before the classification step.

2.5. Classification

A prior distribution and a probability density for the features are combined through Bayes theorem

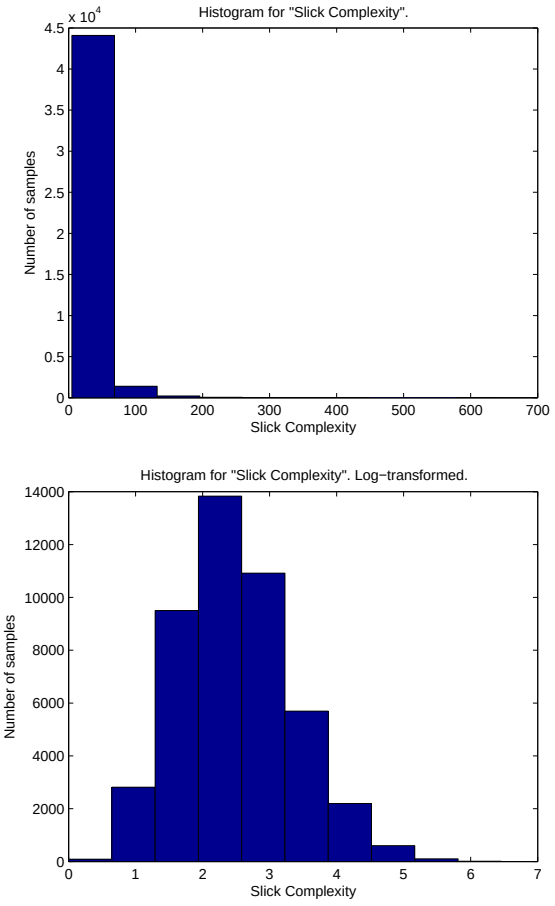

Figure 1. Histograms on the training set for the feature "Slick Complexity". Top: No transform. Bottom: Log-*transformed data.*

to obtain the *posterior* probability for a detected spot being an oil spill. Let c be the unknown class membership of a detected spot (we are dealing with a two class problem: oil spill or look-alike). Then, for $x_i \in \Omega_k$,

$$Pr(c = o|\mathbf{x}_i) = \frac{\pi_o f_{o,\Omega_k}(\mathbf{x}_i)}{\pi_o f_{o,\Omega_k}(\mathbf{x}_i) + (1 - \pi_o) f_{l,\Omega_k}(\mathbf{x}_i)} \quad (3)$$

where π_o is the prior model for the probability that a detected spot is oil. $f_{o,\Omega_k}(\mathbf{x}_i)$ and $f_{l,\Omega_k}(\mathbf{x}_i)$ are the probability densities for the observed features $\mathbf{x}_i$ in classes o = oil spills and l = look-alikes, respectively.

The densities are assumed Gaussian:

$$f_{c,\Omega_k}(\mathbf{x}_i) = \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma_{\Omega_k}|^{\frac{1}{2}}} \times \exp\{-\frac{1}{2}(\mathbf{x}_i - \mu_{c,\Omega_K})^T \Sigma_{\Omega_k}^{-1}(\mathbf{x}_i - \mu_{c,\Omega_k})\} \quad (4)$$

where $c \in \{o, l\}$, d is the number of features, μ_{c,Ω_k}

is the mean vector for class c and Σ_{Ω_k} is the covariance matrix, common for both classes due to the imbalanced data set.

The features "Slick Complexity", "Power-to-mean Ratio", "Slick Local Contrast", "Slick Width", "Regions in Small Neighbourhood", "Slick Border", "Smoothness Contrast" and "Slick Variance" are included in feature vector $\mathbf{x}_i$.

2.5.1. Covariance Matrix Estimation

Regularized covariance matrices are used in the classifier. With the subclass division, the density for class c within a subclass Ω_k is then given by $f_{c,\Omega_k}(\mathbf{x}_i) = N(\mu_{c,\Omega_k}, \widetilde{\Sigma}_{\Omega_k})$.

Applying Gaussian densities and regularized covariance matrices $\widetilde{\Sigma}_{\Omega_k}(\rho_{\Omega_k})$, leads to a general family of covariances indexed by ρ_{Ω_k}. Regularization of the common covariance matrices can be expressed as follows:

$$\widetilde{\Sigma}_{\Omega_k}(\rho_{\Omega_k}) = \rho_{\Omega_k}[diag\{\Sigma_{\Omega_k}\}] + (1 - \rho_{\Omega_k})\Sigma_{\Omega_k} \quad (5)$$

where ρ_{Ω_k} is the regularization parameter and Σ_{Ω_k} is the fully estimated common covariance matrix. Here $\rho_{\Omega_k} \in [0, 1]$ allows a continuum of models, and needs to be estimated from experiments as described in [2].

2.6. Confidence Estimation

After first applying a regularized statistical classifier within each subclass, each slick with a higher posterior probability of being an oil spill than a look-alike is automatically assigned a confidence level. We have developed an automatic confidence estimator as a second step in our two-step classification approach (see Fig. 2). The confidence estimator will automatically assign a slick one of four confidence levels: *High, Medium, Low* or *Very Low*.

If the system is operating on level *Medium*, all slicks with *High* and *Medium* confidence are reported. If the system is operating on level *Low*, all slicks with *High, Medium* and *Low* confidence are reported. All slicks that are not detected as *High, Medium* or *Low* are given *Very Low* confidence. When operating on level *Very Low* all slicks detected as oil spills in the first step of the classification approach (see Fig. 2) are reported.

Kongsberg Satellite Services (KSAT) manual oil spill service chain is described in [9]. To determine the confidence level of a slick, the operators use a set of guidelines. We used these guidelines as a starting point when designing our automatic procedure for confidence estimation. However, experiments showed that it is not sufficient to base the design only on these guidelines [7]. In addition to translating several of the criteria into computed features, we included additional features that we found important for reliable confidence estimation. The development and training of the confidence estimator is described in more detail in [2].

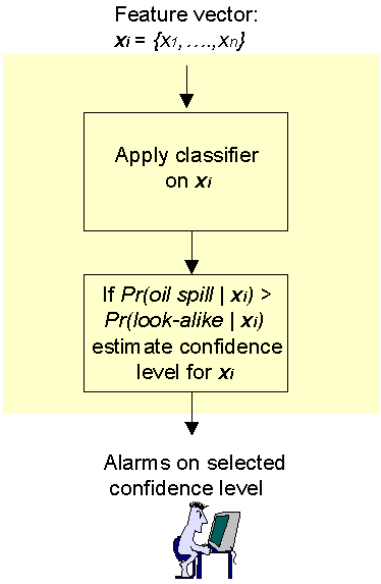

Figure 2. The two-step classification approach.

3. RESULTS

The results from estimating the subclass boundaries are presented in Sect. 3.2. The final classification results are presented in Sect. 3.3.

3.1. Data set

The data set applied in this study consist of 103 ENVISAT ASAR WSM images. The training set contains 76 scenes, while the test set contains 27 scenes. The images are collected mainly from the German and Finnish Baltic Sea and the German North Sea from 2003 to 2005. Aircraft verifications, collected during a combined satellite and aircraft campaign, were available for the test set.

3.2. Results - Optimal Subclass Division

Tab. 4 shows the classification results from the first step of the classification approach (see Fig. 2) for no subclasses, 2, 4, 6 and 8 subclasses. For $K = 8$,

Table 4. Classification accuracies on the test set applying no subclasses, 2, 4, 6 and 8 subclasses, and common regularized covariance matrices.

$K = 1$: **No subclasses.**		
	Classified as Oil spill	Classified as Look-alike
Marked as Oil Spill	39 (95.1%)	2 (4.9%)
Marked as Look-alike	1196 (9.8%)	11049 (90.2%)
$K = 2$: **2 subclasses.**		
	Classified as Oil spill	Classified as Look-alike
Marked as Oil Spill	39 (95.1%)	2 (4.9%)
Marked as Look-alike	1494 (12.2%)	10751 (87.8%)
$K = 4$: **4 subclasses.**		
	Classified as Oil spill	Classified as Look-alike
Marked as Oil Spill	39 (95.1%)	2 (4.9%)
Marked as Look-alike	1676 (13.7%)	10569 (86.3%)
$K = 6$: **6 subclasses.**		
	Classified as Oil spill	Classified as Look-alike
Marked as Oil Spill	39 (95.1%)	2 (4.9%)
Marked as Look-alike	1606 (13.1%)	10639 (86.9%)
$K = 8$: **8 subclasses.**		
	Classified as Oil spill	Classified as look-alike
Marked as Oil Spill	38 (92.7%)	3 (7.3%)
Marked as Look-alike	1607 (13.1%)	10638 (86.9%)

the optimiser was given $\theta = [p = 0.04, m_1 = 0.3, m_2 = 0.5, m_3 = 0.8, m_4 = 1.2]$ as initial values and returned $\hat{\theta} = [p = 0.1285, m_1 = 0.2513, m_2 = 0.3210, m_3 = 0.4646, m_4 = 2.5844]$ which should have given 10 subclasses. However some of the subclasses were empty or had very few samples. After merging some of the subclasses, 8 subclasses were left. For $K = 6$, $\theta = [p = 0.04, m_1 = 0.5, m_2 = 0.8]$ was applied as an initial guess, while the optimal estimate was $\hat{\theta} = [p = 0.1510, m_1 = 0.2996, m_2 = 0.5171]$. For $K = 4$, $\theta = [p = 0.04, m_1 = 0.5]$ was applied as the initial guess, and $\hat{\theta} = [p = 0.1510, m_1 = 0.3971]$ was the optimal estimate. For $K = 2$, $\theta = [p = 0.04]$ was applied as the initial guess, and $\hat{\theta} = [p = 0.1510]$ was the optimal estimate.

Tab. 4 shows that $K = 1$ (no subclasses) gives the best classification result. Fig. 3 presents the two oil spills misclassified as look-alikes when applying $K = 1$. Tab. 5 presents the number of training samples within each of the two classes and the regularization parameter ρ_{Ω_1}. As $\rho_{\Omega_1} = 0$, this means that there is enough data to get a good estimate of the common covariance matrices (see eq. 5) when applying $K = 1$ (all ρ_{Ω_k} were also estimated to be 0 for 2, 4 and 6 subclasses).

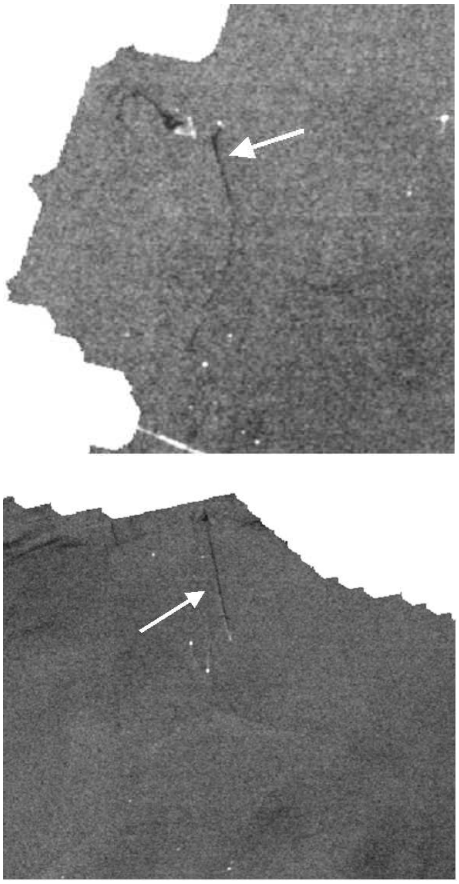

Figure 3. Oil spills classified as look-alikes when $K = 1$ (no subclasses) are applied.

3.3. Results - Confidence Estimation

Applying $K = 1$ (no subclasses), Tab. 6 shows the classification accuracies for the four confidence levels. The trend found in [2], with respect to the automatic confidence estimator, is that the surroundings of the detected slicks get more and more inhomogeneous and the number of look-alikes present increases for lower confidence levels.

4. CONCLUSION

In earlier studies [10, 2], the feature space was divided into 10 subclasses. The division boundaries were selected from experiments on the training set based on a manual approach. Here, we have searched for the optimal division boundaries based on an optimiser minimizing the negative log-likelihood function.

We have compared the classification accuracies from applying no subclasses with dividing the feature space into 2, 4, 6 and 8 subclasses. The results

Table 5. The number of training samples within each class and the estimated ρ_{Ω_1} when applying $K = 1$ (no subclasses).

	Training Set	
ρ_{Ω_1}	**Oil Spills**	**Look-alikes**
0	285	18779

Table 6. Classification accuracies on the test set.

High		
	Classified as Oil Spill	Classified as Look-alike
Marked as Oil Spill	13 (31.7%)	28 (68.3%)
Marked as Look-alike	12 (0.1%)	12233 (99.9%)
Medium		
	Classified as Oil Spill	Classified as Look-alike
Marked as Oil Spill	21 (51.2%)	20 (48.8%)
Marked as Look-alike	42 (0.3%)	12203 (99.7%)
Low		
	Classified as Oil Spill	Classified as Look-alike
Marked as Oil Spill	32 (78.0%)	9 (22.0%)
Marked as Look-alike	122 (1.0%)	12123 (99.0%)
Very Low		
	Classified as Oil Spill	Classified as look-alike
Marked as Oil Spill	39 (95.1%)	2 (4.9%)
Marked as Look-alike	1196 (9.8%)	11049 (90.2%)

show that no division of the feature space gives the best performance result on the test set.

As the final results from our two-step classification approach are comparable to earlier results where 10 subclasses were applied [2], this study shows that when there is enough data in the training set both a division into 10 subclasses and no subclasses could be applied in combination with a regularized classifier. A regularized classifier is preferred (compared to a classifier with diagonal covariance matrices as applied in early versions of the algorithm [10]) to avoid a low detection rate for oil spills and a larger amount of false alarms.

For *Low* confidence, 78.0% of the oil spills are here correctly classified while 99.0% of the look-alikes are correctly classified. This confidence level gives a good trade-off between detecting significant oil spills and having a low number of false alarms.

REFERENCES

1. Brekke, C. & Solberg, A. H. S. (2006). Segmentation and feature extraction for oil spill detection in ENVISAT ASAR images. Submitted.

2. Brekke, C. & Solberg, A. H. S. (2007). Classifiers and Confidence Estimation for Oil Spill Detection in ENVISAT ASAR Images. Submitted.

3. Del Frate, F. Petrocchi, A., Lichtenegger, J. & Calabresi, G. (2000). Neural Networks for Oil Spill Detection Using ERS-SAR Data. *IEEE Trans. on Geosci. and Remote Sensing*, **38**(5), 2282–2287.

4. Fiscella, B., Giancaspro, A., Nirchio, F., Pavese, P. & Trivero, P. (2000). Oil spill detection using marine SAR images. *Int. J. of Remote Sensing*, **21**(18), 3561–3566.

5. Keramitsoglou, I., Cartalis, C. & Kiranoudis, C. T. (2006). Automatic identification of oil spills on satellite images. *Environmental Modelling & Software*, 21, 640–652.

6. Nirchio, F., Sorgente, M., Giancaspro, A., Biamino, W., Parisato, E., Ravera, R. & Trivero, P. (2005). Automatic detection of oil spills from SAR images. *Int. J. of Remote Sensing*, **26**(6), 1157–1174.

7. Solberg, A. H. S. (2005). Automatic oil spill detection and confidence estimation. In *Proc. Int. Symp. Remote Sensing of Environment, St. Petersburg, Russia*, 943–945.

8. Solberg, A. H. S., Brekke, C. & Husøy, P. O. (2007). Oil spill detection in Radarsat and ENVISAT SAR images. *IEEE Trans. on Geosci. and Remote Sensing*, **45**(3), 746–755.

9. Solberg, A., Clayton, P. & Indregard, M. (2005). D2-Report on benchmarking oil spill recognition approaches and best practice. Tech. report, Oceanides project, EC, Archive No. 04-10225-A-Doc, Contr. No: EVK2-CT-2003-00177.

10. Solberg, A. H. S., Storvik, G., Solberg, R. & Volden, E. (1999). Automatic Detection of Oil Spills in ERS SAR Images. *IEEE Trans. on Geosci. and Remote Sensing*, **37**(4), 1916-1924.