

# A Comparative Study of Operational Vessel Detectors for Maritime Surveillance Using Satellite-Borne Synthetic Aperture Radar

Mattia Stasolla, Jordi J. Mallorquí, *Member, IEEE*, Gerard Margarit, Carlos Santamaria, and Nick Walker

**Abstract**—This paper presents a comparative study among four operational detectors that work by automatically post-processing synthetic aperture radar (SAR) images acquired from the satellite platforms RADARSAT-2 and COSMO-SkyMed. Challenging maritime scenarios have been chosen to assess the detectors' performance against features such as ambiguities, significant sea clutter, or irregular shorelines. The SAR images which form the test data are complemented with ground truth to define the reference detection configuration, which permits quantifying the probability of detection, the false alarm rate, and the accuracy of estimating ship dimensions. Although the results show that all the detectors perform well, there is no perfect detector, and a better detection system could be developed that combines the best elements from each of the single detectors. In addition to the comparison exercise, the study has facilitated the improvement of the detectors by highlighting weaknesses and providing means for fixing them.

**Index Terms**—Maritime security, maritime surveillance, satellite imaging, ship detection, synthetic aperture radar (SAR).

## I. INTRODUCTION

SHIP detection by post-processing satellite synthetic aperture radar (SAR) images is a remote sensing-based application that has been a subject of various studies [1]–[4]. Having practical applications in the security maritime domain, ship detection using SAR data has gained popularity thanks to the sensors' large swath widths and their ability to function during all weathers and during the night time as well as when there is

daylight. In addition, in many circumstances the backscattered signal from a ship is much larger than the sea background, so at the most basic level detection is achieved by searching for pixels with amplitudes greater than a given scattering threshold [5]. However, in reality there are several phenomena that add confusion to the overall process and bias the perfect model into something unpredictable [6], [7].

In recent years vessel detection performance has been improved by evolutions in technology, for example by increasing the number of information channels (through polarimetry), improving the resolutions (submetric resolutions are possible) and increased coverage. In addition, new platform and constellation configurations have been implemented to provide response to the more challenging demands that users are requesting, such as reduced latency times (between data acquisition and detection output) or detection capability for smaller ships of interest [8], [9].

Despite this technological evolution, ship detection is not currently yet free from complex artifacts that reduce performance for a range of scenarios, which are of interest for many users. The new techniques that have been proposed over recent years to exploit all the technological potentialities have provided a quality step forward in terms of robustness and probability of detection (PoD). However, they are not yet robust enough against ambiguities, strong scatterers, irregular shorelines or sea clutter [10]–[12].

This paper addresses these problems, which most adversely affect ship detection performance, and presents a comparative study among four operational ship detectors. The goal was to quantify the performance of each of them for a set of challenging configurations and propose improvement actions to overcome the drawbacks or at least to reduce their influence in the final results.

The work presented in this paper has been carried out in the framework of the FP7 NEREIDS project, an European Commission co-financed project with the main goal to improve ship monitoring techniques [13]. Four detectors have been analyzed: 1) Searching for Unidentified Maritime Objects (SUMO) from Joint Research Centre of the European Commission (JRC) [14]; 2) UPC-WT from Universitat Politècnica de Catalunya (UPC) [15]; 3) Ship Detection and Categorization Runtime (SIDE-CAR) from GMV Aerospace and Defence (GMV) [16]; and 4) R&G from eOsphere Limited (eOsphere) [17].

These are only some of the providers that can currently offer operational ship detection by satellite SAR. Nevertheless, the detection systems here analyzed can be considered a proper

Manuscript received September 30, 2015; revised March 10, 2016; accepted March 29, 2016. Date of publication June 25, 2016; date of current version July 05, 2016. This work was supported by the European Union Seventh Framework Programme (FP7/2007-2013) under Grant 263468 (NEREIDS). The work at UPC was also supported by the the Spanish Science, Research and Innovation Plan (Ministerio de Economía y Competitividad) under Grant TIN2014-55413-C2-1-P. RADARSAT-2 Data and Products ©MacDonald, Dettwiler and Associates Ltd. 2013 - All Rights Reserved – RADARSAT is an official mark of the Canadian Space Agency, provided under COPERNICUS by the European Union. COSMO-SkyMed images ©ASI 2013, distributed by e-GEOS S.p.A., provided under COPERNICUS by the European Union and ESA, all rights reserved. (*Corresponding author: Mattia Stasolla.*)

M. Stasolla and C. Santamaria are with the European Commission, Joint Research Centre, Institute for the Protection and Security of the Citizen, Maritime Affairs Unit, Ispra, Varese 21027, Italy (e-mail: mattia.stasolla@unipv.it).

J. J. Mallorquí is with the Remote Sensing Lab, Department of Signal Theory and Communications, Universitat Politècnica de Catalunya, Barcelona 08034, Spain.

G. Margarit is with the Remote Sensing Application and Services Division, GMV Aerospace and Defence, S.A.U., Barcelona 08006, Spain.

N. Walker is with the eOsphere Limited, Satellite Applications Catapult (R103), Harwell Space Cluster, Oxfordshire OX11 0QR, U.K.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JSTARS.2016.2551730

TABLE I  
SHIP DETECTION SYSTEMS

Name	Organization	Sensors supported	Land masking	Ship detection theory	Additional features
SUMO	Joint Research Centre, European Commission (JRC)	Envisat, ERS2, RADARSAT-1, RADARSAT-2, COSMO-SkyMed, TerraSAR-X	Shapefile with buffer zone	CFAR with K-distribution background	Azimuth ambiguities; Ships that are too long or too wide are rejected
SIDECAR	GMV Aerospace and Defense, S.A.U. (GMV)	Envisat, ERS2, RADARSAT-1, RADARSAT-2, COSMO-SkyMed, TerraSAR-X, ALOS-PALSAR	Shapefile with buffer zone; Shapefiles automatically extracted from image	Cluster-based approach based on Wavelet Transform (WT)	Azimuth ambiguities; Geodetic Active Contours to delineate target contour; Inclusion of target categorization with associated confidence
UPC-WT	Remote Sensing Lab (RSLab), Universitat Politècnica de Catalunya (UPC)	Envisat, ERS2, RADARSAT-1, RADARSAT-2, COSMO-SkyMed, TerraSAR-X, ALOS-PALSAR	Shapefile with buffer zone; Shapefiles automatically extracted from image	CFAR with Gaussian-distribution background based on WT	Azimuth ambiguities; Ships that are above or below a given size are rejected.
R&G	eOsphere Limited	Envisat, ERS2, RADARSAT-1, RADARSAT-2, COSMO-SkyMed, TerraSAR-X, ALOS-PALSAR	Shapefile with buffer zone	Median filters, to separate targets from background; Pauli decomposition channels for quad-pol	No Azimuth ambiguities; Histogram of the detected pixel for outliers and fitting an ellipse to the remaining detected pixels.

representative sample<sup>1</sup> so that their performance assessment would provide very useful information about the actual technological capacity in this field.

Moreover, the latest detailed performance assessment of ship detection systems reported in literature dates back to 2007 [18], when a comparative exercise among nine different detectors was run within the DECLIMS project to know what quality should have been expected at that time from satellite ship detection services, and to understand how to improve their products. The present work would therefore be helpful to update those results and provide new insights.

This paper is organized as follows. Section II provides a brief description of the four detection systems under analysis. In Section III, the methodology employed for the comparative study is discussed, focusing on the scope of the exercise and the evaluation procedure. Section IV is devoted to the description of the dataset employed for the analysis and the review of the results. Final comments are given in Section V.

## II. BACKGROUND

Table I summarizes the main characteristics, at the time of the study, of the four operational detectors under review, whose brief description is provided hereafter. As can be seen, the softwares are based on different detection algorithms and they can support a number of sensor/image products. They are also supplied with advanced processing features, such as land masking and the removal of azimuth ambiguities.

### 1) SUMO

The JRC has developed in the past years an algorithm for vessel detection called SUMO, shown in operation in Fig. 1. The SUMO ship detector adopts a fast version of a K-distribution constant false alarm rate (CFAR) detector [14]. SUMO has been implemented in Java and is available for all computer platforms

<sup>1</sup>Given that each detector has its own specific features, in many cases, the detection theories overlap.

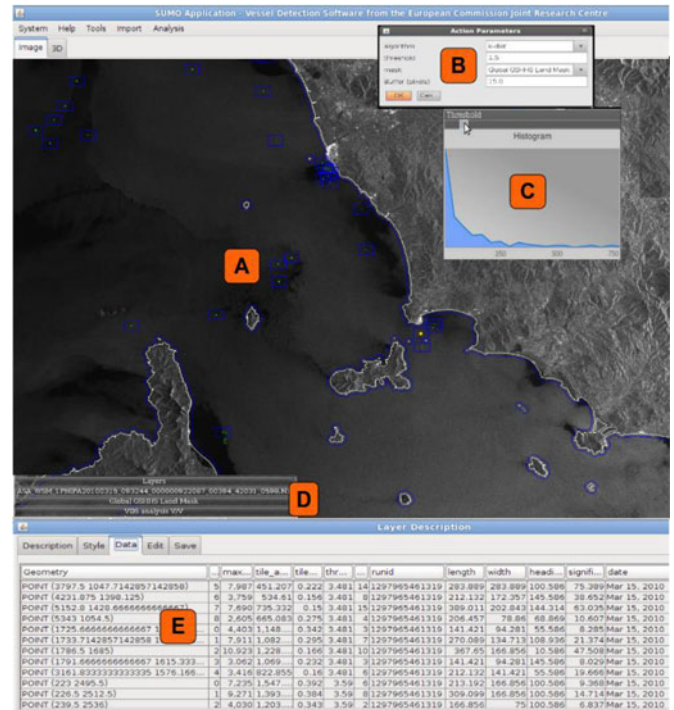


Fig. 1. SUMO: Snapshot in operation. (A) Main SUMO application window and in blue squares are shown the detected vessels, (B) K-distribution threshold and buffer zone dialog entry parameters, (C) Histogram versus threshold window which shows the impact of threshold selection on the detected targets, (D) Process layers, namely the Image Name, the Global Self-consistent Hierarchical High-resolution Shoreline (GSHHG) land mask for this image and the vessel detection analysis results which is shown in more detail in (E).

supporting the Java Virtual Machine [19]. The main advantage of the algorithm is the performance in terms of detection speed and robustness.

The K-distribution model depends on three parameters [20]: the mean, the number of looks, an instrument factor essentially describing the noise that is introduced by the radar image, and the shape factor, which determines the width of the distribution

and it reflects the actual variation of radar backscatter values on the sea surface. In SUMO the size of the subimage tile to estimate these parameters is of the order of  $200 \times 200$  pixels.

Once the parameters of the probabilistic distribution have been estimated, and a value for the probability of false alarm (PFA) is chosen, the false alarm threshold can be determined. Since the estimation of the unknown parameters might be biased by the presence of high-valued target pixels in the tile, these ones are excluded from the calculation. This operation is named clipping. In practice, the 1% highest pixels in the tile are discarded from the mean and order parameter computation. However, strictly applying the threshold calculated after the clipping process gives many false alarms. Presumably this is because the K-distribution does not perfectly represent the sea clutter in all situations, e.g., when the presence of ocean waves or atmospheric phenomena bias the tile statistics. Therefore a second, higher, threshold is defined based on a significance parameter that measures the difference between the pixel value and the mean of the tile.

In order to avoid detection of inland targets, land masking is applied and SUMO uses the publicly available GSHHG database [21]. The land mask is not always accurate, depending on several factors, such as the original source and scale of the shoreline information, tidal differences, etc. Hence, a buffer zone having a constant distance from the land (shore) is applied and the detection takes place outside the masked and buffered zones.

Following the detection, pixel clustering is applied and detected pixels are agglomerated into targets (vessels). Then, for each vessel, the number of pixels, position, length, width and heading are estimated. Finally, detections likely to be azimuth ambiguities are removed by using the SAR parameters. This step makes use of the deterministic distance that separates an ambiguity from the scatterer that originates that ambiguity. If a detection is located at that ambiguity distance from a much stronger scatterer, that detection is classified as an ambiguity and removed.

At the moment of the exercise, SUMO supported the following SAR imagery formats: Envisat Images and ERS2, RADARSAT-1 (CEOS format), RADARSAT-2, TerraSAR-X, COSMO-SkyMed (later on Sentinel-1 and ALOS-2 have been added).

The detection output has the form of comma separated values (csv), shapefile, xml, gml (KSAT schema), kmz (Google Earth), postgis, or thumbnails.

## 2) UPC-WT

UPC has developed during the last years an algorithm for vessel and coast detection that uses the wavelet transform (WT) through a multiscale analysis based not only on the intensity on SAR images but considering also its localized statistical behavior. At fine scales, the WT provide information about the variation of a function around a point. Thus, irregularities (edges) are sharpened in the particular direction of each subband (i.e., vertical, horizontal, and diagonal). The WT does not remove from the subbands the most local dependencies due to regular spatial structures and patterns. The different scales and bands are combined differently depending on the application.

*Vessel Detection:* The ship is usually noticeable in the three subbands after the two-dimensional (2-D) WT, but the values of the pixels of the background clutter are in general randomly distributed. For vessel detection, the algorithm works by exploiting these characteristics by spatially multiplying the four components (low-pass filter version of the original image plus three subbands) resulting from each iteration and proceeding to the detection directly in the wavelet domain. This is known as intrascalar processing of the WT. With this technique there is an enlargement of the dynamic range due to the multiplication process involved. The histogram of the resulting image is adjusted to a Gaussian, it has been seen that this distribution adjusts very well to the sea clutter after the WT process quite independently of the original clutter distribution. Depending on the selected PFA the proper threshold is selected, as in a classical CFAR approach. A complete description of the algorithm can be found in [15] and [22].

If the coast is not perfectly masked, it has an important influence when calculating the detection thresholds under a CFAR approach. The presence of the coast modifies the histogram and thus the threshold. The adjustment of the histogram to a double Gaussian helps to reduce the chances of an incorrect estimation of the threshold.

All detections are post-processed with morphological filters [23] to eliminate targets which size makes them to be non-realistic vessels (for instance a large non-masked island or too small targets) and an ambiguity detector to eliminate those detections associated to ambiguities, either from vessels or coastal features. The implemented algorithm also includes modules to make some improvements in the image to facilitate the classification of the detected vessels. For instance, the singular value apodization (SVA) is a nonlinear filter able to eliminate the side-lobes of strong scatters visible in low reflectivity areas [24]. The software also includes tools for vessel size and orientation.

*Land Masking:* The inaccuracies of GSHHG shoreline public databases can produce many false alarms when working with images containing coast. Usually, the way to deal with such inaccuracies is to expand the coastline towards the sea, for instance 100 m. This solution can be effective with simple coastlines but it fails with convoluted ones. One option is to extract the coastline from the image itself. UPC-WT uses the properties of the interscalar processing of WT to enhance the edges to carry out a coastline extraction [25]. Fig. 2 shows the final result with the extracted coastline. As can be seen, it perfectly matches the SAR image and it is more precise than the GSHHG one.

The detection report for UPC-WT can be provided in two output formats: csv and kmz.

## 3) SIDECAR

GMV has developed a Ship Monitoring Service (SIMONS) that provides maritime surveillance by automatically post-processing any type of Earth Observation (EO) image and Automatic Identification System (AIS)<sup>2</sup> streams [16]. The EO-based ship detection and categorization stage is covered by SIDECAR that can post-process either any type of SAR (any

<sup>2</sup> AIS is an automatic tracking system introduced for identifying and locating vessels [26].



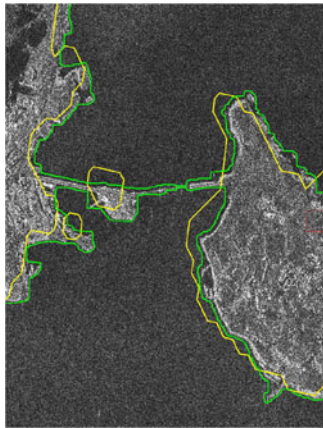


Fig. 2. UPC-WT: WT coastline extraction (green) compared with GSHHG coastline (yellow).

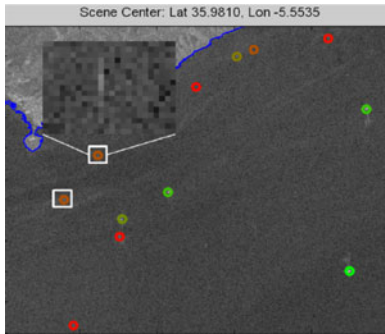


Fig. 3. SIDECAR: Snapshot of the ship detection report and coastline isolation (blue line). Confidence is color-coded from red (0—minimum) to green (1—maximum).

imaging mode, sensor, and format) or optic image. SIDECAR has been designed to be modular, flexible, reliable, and unsupervised (no human operator support). Four main high-level stages can be differentiated:

*Data Acquisition:* This stage consolidates input data into system repository.

*Information Extraction:* This stage isolates and estimates all the information that is needed from EO images. There are three main submodules:

- 1) *Coastline Delineation (CD):* The module isolates the coastline present in a SAR image by using external shape files and/or an algorithm that combines WT [27] and Geodesic Active Contours (GAC) [28], [29]. WT enhances the scattering information of land, while GAC delineates the contour via an energy-based minimization function.
- 2) *Ship Detection (SD):* The module detects the ships present in a EO image and derives the so-called Vessel Detection System (VDS) reports, as shown in Fig. 3. SAR images are post-processed via a multiscale clustering method based on WT. Differently from UPC-WT (which makes use of WT only), here the detector first combines the intermediate WT products by applying a pointwise

multiplication among them, and then identifies clusters of pixels that allow to detect targets regardless of their proximity and morphology. This also permits to estimate a confidence parameter that quantifies how reliable a detected target is. Finally, azimuth ambiguities are removed.

- 3) *Ship Categorization (SC):* The module analyzes the distribution of the reflectivity values along the SAR ship signature to make ship categorization via fuzzy logic (FL) [30]. Whenever possible, a wake detector is used to determine ship speed and course, increasing the discrimination sensitivity.

*Added-Value:* This stage provides added-value features:

- 1) AIS processing that correlates ship detection reports with AIS polls. FL is used to quantify the correlation percentage and refine the SAR-based VDS position by compensating the slant-range projection and the azimuth shifts caused by target dynamics.
- 2) Anomaly detection by processing cooperative data and the VDS reports.

*Result Dissemination.* This stage delivers final products to users via a large set of output formats and channels. Plain text files or complex xml files can be selected. Automatic integration with SIMONS is possible through dedicated databases.

#### 4) R&G

The R&G ship detection algorithm developed at the eOsphere Limited was originally designed to exploit fully polarimetric data, such as that available from the RADARSAT-2 satellite. However, the algorithm was also adapted for use with dual and single polarization data [17] that was available as a part of the NEREIDS project. For the single polarization case, i.e., for the results presented in this paper, the algorithm reduced to the use of a median filter followed by a series of morphological operations.

In order to find an estimate of the background signal from the ocean, each pixel under consideration is surrounded by a guard area to ensure that no pixel of an extended target could be considered as a part of the background. The width of the guard region is set by the user to a suitable value based on the expected size of possible vessels. In order to counteract the influence of outlying values, the median value of the background values is calculated and compared against the pixel under consideration. The calculation of the median is generally slower than for the mean; however, the resulting algorithm provides more reliable results.

Pixels are designated as “detected” when they are more than  $X$  times that of the median of the values from the background. During the course of the experiment a range of different values of  $X$  were tested and a value of  $X = 3$  was found to give a good level of target detection versus unwanted false alarms.

The resulting binary detection image is then filtered using morphological operations [31] to remove features in the detection image that would be unlikely to occur in reality.

The first morphological operation is the *fill* function, which fills isolated interior pixels, i.e., individual 0s surrounded by

1s. The second morphological operation is the *bridge* function, which sets 0-valued pixels to 1 if they have two nonzero neighbors that are not connected. Moreover, in order to prevent the detection algorithm from being applied to land areas, a masking procedure is employed utilizing the freely available GSHHG shoreline.

The output from the “classification” module is generated by extracting properties from the connected regions identified in the detection image, such as the area, centroid, length, width and orientation, and it can be made available as csv files or shapefile format.

### III. METHODOLOGY

For operational services, ship detection is normally carried out in a semi-automated way, i.e., the SAR images are first automatically analyzed and then passed to a human operator that validates and corrects the detection report. This provides a reasonable trade-off between quick response time and high accuracy. Nevertheless, as the volume of data to be processed is increasing more and more, the current need is to offer reliable and effective services in a fully automatic way.

The idea behind this exercise is therefore to assess the performance of the four detection systems described in the previous section when they are operating without human intervention. The detectors have been run in a fully automatic way, meaning that manual global adjustments (e.g., manually setting a certain value for a global threshold or choosing a certain buffer area from the coast) are acceptable, but manually editing the detected targets (e.g., adding or deleting targets) is not allowed.

Along with the performance assessment in terms of detection accuracy, the scope was also to better understand the strong/weak points of the detectors.

To this end we chose a set of images showing some of the typical features that characterize SAR maritime images. Specifically, we were interested in assessing the detectors’ behavior in the presence of the following five challenging situations:

- 1) *Ambiguities*: Ambiguities are an aliasing effect caused by the periodic sampling of the scene backscatter inherent to pulsed radar systems [32]. In SAR images this effect results in “ghost” replicas (red circles) of a real target (blue circle) at fixed positions, as shown in Fig. 4(a). To avoid false alarms, such artifacts should be properly discarded.
- 2) *Coastline*: When the scene spotted is a coastal area, the inland targets produce a large number of undesired detections, that need to be masked. Nevertheless, the effectiveness of the filtering may significantly vary depending on the type of the land mask used (e.g., an external coastline vector file, such as GSHHG shapefile, or image-based coastline discrimination); thus it is fundamental to test the capabilities offered by the detectors. Fig. 4(b) shows a clear example of mismatch between the external source employed for land masking and the actual coastline profile, that caused the detection of inland targets (red marks).
- 3) *Large Targets*: Another typical issue is the over-detection of ships due to extended targets. When a vessel covers

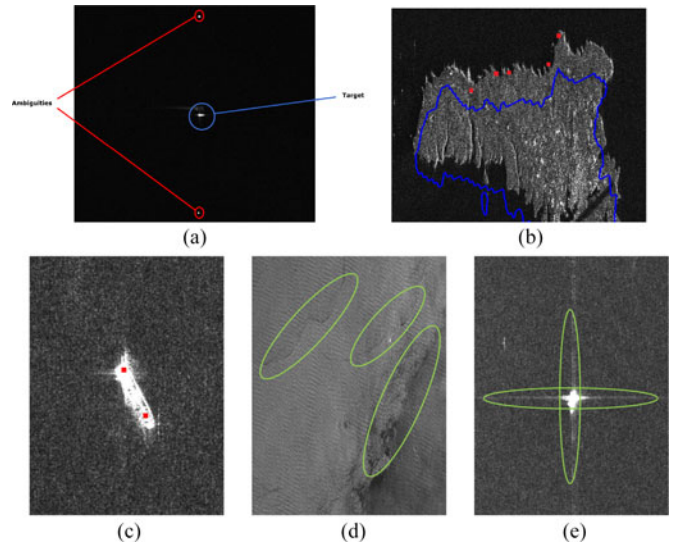


Fig. 4. Challenging situations. (a) Ambiguities. (b) Coastline. (c) Large targets. (d) Sea clutter. (e) Sidelobes.

a wide area, e.g., in high resolution SAR images, it may happen that different parts of its signature are detected as separate targets. An example is shown in Fig. 4(c), where both bow and stern of a tanker (red marks) have been counted as detections.

- 4) *Sea Clutter*: with the generic term of sea clutter we here refer to some of the most common features in SAR images related to the maritime environment, such as ocean waves, wind fronts, rain cells, surfactants, etc., [33]. As can be seen in Fig. 4(d), these features severely alter the image statistics, potentially limiting the accuracy of the detection systems.
- 5) *Sidelobes*: In SAR imaging, energy from strong scatterers can be detected at some distance away from their locations due to the existence of antenna sidelobes [34]. This energy typically appears as streaks in the range/azimuth directions and reduces the efficiency of automated target recognition algorithms. In addition, a bright scatterer may potentially mask the radar return from nearby, less intense scatterers.

As the scope of this exercise was to evaluate whether the detectors—run in a fully automatic way—have the capability to provide results comparable to those currently provided in operational activities by a human operator, we have compared their outputs to a manual ground truth (GT) created by visual inspection of the images. To obtain this GT, the operator has verified (visually and according to AIS data available) all the detections provided by the four systems and manually added those targets (e.g., small ships) that were not detected.

In particular, to quantitatively assess the accuracy of the detectors, the following metrics have been used:

- 1) PoD, defined as

$$\text{PoD} = \frac{\text{ND}}{\text{NT}} \times 100 \quad (1)$$

TABLE II  
DATASET DETAILS (I)

Image	Sensor	Product	Mode	Polarization	Resolution [m]	Targets (NT)	Pixels (NP)
CSK_20130308_050028	COSMO-SkyMed	DGM_B	HI	HH	5×5	99	18 545×19 171
RS2_20130318_054517	RADARSAT-2	SLC	WF	HH	5.2×7.7	80	15 801×29 359
CSK_20130606_045549	COSMO-SkyMed	SCS_B	HI	HH	3×3	19	19 604×20 188
RS2_20130606_050338	RADARSAT-2	SLC	S	HH	13.5×7.7	29	6778×21 069
RS2_20130610_044645	RADARSAT-2	SGF	S	HH	26.8×24.7	5	8215×8561

TABLE III  
DATASET DETAILS (II)

Image	Detector	Ambiguities	Coastline	Large targets	Sea clutter	Sidelobes
CSK_20130308_050028	SUMO, SIDECAR, UPC-WT	X		X	X	X
RS2_20130318_054517	SUMO, SIDECAR, UPC-WT, R&G		X	X	X	
CSK_20130606_045549	SUMO, SIDECAR, UPC-WT	X	X	X	X	
RS2_20130606_050338	SUMO, SIDECAR, UPC-WT, R&G	X	X	X	X	
RS2_20130610_044645	SUMO, SIDECAR, UPC-WT, R&G	X		X	X	X



Fig. 5. CSK\_20130308\_050028: Overview.

2) FAR), defined as

$$FAR = \frac{NF}{NP} \quad (2)$$

where ND is the number of correct detections, NF is the number of false alarms, NT is the number of GT targets, and NP is the number of image pixels.

To have a clearer picture of the specific weaknesses of the detectors, we have also disaggregated the false alarms with respect to the type of challenging scenarios that caused them.

In addition to the assessment of the detection performances, part of the exercise is also dedicated to the analysis of the detectors' capabilities to estimate the length, width and heading of the vessels. The estimations have been compared against the actual values derived from AIS data.

TABLE IV  
CSK\_20130308\_050028: RESULTS

	SUMO	SIDECAR	UPC-WT	(R&G)
ND	92	68	93	–
NF	0	5	32	–
PoD	92.93%	68.69%	93.94%	–
FAR	<2.81E–09	1.41E–08	9.00E–08	–

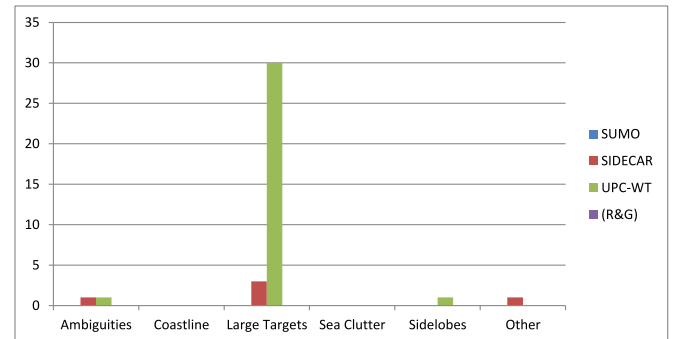


Fig. 6. CSK\_20130308\_050028: Number of false alarms versus challenging situations.

#### IV. RESULTS

The performance assessment of the detectors has been carried out over a set of five images, two COSMO-SkyMed (CSK) images and three RADARSAT-2 (RS2) images. They have been acquired over the Atlantic Ocean off West Africa in March 2013 and over the Mediterranean Sea in June 2013.

For the sake of clarity, in Table II we have reported the main products' details. Table III lists for each image the detectors used for the analysis and the challenging situations to deal with (marked with an "X"). As can be seen, all of them are single



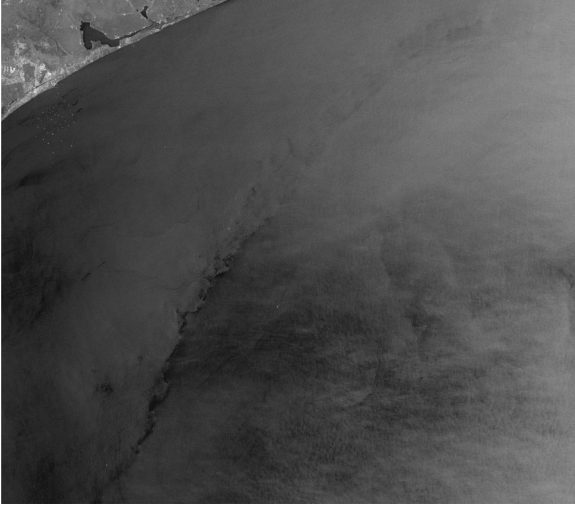


Fig. 7. RS2\_20130318\_054517: Overview.

polarization (HH), acquired in different modes and with a spatial resolution ranging from 3 m to about 27 m.

#### A. Vessel Detection

1) *CSK\_20130308\_050028*: The first image is a multi-look detected CSK single-pol HIMAGE product, with resolution of  $5\text{ m} \times 5\text{ m}$ . It has been acquired in March 2013 over the Bight of Bonny, the easternmost part of the Gulf of Guinea (see Fig. 5). The scene is characterized by areas with marked differences in sea clutter intensity, along with the presence of large targets, sidelobes, and ambiguities. The number of verified vessels is 99.

The results reported in Table IV show that all the three detectors (the image was not processed with R&G) achieved good values of PoD, with SUMO and UPC-WT even over 90%. The FAR is below  $2.81\text{E}-09$  for SUMO (this is the value of FAR corresponding to one single false alarm), for SIDECAR it is in the order of  $1\text{E}-08$  and for UPC-WT in the order of  $1\text{E}-07$ .

Looking at the histogram<sup>3</sup> of disaggregated false alarms in Fig. 6, it can be noticed that UPC-WT and SIDECAR have 30 and 3 false alarms related to large targets, respectively. The reason is that some of the big vessels in the image have been detected as multiple targets. This happened also to UPC-WT for one ship signature featuring strong sidelobes. As regards ambiguities, only a few errors have been reported for SIDECAR and UPC-WT. SUMO was able to cope with all the challenging situations. It is worth noting that the sea clutter apparently did not affect the analysis.

2) *RS2\_20130318\_054517*: The second image of the dataset is a single-look complex Wide Fine RS2 product, still from the Gulf of Guinea, namely its western part off Lomé, Togo (see Fig. 7). The spatial resolution provided is  $5.2\text{ m} \times 7.7\text{ m}$ . In this scene, taken 10 days after the previous one, the main

TABLE V  
RS2\_20130318\_054517: RESULTS

	SUMO	SIDECAR	UPC-WT	R&G
ND	78	33	73	73
NF	9	30	6	0
PoD	97.50%	41.25%	91.25%	91.25%
FAR	$1.94\text{E}-08$	$6.47\text{E}-08$	$1.29\text{E}-08$	$< 2.16\text{E}-09$

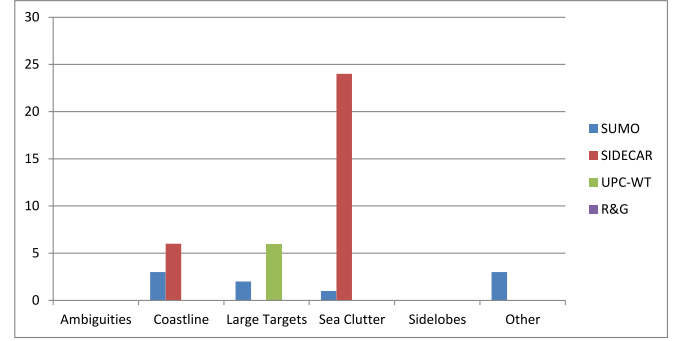


Fig. 8. RS2\_20130318\_054517: Number of false alarms versus challenging situations.

challenging situations are the presence of large targets and sea clutter. Moreover, many of the 80 GT vessels were moored at the port, so part of the coast had to be masked to avoid possible false alarms.

In Table V we can find the detection results. SUMO, UPC-WT and R&G detected almost all the targets, with PoD values higher than 90%. The number of false alarms for these three detectors is quite small, ensuring a FAR below  $2\text{E}-08$ . As regards SIDECAR, we can notice that it has a low PoD (about 40%) and a higher FAR (about  $6\text{E}-08$ ).

Such high FAR is basically due to sea clutter: as clearly shown by the strong red peak in Fig. 8, almost all SIDECAR's false alarms belong to the fourth group. The remaining ones are instead associated to coastline inaccuracy. From the same histogram, we can also see that the issues for the other detectors were different. SUMO was unable to mask some of the targets along the coast and, as well as UPC-WT, it did not properly handle large targets. With zero false alarms, R&G behaved well in all the situations.

3) *CSK\_20130606\_045549*: The third image is a single-look complex CSK HIMAGE product with a resolution of  $3\text{ m} \times 3\text{ m}$ . It was acquired over the Mediterranean sea off Lampedusa in June 2013. In this scene, which holds 19 targets, we can find all the challenging scenarios but sidelobes (see Fig. 9).

As reported in Table VI, the performances of all the three detectors (as for the previous CSK image, R&G was not run) are quite poor, with a maximum PoD of 52% reached by SUMO. Such a significant drop-off of the detection accuracy can be explained considering that many of the vessels in this area were small fishing ships, which, due to their dimensions, were more difficult to be detected. As regards the false alarms, SUMO

<sup>3</sup>Along with the five groups corresponding to the challenging situations, the histograms in this section also report a sixth column ("Other") which takes into account all those detections that could not be directly linked to the other groups or that could not be properly verified.

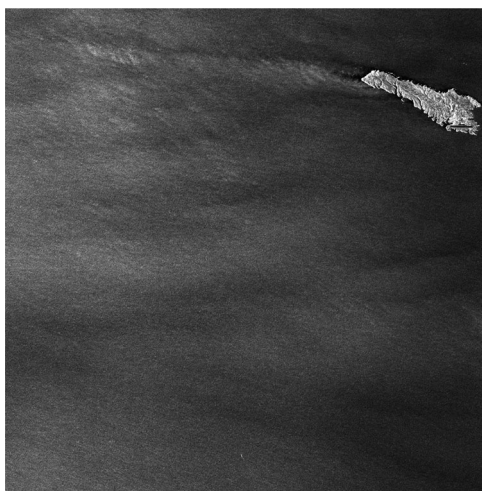


Fig. 9. CSK\_20130606\_045549: Overview.

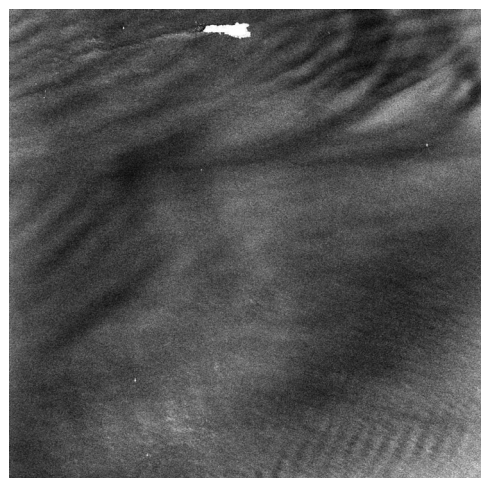


Fig. 11. RS2\_20130606\_050338: Overview.

TABLE VI  
CSK\_20130606\_045549: RESULTS

	SUMO	SIDECAR	UPC-WT	(R&G)
ND	10	6	9	—
NF	60	16	5	—
PoD	52.63%	31.58%	47.37%	—
FAR	1.52E−07	4.04E−08	1.26E−08	—

TABLE VII  
RS2\_20130606\_050338: RESULTS

	SUMO	SIDECAR	UPC-WT	R&G
ND	26	11	18	9
NF	8	8	10	1
PoD	89.66%	37.93%	62.07%	31.03
FAR	5.60E−08	5.60E−08	7.00E−08	7.00E−09

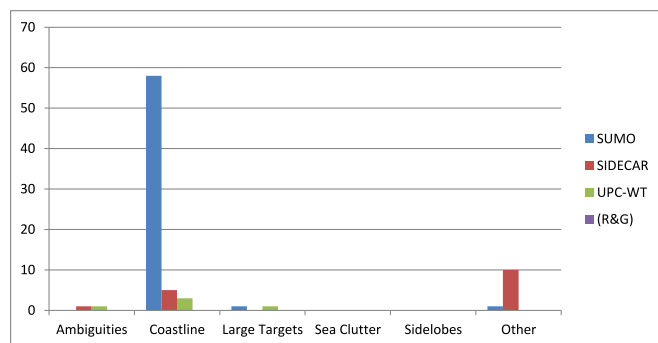


Fig. 10. CSK\_20130606\_045549: Number of false alarms versus challenging situations.

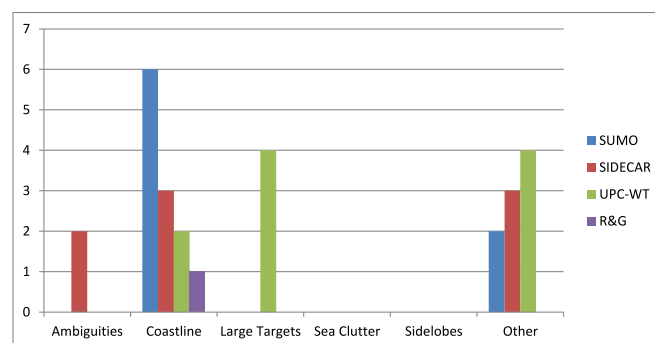


Fig. 12. RS2\_20130606\_050338: Number of false alarms versus challenging situations.

shows the highest number of errors, which caused a FAR of  $1\text{E}-07$ . The other two detectors could reach a rate about 10 times lower.

According to Fig. 10, SUMO's results were totally biased by the large number of inland targets that were not filtered out. The reason is that the external land mask used to fit the actual coastline was not sufficiently accurate in terms of both shape and georeferentiation. In fact, SIDECAR and UPC-WT, that are supplied with their own image-based coastline filters, could significantly avoid this problem.

4) *RS2\_20130606\_050338*: The fourth image is again an acquisition from the area over the island of Lampedusa (see Fig. 11), taken just a few minutes after the previous image. The

product is a single-look complex Standard RS2 image, with resolution of  $13.5\text{ m} \times 7.7\text{ m}$ . As before, sidelobes are the only challenging situation missing. Due to the wider area covered we have verified the presence of 10 more vessels.

As presented in Table VII, the best PoD has been achieved by SUMO (around 90%), followed by UC-WT (60%), and SIDECAR (38%). R&G obtained the lowest accuracy (31%). The number of false alarms is almost the same for SUMO, SIDECAR, and UPC-WT, which therefore all have a FAR around  $6\text{E}-08$ . R&G behaved better and could reach a FAR 10 times lower. As before, we can here ascribe the limited detection accuracy to the small dimensions of a number of observed targets.

The graph in Fig. 12 shows that, once again, SUMO suffered more than the other detectors from inland targets. SIDECAR,





Fig. 13. RS2\_20130610\_044645: Overview.

TABLE VIII  
RS2\_20130610\_044645: RESULTS

	SUMO	SIDECAR	UPC-WT	R&G
ND	3	4	5	4
NF	0	0	8	10
PoD	60.00%	80.00%	100.00%	80.00
FAR	$< 1.42E-08$	$< 1.42E-08$	$1.14E-07$	$1.42E-07$

along with a few false alarms in the coastline group, has two occurrences within ambiguities. UPC-WT included some inland targets but mainly did not recognize large targets. Only one false alarm (but out of 10 total detections) due to coastline inaccuracy for R&G.

5) *RS2\_20130610\_044645*: The last image is a multi-look detected Standard RS2 product, with a resolution of  $26.8\text{ m} \times 24.7\text{ m}$ . The area spotted, shown in Fig. 13, is located about 200 km West of Crete, and contains only five ships. The detectors had to cope with ambiguities, large targets, sea clutter, and sidelobes.

From the statistical point of view, missing even one single target in this image could bias significantly the results. This is well explained in Table VIII, where we can see that SUMO has a PoD of only 60%, as it did not detect two targets. SIDECAR and R&G missed only one target, so their PoD is 80%. The best performance is carried out by UPC-WT, which detected 100% of the targets. Looking at the false alarms, SUMO and SIDECAR have no wrong detections, while almost two thirds of UPC-WT and R&G's total detections are incorrect. This led to a FAR of about  $1E-07$ .

More in detail (see Fig. 14), UPC-WT did not properly manage the sea clutter, whereas R&G was hindered by ambiguities and extended targets.

6) *Overall Results*: In order to provide a more comprehensive evaluation of the performances, all the results from the single images have been merged to create a bigger unique dataset

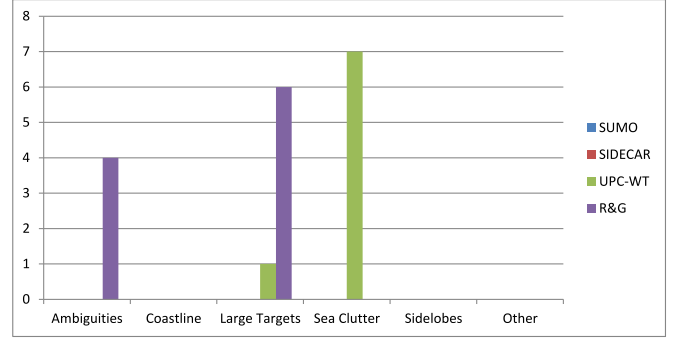
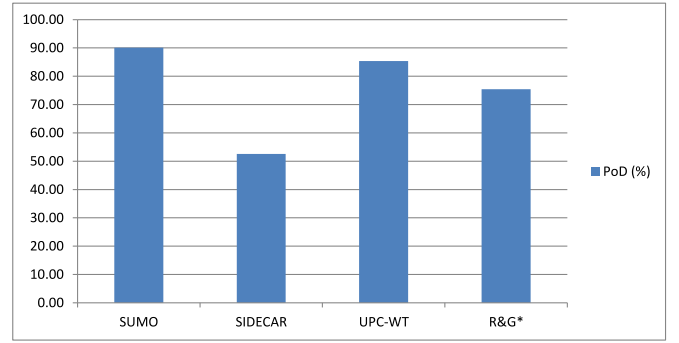
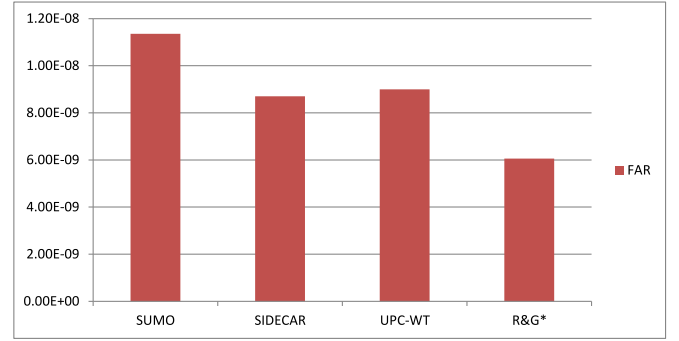


Fig. 14. RS2\_20130610\_044645: Number of false alarms versus challenging situations.



(a)



(b)

Fig. 15. Overall results (I). (a) Probability of detection. (b) False alarm rate. \*R&amp;G results have been normalized to the full dataset.

of 232 verified targets: 114 from the RADARSAT-2 images and 118 from the COSMO-SkyMed images.

All the metrics have been calculated accordingly, taking into account that R&G was not run on the COSMO-SkyMed subset (the corresponding graphs have been normalized by a factor of about 2).

In particular, we can see in Fig. 15(a) the performances of the four detectors in terms of PoD. SUMO is about 90%, SIDECAR is slightly above 50%, UPC-WT has 85% of correct detections, and R&G reaches 75%.

As regards the FAR, the results are visualized in Fig. 15(b). SUMO has a rate slightly above  $1E-08$ , SIDECAR and UPC-WT provide a value around  $8E-09$ , R&G was able to achieve  $6E-09$ .

To read the above results from a different point of view, we can observe the histogram in Fig. 16(a). It reports the number

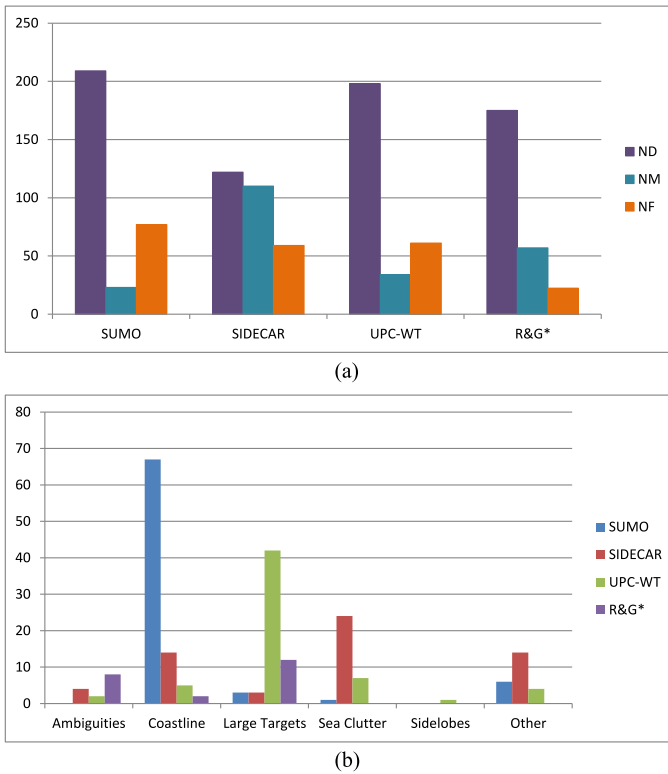


Fig. 16. Overall results (II). (a) Number of correct detections (ND); missed targets (NM); false alarms (NF). (b) Number of false alarms versus challenging situations. \*R&G results have been normalized to the full dataset.

of correctly detected targets (ND), the number of targets that have not been detected (NM) and the number of false alarms (NF). SUMO missed 23 targets out of 232, UPC-WT 34 targets and SIDECAR more than hundred (110). R&G did not detect 28 targets out of 114, which would mean 57 for the normalized dataset. These results are in line with the above mentioned percentages of PoD.

The additional information provided by the histogram is indeed the number of false alarms. SUMO output 77 wrong detections, SIDECAR and UPC-WT had around 60 false alarms, and R&G failed 22 times (normalized by a factor of 2). As done for the single images, it is interesting to give a look to the disaggregated histogram, in Fig. 16(b). What is immediately clear is that almost every group is characterized by a major error peak, i.e., every detector suffers more than the others a specific challenging situation.

As expected from the previous discussion, the most significant drawback for SUMO is the poor masking of coastal areas, which caused 67 out of 77 false alarms. For SIDECAR, the larger part of wrong detections is due to sea clutter (24 out of 59) and only partly to coastline (14). Once again, we can see that UPC-WT is mainly hindered by large targets, which produce multiple detections and result in 42 false alarms out of 61. R&G does not have a clear peak in any of the groups, but some extended targets were counted as separate targets and none of the ambiguities have been recognized. Finally, all the detectors suitably handle sidelobes, which have been reported as targets only in one case.

TABLE IX  
CHALLENGING SCENARIOS VERSUS DETECTORS' CAPABILITIES

Detector	Ambiguities	Coastline	Large targets	Sea clutter	Sidelobes
SUMO	X		X	X	X
SIDECAR			X		X
UPC-WT		X		X	X
R&G		X		X	X

### B. Combination of the Vessel Detection Systems

The results presented in the previous section show good performance for all of the detectors, but they also reveal that none of them could consistently cope with all the challenging scenarios analyzed.

In particular, SUMO has presented weakness against shorelines that are not properly delineated by the external files that are used to mask the land out. SIDECAR has shown difficulties to deal with structured clutter, such as the one caused by wave crests or wind-induced effects, and partly with coastline inaccuracies and ambiguities. UPC-WT had mainly problems with large vessels, which were often detected as multiple targets. That happened to R&G as well, which also suffered the lack of a tool for ambiguity discrimination.

Table IX schematizes the detectors' capabilities, marking with an 'X' the challenging situations where they have been proven to be effective.

As can be noticed, each of the detectors has its specific shortcomings, basically failing where the others are successful and vice versa. This suggests that the overall results could be improved if the best elements of each detector were combined.

In the past years, a number of strategies have been proposed in literature to merge the decision of different classifier to arrive at improved classification results. These include: statistical approaches [35], formulations based on the Bayesian theory, the Dempster-Shafer framework [36], neural networks [37]. In addition to them, a very basic yet very popular combination scheme is the so-called *majority vote* [38], where the final decision is taken according to what has been decided by the majority of the input classifiers. It is the simplest method for implementation, but it has been shown to be as effective as the other more complicated schemes [39]. Moreover, it does not assume prior knowledge of the behavior of the individual classifiers and it does not require training on large quantities of representative results [40].

Considering these upsides, and due to the particular nature of our problem (the combination of independent binary classifiers in order to reach a unique binary decision), we applied the majority vote approach to merge the outputs of the four detection systems into a combined classifier (from now we will refer to it as MAJORITY) and improve the overall performance.

The problem can be formalized as follows.

We want to combine the binary decisions  $[x_1, \dots, x_N]$  of  $N$  input detectors so that the final binary decision  $y$  represents the majority of the detection results. If we assign the value +1 to the event of a target detection, and the value -1 to a non-detection,

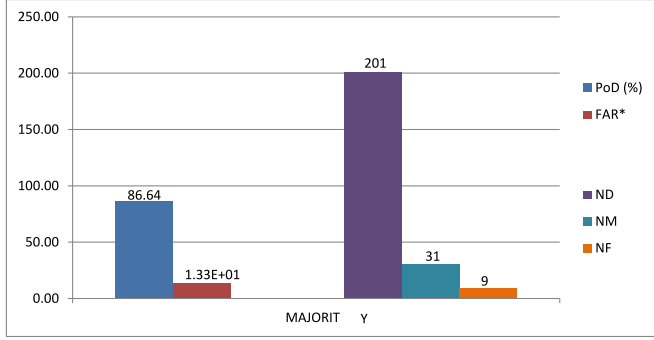


Fig. 17. Overall results: MAJORITY. \*For the sake of clarity, the actual value of FAR ( $1.33E-09$ ) has been magnified by a factor  $1E10$ .

the combined decision can be obtained as [41]:

$$y = \text{sign} \left( \sum_{i=1}^N x_i \right) \quad (3)$$

where  $x_i \in \{\pm 1\}$  is the independent decision of the  $i$ -th input detector.

Despite its very simple formulation, the majority vote scheme also gives the possibility to combine the input detectors in a more complex way, in order to better take into account their specific accuracies. To this end, it is simply needed to introduce for each input a weighting factor  $w_i$ , so that (3) becomes:

$$y = \text{sign} \left( \sum_{i=1}^N w_i x_i \right) \quad (4)$$

This is the so-called *weighed majority vote*, that could be for example used when the detectors provide a “confidence parameter” for their outputs or if the final decision should be biased by one (or more) specific input.

According to the discussion above, we therefore merged the detection outputs of SUMO, SIDECAR, UPC-WT and R&G by applying (3) in order to obtain the combined results for MAJORITY.<sup>4</sup> These have been then compared to the GT to calculate the corresponding PoD, FAR, ND, NM, and NF. If we look at the bars in Fig. 17, we can see that MAJORITY was able to detected 201 targets out of 232, and therefore it missed 31 targets. This means that it performs better than all the detectors but SUMO, as shown also by its PoD, which is close to 87%. As regards the false alarms, instead, we can see that the merging of the single outputs allowed to reduce the number of NF to 9, which is 2–8 times less than those of the other detectors. This allowed to reach a level of FAR which is significantly lower than any other.

We can therefore conclude that, even though MAJORITY could not provide the highest PoD (being anyway quite close), in terms of overall accuracy—high PoD and low FAR—it certainly provides the best performance.

<sup>4</sup>We recall that, for the COSMO-SkyMed subset  $N = 3$ , whereas for the RADARSAT-2 one  $N = 4$ . We opted to always take a majority decision based on the agreement of at least two detectors.

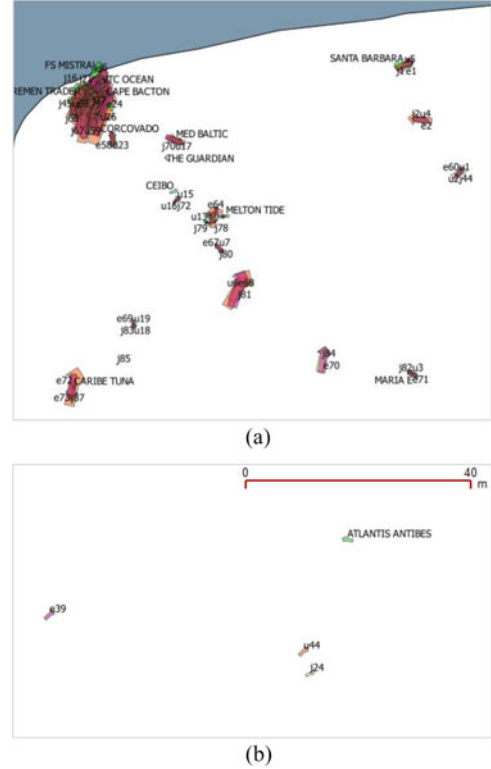


Fig. 18. RS2\_20130318\_054517: Detection results. The vessels are displayed as arrows, oriented according to the reported or detected heading and the length of each arrow is proportional to the reported or detected vessel length. (a) Overall detections. (b) Focusing on one vessel as reported by AIS (ATLANTIS ANTIBES) and detected by SUMO (j24), UPC-WT (u44) and R&G (e39).

### C. Vessel Length/Width/Heading Estimation

The SAR data acquired off the coast of Togo, image RS2\_20130318\_054517, was also exploited to test the accuracy of the additional ship parameters provided by the detection algorithms. From the AIS data transmitted by the many vessels in the region, we extracted the information about the actual ships’ length, width and heading, and we compared them against the estimations from SUMO, UPC-WT, and R&G (GMV results were not available).

Fig. 18(a) shows an overview of all the detections together with the results from the AIS data. The vessels are displayed as arrows, oriented according to the reported or detected heading and the length of each arrow is proportional to the reported or detected vessel length.

The number of recorded detections for the AIS system and for each of the algorithms is shown in Table X. It can be seen that the number of AIS-transmitting ships and the number of detections for each of the algorithms are fairly similar. Only 70 out of 80 verified targets transmitted AIS signals.

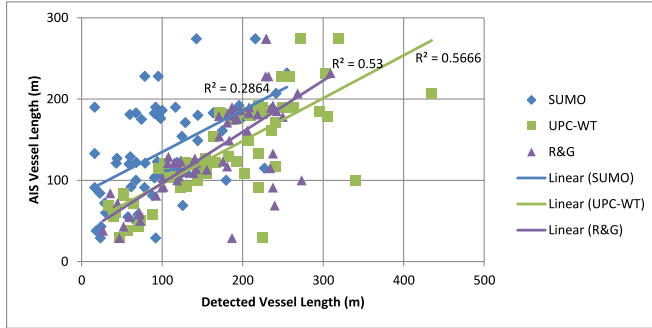
Fig. 18(b) shows the result of focusing on one vessel, the Atlantis Antibes, as reported by AIS and detected by SUMO, UPC-WT, and R&G.

Of the 70 vessels reported by the AIS, there were 63 which were also detected by all three SUMO, UPC-WT, and R&G. Of these 63 sets of data it was found that four were missing

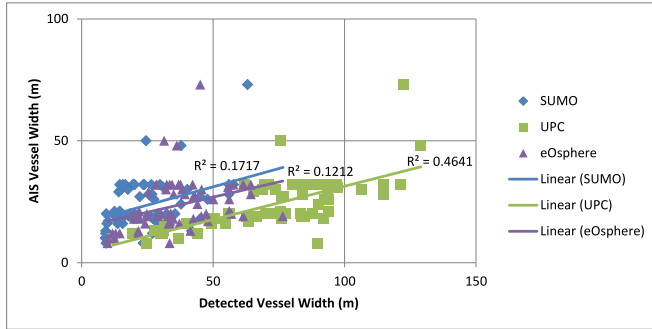


TABLE X  
RS2\_20130318\_054517: LENGTH/WIDTH ESTIMATION

	AIS	SUMO	UPC-WT	R&G
# of detections	70	78	73	73
# of correlated targets			63	
# of reliable targets			59	
Average length difference [m]		-28.83	41.50	24.66
Average length difference (absolute) [m]		46.44	47.00	31.73
Average width difference [m]		2.12	48.27	12.44
Average width difference (absolute) [m]		10.51	48.27	14.80



(a)



(b)

Fig. 19. RS2\_20130318\_054517: Length/width estimation. (a) Reported AIS vessels lengths versus algorithm detected vessel lengths. (b) Reported AIS vessels widths versus algorithm detected vessel widths.

some information from their AIS reports (usually the vessel dimensions), which made them unusable in the analysis.

This left 59 data sets to be included in the analysis where there was simultaneously an AIS report and detections from all three algorithms.

1) *Vessel Length*: Fig. 19(a) shows a scatterplot for the 59 AIS reports for the vessel lengths, versus the vessel lengths as measured by the three different algorithms. It can be seen that there is a degree of correlation between the reported and detected vessel lengths; however, if it is assumed that the AIS reported lengths are relatively accurate, then there is a degree of error from each of the algorithms.

Table X shows, for each of the three algorithms, the mean difference between the AIS reported vessel lengths and the detected vessel lengths, and also the mean of the absolute

difference between the AIS reported vessel lengths and the detected vessel lengths.

The fact that the average length differences between AIS and the algorithm are positive for UPC-WT, and R&G shows that these algorithms, in general, are overestimating the vessel lengths. This tendency can be observed in the scatterplot in Fig. 19(a), where the linear regression lines for these two algorithms are below the 45° diagonal line that would represent equal values of AIS and algorithm length.

In contrast, SUMO has a negative average length difference between itself and the AIS data, which indicates that, on average, it is underestimating the vessel lengths. From Fig. 19(a), it can be seen that this tendency to underestimate vessel lengths is more pronounced at smaller vessel sizes, but that the estimates become more accurate for larger vessels.

The mean of the absolute differences between the AIS reported vessel lengths and the detected vessel lengths, shown still in Table X, gives an overall summary of how accurate the detection algorithms are. It can be seen that R&G provides the most accurate results and that the mean of the averages of the absolute differences is 41.72 m, which indicates that there is room for improvement, especially because the average value of the reported AIS vessel lengths is 138.87 and therefore an error of 41.72 m represents a percentage of 30%.

2) *Vessel Width*: Fig. 19(b) shows a scatterplot for the 59 AIS reports for the vessel widths, versus the vessel widths as measured by the three different algorithms.

Table X shows, for each of the three algorithms, the mean difference between the AIS reported vessel widths and the detected vessel widths and also the mean of the absolute difference between the AIS reported vessel widths and the detected vessel widths.

It can be seen that all three algorithms are, on average, providing overestimates of the vessel widths; however, SUMO has a very low average, indicating that, in general, it is underestimating almost as often as overestimating. The means of the absolute differences provide an overall measure of the performance, where it can be seen that SUMO is the most accurate, followed by R&G, with the UPC-WT falling significantly behind.

The average absolute error for the three algorithms is 24.52 m, which, in fact, is greater than the average value of the vessel widths, 23.97 m, as reported by the AIS.

3) *Vessel Heading*: Fig. 20 shows a rose-diagram histogram of the vessel heading values for the results from the three algorithms and those received from the AIS signals.

It can be seen that the AIS heading values are relatively evenly distributed throughout the full range of angles. This is because the AIS heading values are derived from the course over ground (CoG) parameter of the AIS messages, and the fact that most of the vessels in this test set are stationary. The CoG signal for a stationary vessel will be an almost random value dependent on small motions of the vessel. Therefore, it was not possible to use the AIS data to evaluate the accuracy of the heading estimation. However, Fig. 20 is able to demonstrate there is a very high degree of correspondence between the three different detection

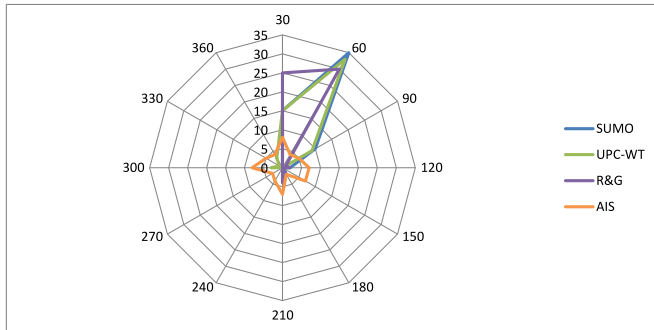


Fig. 20. RS2\_20130318\_054517: Heading estimation.

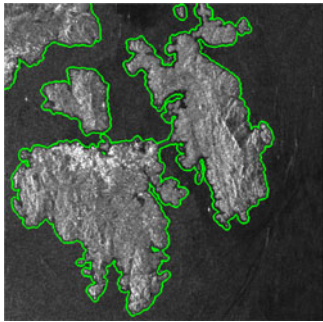


Fig. 21. SUMO improvements: Sentinel-1 (Copernicus data, 2015) image of the islands of Maddalena and Caprera in Sardinia, geocoded using orbit state vectors. In green, the OpenStreetMap landmask buffered by 100 m.

algorithms, where nearly all of the detected heading values are found to lie in the  $30^{\circ}$ – $60^{\circ}$  range.

This general effect was illustrated in Fig. 18(b) where, for the vessel Atlantis Antibes, the three detected vessels appear to be in the same direction, but the AIS signal indicates the vessel is almost at right-angles to the detected vessels.

#### D. Improvements

As we have already discussed, the results presented in Section IV-A have clearly pointed out the specific shortcomings of the single detectors. This helped to focus on targeted software developments that could improve the detection performances.

More in detail, the SUMO system has been supplied with an additional external shapefile for land masking: along with the GSHHG shoreline, it is now possible to load the data from OpenStreetMap, the collaborative project providing an openly licensed map of the world [42]. Moreover, a new geocoding algorithm based on orbit state vectors has been introduced in order to minimize the co-registration errors between the image and the mask [43]. An example of the improvements brought by this new feature is depicted in Fig. 21. GMV has improved their detection algorithm to better handle images with strong sea clutter: now the computation of the confidence parameter for the detection of the targets has been suitably adjusted to reduce false alarms, as shown in Fig. 22. UPC found a bug in the software which did not properly process some small detected blob-like features that could arise after the thresholding phase (see Fig. 23), causing the multiple detections. The problem was fixed by applying further morphological filters. As regards eO-

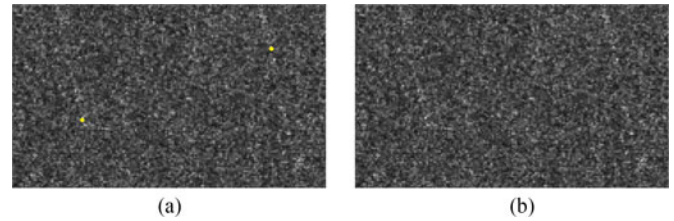


Fig. 22. SIDEAR improvements: (a) Snapshot of a RADARSAT-2 ultrafine image with 3 m of resolution where the sea clutter have generated false alarms (yellow dots). (b) Removal of the false alarms.



Fig. 23. UPC-WT improvements: (a) Image after scaling with the sea clutter statistics and thresholding: Disconnected areas belonging to the same vessel can be clearly seen. (b) Same image after applying the morphological filters: Disconnected areas have been joined.

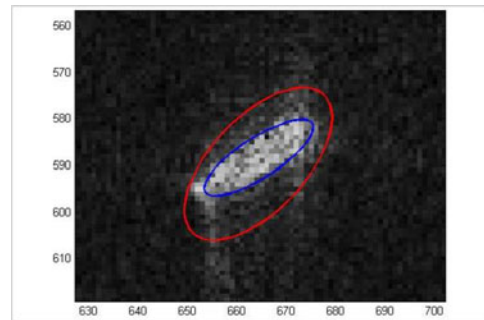


Fig. 24. R&G improvements: Reduction of smearing and side-lobes in target signature. Red ellipse has been fitted to the length and width derived using the initial algorithm and the blue ellipse has been fitted to the length and width of the output from the new module.

sphere, a new addition to the algorithm was incorporated which reduces the effect of smearing and side-lobes on the target signatures. This would have a two-fold effect: on one hand, to limit the number of false alarm in the presence of extended targets, and on the other hand to improve the estimation of lengths and widths. Fig. 24 shows an example of the difference between the algorithm before (red ellipse) and after the implementation of the sidelobe reduction module (blue ellipse).

In order to test these new software developments and measure their effectiveness, the upgraded versions of SUMO, SIDEAR, UPC-WT, and R&G have been rerun over the same dataset described in Section IV-A. In this way, it has been possible to directly compare the results, and have a clear and immediate picture of the improvements carried out.

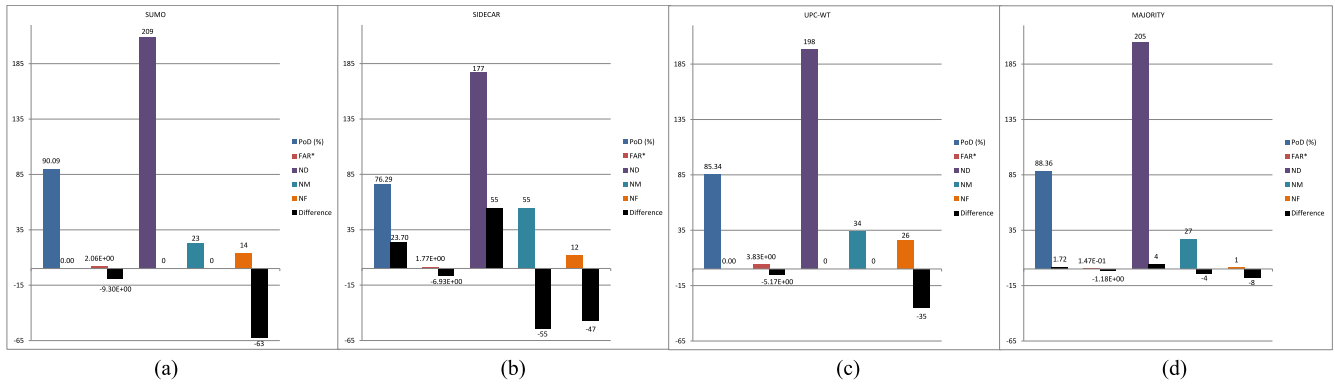


Fig. 25. Overall results: Upgraded detectors. (a) SUMO, (b) SIDECAR, (c) UPC-WT, (d) MAJORITY. \*For the sake of clarity, the actual values of FAR have been magnified by a factor  $1E09$ .

Fig. 25 shows, for each detector,<sup>5</sup> a column graph reporting—along with PoD, FAR, ND, NM, and NF—also a black bar that measures the difference between the current and the previous results, allowing to immediately visualize whether the upgraded detectors performed better (or worse) than the older versions.

As can be seen, for SUMO and UPC-WT the number of detected targets does not change (so it is for NM and PoD, then), but the number of false alarms drops off to 14 and 26, respectively. As a consequence, the FAR is reduced by a factor of 5.5 and 2.3, respectively. This clearly confirms that both the new geocoding algorithm (SUMO) and the morphological filtering of big targets (UPC-WT) brought the desired effects, significantly improving the performances. Nevertheless, the detector which experiences the most noticeable improvements is SIDECAR: the PoD increases from 52% to 76%, the number of missed targets is halved, and the false alarms are reduced to 12, producing a FAR of about  $2E-09$ .

To conclude the analysis we have also compared the results of the detector MAJORITY. The combination of the four new detectors results in an increase of the PoD, which raises to 88% and almost matches the maximum value reached by SUMO (they detect 205 and 209 targets, respectively). Yet, the most remarkable point is the number of false alarm: as can be seen, it is basically reduced to zero, providing a very low FAR level, which is about  $1.5E-10$ . Once again, we can therefore conclude that MAJORITY is the best detector in terms of overall accuracy.

## V. CONCLUSION

This paper has presented a comparative study among four operational ship detectors (JRC's SUMO, GMV's SIDECAR, UPC's UPC-WT, and eOsphere's R&G) which operate on satellite acquired SAR images.

The overall goal was to provide insights about the current technological capacity in automatic ship detection by satellite SAR. The analysis has shown that the detectors behave generally well, but none of them is completely robust to all of the challenging scenarios that have been chosen for the test, namely: SAR

azimuth ambiguities, coastline effects (especially for intricate coastlines), large targets (which sometimes become separated into multiple detected targets), sea clutter, and sidelobes (which can significantly alter the signature of the target).

Nevertheless, the study also pointed out that the detectors have specific shortcomings (they basically fail where the others are successful and vice versa) that make them complementary. This was the starting point for the implementation of a merging scheme based on the majority vote approach that would allow outputs from more than one algorithm to be combined to improve the overall results.

As regards the second part of the analysis, i.e., the estimation of the length and width of the vessels from SAR images, currently this can only be done with limited accuracy. This is important because it has a significant impact on the ability to successfully classify vessels. Further improvements are therefore needed, even though this task is strictly dependent on the resolution of the images, and may not be feasible in all cases. Concerning the heading estimation (i.e., the direction of the vessel), we can conclude that it would be possible to have reliable estimates for vessels that are significantly larger than the image resolution. Nevertheless, as that will be the estimated heading of the vessels at precisely image acquisition time, care is needed when comparing the SAR-derived heading with the AIS-reported heading (or CoG), especially if the ship is moving at a low speed.

Finally, it is worth noting that the conclusions of this study have served to oversee the development of more robust SAR based vessel detectors. Modifications in the detection algorithms have been successfully implemented to overcome their weaknesses and improve the overall performances.

## REFERENCES

- [1] K. Eldhuset, "An automatic ship and ship wake detection system for spaceborne SAR images in coastal regions," *IEEE Trans. Geosci. Remote Sens.*, vol. 34, no. 4, pp. 1010–1019, Jul. 1996.
- [2] R. Touzi, R. Raney, and F. Charbonneau, "On the use of permanent symmetric scatterers for ship characterization," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 10, pp. 2039–2045, Oct. 2004.
- [3] S. Brusch, S. Lehner, T. Fritz, M. Soccori, A. Soloviev, and B. van Schie, "Ship surveillance with TerraSAR-X," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 3, pp. 1092–1103, Mar. 2011.

<sup>5</sup>As the upgraded version of R&G did not provide results different from those previously presented, the corresponding graphs have not been here replicated.



- [4] D. Tao, S. Anfinssen, and C. Brekke, "Robust CFAR detector based on truncated statistics in multiple-target situations," *IEEE Trans. Geosci. Remote Sens.*, vol. 1, no. 54, pp. 117–134, Jan. 2016.
- [5] M. Skolnik, *Radar Handbook*, (Electronics Electrical Engineering), 3rd ed., New York, NY, USA: McGraw-Hill Education, 2008.
- [6] V. Kerbaol and F. Collard, "SAR-derived coastal and marine applications: From research to operational products," *IEEE J. Ocean. Eng.*, vol. 30, no. 3, pp. 472–486, Jul. 2005.
- [7] M. Vespe and H. Greidanus, "SAR image quality assessment and indicators for vessel and oil spill detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 11, pp. 4726–4734, Nov. 2012.
- [8] F. Caltagirone *et al.*, "The COSMO-SkyMed dual use earth observation program: Development, qualification, and results of the commissioning of the overall constellation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 7, pp. 2754–2762, Jul. 2014.
- [9] R. Torres P. Snoei, D. Geudtner, D. Bibby, M. Davidson, E. Attema, P. Potin, B. Rommen, N. Flourey, M. Brown *et al.*, "GMES Sentinel-1 mission," *Remote Sens. Environ.*, vol. 120, pp. 9–24, 2012.
- [10] G. Di Martino, A. Iodice, D. Riccio, and G. Ruello, "Filtering of azimuth ambiguity in stripmap synthetic aperture radar images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 9, pp. 3967–3978, Sep. 2014.
- [11] A. Buono, F. Nunziata, L. Mascolo, and M. Migliaccio, "A multipolarization analysis of coastline extraction using X-band COSMO-SkyMed SAR data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 7, pp. 2811–2820, Jul. 2014.
- [12] X. Zhao, Y. Jiang, and W.-Q. Wang, "Efficient clutter suppression in SAR images with shedding irrelevant patterns," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 9, pp. 1828–1832, Sep. 2015.
- [13] [Online]. Available: [www.cordis.europa.eu/project/rcn/99070\\_en](http://www.cordis.europa.eu/project/rcn/99070_en). Accessed on Mar. 10, 2016.
- [14] N. Kourti, I. Shepherd, G. Schwartz, and P. Pavlakis, "Integrating spaceborne SAR imagery into operational systems for fisheries monitoring," *Can. J. Remote Sens.*, vol. 27, no. 4, pp. 291–305, 2001.
- [15] M. Tello, C. Lopez-Martinez, and J. Mallorqui, "A novel algorithm for ship detection in SAR imagery based on the wavelet transform," *IEEE Geosci. Remote Sens. Lett.*, vol. 2, no. 2, pp. 201–205, Apr. 2005.
- [16] G. Margarit, J. A. Barba Milans, and A. Tabasco, "Operational ship monitoring system based on synthetic aperture radar processing," *Remote Sens.*, vol. 1, no. 3, pp. 375–392, 2009. [Online]. Available: <http://www.mdpi.com/2072-4292/1/3/375>
- [17] A. Marino, N. Walker, and I. Hajnsek, "Perturbation analysis for maritime applications," in *Proc. 9th Eur. Conf. Synthetic Aperture Radar*, Apr. 2012, pp. 509–512.
- [18] H. Greidanus, "Vessel classification benchmark," *European Commission – JRC Tech. Rep. DECLIMS D4*, 2007.
- [19] B. Venners, *Inside the Java Virtual Machine*, 1st ed., New York, NY, USA: McGraw-Hill, 1999.
- [20] C. Oliver and S. Quegan, *Understanding Synthetic Aperture Radar Images*. Raleigh, NC, USA: SciTech Publ., 2004.
- [21] P. Wessel and W. H. Smith, "A global, self-consistent, hierarchical, high-resolution shoreline database," *J. Geophys. Res.—Solid Earth (1978–2012)*, vol. 101, no. B4, pp. 8741–8743, 1996.
- [22] M. Tello, C. Lopez-Martinez, and J. J. Mallorqui, "Automatic vessel monitoring with single and multidimensional SAR images in the wavelet domain," *ISPRS J. Photogrammetry. Remote Sens.*, vol. 61, no. 34, pp. 260–278, 2006.
- [23] R. Haralick, S. R. Sternberg, and X. Zhuang, "Image analysis using mathematical morphology," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 9, no. 4, pp. 532–550, Jul. 1987.
- [24] T. Xiong, S. Wang, B. Hou, Y. Wang, and H. Liu, "A resample-based SVA algorithm for sidelobe reduction of SAR/ISAR imagery with noninteger Nyquist sampling rate," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 2, pp. 1016–1028, Feb. 2015.
- [25] M. Tello Alonso, C. Lopez-Martinez, J. Mallorqui, and P. Salembier, "Edge enhancement algorithm based on the wavelet transform for automatic edge detection in SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 1, pp. 222–235, Jan. 2011.
- [26] International Maritime Organization, Automatic Identification Systems (Model course). *Int. Maritime Org.*, London, U.K., 2006.
- [27] S. Mallat, *A Wavelet Tour of Signal Processing* Burlington, MA, USA: Academic Press, 1999.
- [28] V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," *Int. J. Comput. Vis.*, vol. 22, no. 1, pp. 61–79, 1997.
- [29] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Int. J. Comput. Vis.*, vol. 1, no. 4, pp. 321–331, 1988.
- [30] K. Tanaka, *An Introduction to Fuzzy Logic for Practical Applications*. New York, NY, USA: Springer-Verlag, 1997.
- [31] P. Soille, *Morphological Image Analysis: Principles and Applications*, 2nd ed., Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2003.
- [32] A. Freeman, "On ambiguities in SAR design," presented at the *6th Eur. Conf. Synth. Aperture Radar*, Dresden, Germany, May 2006.
- [33] P. D. Mourad, "Footprints of atmospheric phenomena in synthetic aperture radar images of the ocean surface: A review," in *Air-Sea Exchange: Physics, Chemistry and Dynamics*, G. L. Geernaert, Ed. Dordrecht, The Netherlands: Springer Netherlands, 1999, pp. 269–290.
- [34] B. Smith, "An analytic nonlinear approach to sidelobe reduction," *IEEE Trans. Image Process.*, vol. 10, no. 8, pp. 1162–1168, Aug. 2001.
- [35] Y. Shiraishi and K. Fukumizu, "Statistical approaches to combining binary classifiers for multi-class classification," *Neurocomputing*, vol. 74, no. 5, pp. 680–688, 2011.
- [36] B. Quost, M.-H. Masson, and T. Dencœur, "Classifier fusion in the Dempster-Shafer framework using optimized t-norm based combination rules," *Int. J. Approx. Reason.*, vol. 52, no. 3, pp. 353–374, 2011.
- [37] X. Li, L. Wang, and E. Sung, "Adaboost with SVM-based component classifiers," *Eng. Appl. Artif. Intell.*, vol. 21, no. 5, pp. 785–795, Aug. 2008.
- [38] L. I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*. Hoboken, NJ, USA: Wiley, 2004.
- [39] D. Lee and S. Srihari, "Handprinted digit recognition: A comparison of algorithms," presented at the *3rd Int. Workshop Frontiers Handwriting Recognition*, Buffalo, NY, USA, 1993.
- [40] L. Lam and C. Y. Suen, "Application of majority voting to pattern recognition: An analysis of its behavior and performance," *IEEE Trans. Syst., Man, and Cybern.—Part A—Syst. Humans*, vol. 27, no. 5, pp. 553–568, Sep. 1997.
- [41] D. Berend and A. Kontorovitch, "Consistency of weighted majority votes," in *Proc. Adv. Neural Inf. Process. Syst.* 27, pp. 3446–3454, 2014.
- [42] M. Haklay and P. Weber, "OpenStreetMap: User-generated street maps," *IEEE Pervasive Comput.*, vol. 7, no. 4, pp. 12–18, Oct. 2008.
- [43] G. Schreier, *SAR Geocoding: Data and Systems Karlsruhe*, Germany: Wichmann, 1993.

Authors' photographs and biographies not available at the time of publication.