A near real-time automated oil spill detection and early warning system using Sentinel-1 SAR imagery for the Southeastern Mediterranean Sea

Yi-Jie Yang, Suman Singha & Ron Goldman

Published online: 07 Mar 2024.

Submit your article to this journal ⤤

Article views: 76

View related articles ⤤

View Crossmark data ⤤

Taylor & Francis
Taylor & Francis Group

A near real-time automated oil spill detection and early warning system using Sentinel-1 SAR imagery for the Southeastern Mediterranean Sea

Yi-Jie Yang ⓘ [a,b], Suman Singha ⓘ [a,c] and Ron Goldman[d]

[a]Maritime Safety and Security Lab, Remote Sensing Technology Institute, German Aerospace Center (DLR), Bremen, Germany; [b]Research and Technology Centre Westcoast, Kiel University, Büsum, Germany; [c]National Centre for Climate Research (NCKF), Danish Meteorological Institute (DMI), Copenhagen, Denmark; [d]Israel Marine Data Center (ISRAMAR), Israel Oceanographic and Limnological Research (IOLR), Haifa, Israel

ABSTRACT

The ecological and environmental impact of marine oil pollution underlines the importance and necessity of an oil spill surveillance system. This study proposes an operational automated oil spill detection and early warning system to help take quick action for oil combating operations. Oil slicks in the spaceborne Sentinel-1 synthetic aperture radar (SAR) data are detected by a trained deep learning-based oil object detector. These detected oil objects are segmented into binary masks based on the similarity and discontinuity of the backscattering coefficients, and their trajectory is simulated. The detection process was tested on one-year SAR acquisitions in 2019, covering the Southeastern Mediterranean Sea; the false discovery rate (FDR) and false negative rate (FNR) are 23.3% and 24.0%, respectively. The system takes around 1.5 h from downloading SAR images to providing slick trajectory simulation. This study highlights the capabilities of using deep learning-based techniques in an operational oil spill surveillance service.

1. Introduction

The spreading and drifting of marine oil slicks can impact aquatic wildlife in large areas and have long-term biological and ecological consequences. A spill of 1 ton of oil can cover a radius of 50 m with a thickness of 10 mm within 10 min; as it continues spreading, even such a small spill could influence an area of around 12 km^2 before it disintegrates into smaller fragments (El-Magd et al. 2020). To reduce the negative impact of a spill, countermeasures should be applied as soon as possible. A near real-time (NRT) detection system is necessary to enable such a rapid response. Due to its high maritime traffic, the Eastern Mediterranean Sea is known as an oil pollution 'hotspot' and a vital oil transit centre (Kostianoy, Kostianaia, and Soloviev 2020; Polinov, Bookman, and Levin 2021; Zodiatis, Coppini, et al. 2017). This makes an oil spill surveillance system increasingly necessary.

Remote sensing is widely used for oil spill detection since it acquires information without physical contact. These sensors can be categorized into optical and microwave sensors, and they are usually equipped on shipborne, airborne or spaceborne platforms. Optical images are likely to be influenced by weather conditions and daylight; thus, microwave sensors, such as spaceborne synthetic aperture radar (SAR) and side-looking airborne radar (SLAR), are preferable for a regular surveillance system. With a lower cost for observing a larger area, spaceborne SAR is commonly used in the field. Oil dampens gravity-capillary waves and reduces radar backscatter, resulting in dark formations in SAR scenes. However, other phenomena, such as algae, low wind areas, wind sheltering, and eddies, also manifest as dark regions and are regarded as 'look-alikes' (Hovland, Johannessen, and Digranes 1994; Topouzelis 2008). A previous study showed the possible short- and long-term contribution of oil spills on phytoplankton blooms (Tang et al. 2019), highlighting the importance of distinguishing algae and oil spills in polluted hotspots.

Conventional oil slick detection in SAR images includes dark formation segmentation, feature extraction and classification of oil slicks and look-alikes (Solberg and Solberg 1996). Dark formations are first separated from their surroundings, and their features are then extracted to identify the difference between oil slicks and look-alikes. Previous studies applied artificial neural networks (ANN) to learn features and search for the best feature combinations to classify oil slicks and look-alikes (Mera, Veronica Bolon-Canedo, and Alonso-Betanzos 2017; Stathakis, Topouzelis, and Karathanassi 2006). However, the radar complexity, source and age of oil spills, and weather conditions make it difficult to segment dark formations correctly (Mera, Veronica Bolon-Canedo, and Alonso-Betanzos 2017; Topouzelis et al. 2007). Therefore, some studies segmented dark formations with one ANN and extracted features for classification with another ANN (Singha, Bellerby, and Trieschmann 2013; Topouzelis et al. 2007). Previous studies highlighted the advantages of quad-polarimetric and dual-polarimetric HH-VV SAR observations on discriminating oil slicks from look-alikes (Migliaccio, Nunziata, and Buono 2015; Singha et al. 2016); however, the limited acquisitions of these polarimetric data make it hard to apply in an operational system.

With a large amount of accessible SAR data after the advent of the Sentinel-1 mission by the European Space Agency (ESA) in 2014 and the increasing computational capability, deep learning algorithms have been increasingly applied in oil slick detection. Instead of classifying certain dark formations as oil spills or look-alikes, a previous study applied semantic segmentation methods to categorize each pixel into one of the following classes: sea surface, oil spill, look-alike, ship or land (Krestenitis et al. 2019). There were 1112 Sentinel-1 SAR images with 1250×650 px used, and different models, such as UNet, LinkNet, PSPNet, DeepLabv2 and DeepLabv3+, were compared. However, another study pointed out that oil spills only represented around 1.2% of the total pixels, which might lead to hindrances in directly applying the semantic segmentation method to entire SAR scenes (Shaban et al. 2021). Therefore, the authors proposed a two-step approach; each image patch was classified by Convolutional Neural Networks (CNN), and patches with more than 1 pixel containing oil spills were fed to UNet for segmentation. Another study followed similar procedures but integrated oil spill classification and segmentation into a single framework by applying multi-task Generative Adversarial Networks (GANs) (Fan and Liu 2023).

Applying machine learning techniques can help reduce the work of human interpretation. However, extending the techniques from a limited dataset to an operational service is challenging. For studies using pixel-wise classification or semantic segmentation, one advantage is that the exact locations covered with oil are predicted; however, applying them to an operational service might be computationally intensive as they classify each pixel of the given images, which are the entire SAR scenes. Therefore, previous studies have utilized the oil spill detection tool provided in the Sentinel Application Platform (SNAP) toolbox, which identifies suspicious dark formations with thresholding and generates image patches containing areas of interest (Dhavalikar and Choudhari 2022; El-Magd et al. 2021). After obtaining the image patches, the authors applied dark spot detection, feature extraction and classification methods to acquire the final detections of oil slicks (Dhavalikar and Choudhari 2022).

This study proposes a different approach: The entire SAR scenes are cropped into image patches with preferable dimensions by using sliding windows, and a trained deep learning-based oil object detector examines these patches to retrieve oil slicks. Afterwards, a segmentation method separates the slicks and their surroundings to obtain the exact areas covered with oil. Previous studies have proven the capabilities of applying object detection algorithms, such as faster region-based CNN (Faster RCNN) and You Only Look Once version 4 (YOLOv4), for oil slick detection (Huang et al. 2022; Yang, Singha, and Mayerle 2022). The proposed oil slick detection system integrates into an early warning system, which helps estimate oil-contamination regions and aids in planning an oil combatting response. The European Maritime Safety Agency (EMSA) has provided an NRT 'CleanSeaNet' service since 2007; the operators assess images, identify the possible pollution and send alerts to national authorities (Carpenter 2016; European Maritime Safety Agency 2017; Singha, Vespe, and Trieschmann 2013). Such a service also contributed to a decision support system, which provides oil slick prediction models to assist the response agencies (Zodiatis et al. 2016). This study aims to provide a fully automated system which only requires the operators to confirm the results before sending alerts to its users.

The paper first introduces the structure of the surveillance system in Section 2. Detailed information about the satellite data is shown in Subsection 2.1. Subsection 2.2 introduces a custom-trained object detector from a previous study and explains the improvements done in this study. The utilization of the trained object detector and the segmentation method for oil binary mask generation are described in Subsection 2.3. The explanation of oil trajectory simulation is given in Subsection 2.4. Subsection 3.1 then evaluates the performance of the system and discusses its advantages and disadvantages in different circumstances. The model simulations for oil trajectory are compared to SAR observations as a validation study in Subsection 3.2. Afterwards, Subsection 3.3 demonstrates the operational system and provides a latency test. Finally, Section 4 summarizes the outlook and limitations of the system and outlines possible future implementations.

2. Methodology

Figure 1 shows the structure of the oil spill detection and early warning system, which contains five subsystems: satellite data processing, oil spill detection, forecasts of synoptic conditions, oil slick trajectory simulation and web interface. In the satellite data
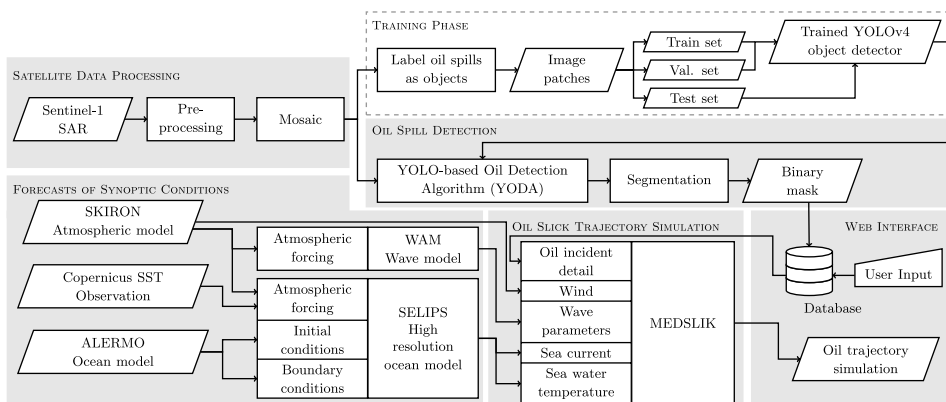
Figure 1. Structure of the oil spill detection and early warning system using the object detector from the training phase and comprising five subsystems: satellite data processing, oil spill detection, forecasts of synoptic conditions, oil slick trajectory simulation and web interface (see Subsection 2 for detailed explanation).

processing subsystem, Sentinel-1 SAR data is acquired and preprocessed with a series of corrections, and the latest coverage of preprocessed SAR over the study area is generated as a mosaic to make it easier to help visualize the whole area and understand the relationship between acquisitions better. Subsection 2.1 covers the information on Sentinel-1 data and the processing details.

Afterwards, the mosaic is delivered to the oil spill detection subsystem, including a YOLO-based oil detection algorithm (YODA) and a segmentation method. YODA automatically generates image patches, examines these image patches with an object detector from the training phase, and returns detections. The details about training the object detector are explained in Subsection 2.2, and the strategy of detecting oil spills over the whole study area is described in Subsection 2.3.1. However, as the detections are defined by bounding boxes, not the exact region covered by oil slicks, the bounding boxes of detections are turned into binary masks by the segmentation method described in Subsection 2.3.2. Subsequently, the simulations of the oil slick trajectory and fate are performed by the MEDSLIK model (Zodiatis et al. 2012), which uses daily forecasts of synoptic conditions and the oil slick binary masks; detailed information regarding oil trajectory simulation is explained in Subsection 2.4. The oil slick trajectory simulation subsystem is connected to an online interface, which enables decision-makers to perform simulations with their annotations and information and visualize the results; examples of the web interface are shown in Subsection 3.3. The results can be exported as geographic information system (GIS) compliant files, allowing users to display them with their preferred software. The system is currently running in the pre-operational phase.

The whole system focuses on the area between longitudes 31.5–36°E and latitudes 31–33.7°N in the Southeastern Mediterranean Sea defined by the dashed boundary in Figure 2, where the ocean circulation model for simulating oil trajectory is applicable. However, to increase the number of oil objects for training the detector, satellite data is collected in an enlarged area with longitude and latitude extended to the west at 30°E and north at 34.7°N, respectively (i.e. the solid boundary in the figure). The heat map in
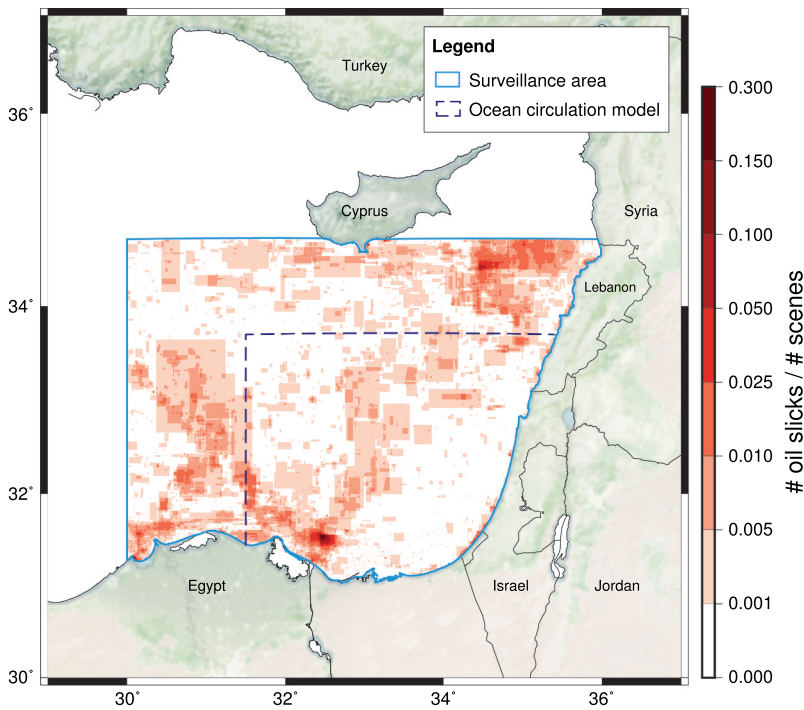
Figure 2. Study area map showing the extent of used data and model, along with a heatmap showing the chance of each pixel containing oil slicks collected from 2015 to 2018 in the previous study (Yang, Singha, and Mayerle 2022). The solid and dashed boundaries define the extent of sentinel-1 SAR acquisitions automatically downloaded by the system and the applicable area for oil trajectory simulation, respectively. Note that the oil slicks were defined by rectangular bounding boxes showing their extents but not the exact coverage. The base map and borders were obtained from Stevens (2020) and Wessel and Smith (1996), respectively.

Figure 2 illustrates the chance of oil slicks occurring in a Sentinel-1 SAR observation on each location based on a collection of manually inspected oil objects from 2015 to 2018 described in the previous study (Yang, Singha, and Mayerle 2022). As each SAR acquisition covers only part of the surveillance area, the number of scenes covering each location varies. There were at least 240, at most 755 and an average of 402 scenes covering each location from 2015 to 2018. Therefore, each pixel value on the heat map shows the ratio of the number of collected oil objects and the number of scenes covering that pixel.

2.1. Satellite data processing subsystem

Sentinel-1 SAR Level-1 Ground Range Detected (GRD) products are obtained from Copernicus Open Access Hub. Without a special subscription, the data is usually available within 24 hours after observation (European Space Agency 2013). Sentinel-1 SAR was a constellation of two satellites, Sentinel-1A and Sentinel-1B. However, after the Sentinel-1B anomaly on 23 December 2021, ESA and the European Commission (EC) announced the end of the Sentinel-1B mission on 3 August 2022 (European Space Agency 2022a). Thus, the current Sentinel-1 mission includes only Sentinel-1A data,

(a) Descending track.   (b) Ascending track
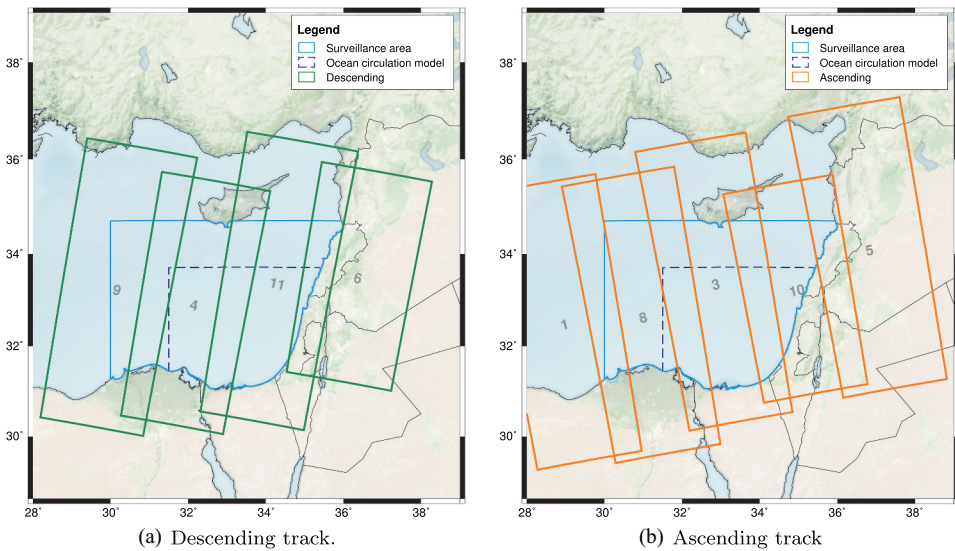
Figure 3. Coverage of sentinel-1A tracks in the study area defined in Figure 2, along with the estimated passing day index starting with 24 December 2021 in a 12-day period (i.e. 24 December 2021 has day index 1). The descending tracks pass the study area at around 03:30–04:05 UTC and the ascending ones pass at around 15:30–16:05 UTC.

and the revisiting period increased from 6 to 12 days. Figure 3 indicates the coverage of Sentinel-1A tracks over the study area along with the passing day index in a 12-day period starting with 24 December 2021 (i.e. the passing day index of 24 December 2021 is 1). The descending and ascending tracks pass the study area at around 03:30–04:05 and 15:30–16:05 UTC, respectively; they are usually available on Copernicus Data Hub at around 22:00 UTC. With the plan for the future launch of Sentinel-1C and Sentinel-1D (European Space Agency 2022b), it is worthwhile to have an operational system relying on the Sentinel-1 mission. However, adapting the system for using other SAR missions is also possible.

After download, Sentinel-1 SAR data is preprocessed with a series of corrections, including border noise removal, thermal noise removal, calibration, ellipsoid correction and conversion to decibels (dB). These are applied automatically using the SNAP Python API provided by ESA (European Space Agency 2020). Afterwards, a mosaic of different preprocessed scenes covering the study area is generated to make it easier to check the location of the new coming scenes by comparing them with the previous ones. Skipping the mosaic step and directly working on preprocessed images in SAR geometry is also possible. However, in the training stage and the pre-operational phase of the system, mosaics provide a more straightforward way for experts in different fields to compare SAR with additional information (e.g. wind speed). Therefore, the training phase and the oil spill subsystem use mosaics as input.

2.2. Training phase

In a previous study (Yang, Singha, and Mayerle 2022), oil slicks inside mosaics were examined and manually annotated as oil objects with bounding boxes with a total of 9768 oil objects

from 2015 to 2018; they were collected in *oil set*. Based on the positions of these oil objects, the original mosaics were cropped into image patches with dimensions of $N \times N$ px, where N equals the maximum of the model input size (i.e. 640×640px) and the edge lengths of the object bounding box. Afterwards, these labelled image patches were used to custom train the deep learning-based YOLOv4 object detection algorithm (Bochkovskiy, Wang, and Liao 2020) on one-class objects (i.e. oil objects). A final detector reached an average precision (AP) on the validation and test sets of 69.10% and 68.69%, respectively.

During training, the detector learns the features of objects and negative samples where no object is located. However, each image patch contained at least one oil object; hence, the detector only learned the patterns of look-alikes and the background signatures when there were oil objects nearby. In the preliminary stage of this study, applying the object detector trained with only *oil set* to YODA returned false detections focusing on dark signatures which were not oil slicks. To further improve the detector on targeting only oil slicks but not look-alikes, besides *oil set*, image patches without oil objects but with look-alikes or other remarkable signatures (e.g. radio frequency interference) were collected as *no-oil set* and used for training. Note that image patches in *no-oil set* were not given any annotation but were regarded as negative samples for object detectors to learn. Figure 4 gives some examples of these image patches. Take the left-most image patch as an example, the darker pixels show look-alikes, and the brighter pixels show clean sea surfaces; both are learned as negative samples. Therefore, though *no-oil set* did not intentionally collect clean sea surfaces, they were considered and learned in training. Table 1 shows the numbers of image patches in *oil set* and *no-oil set*.

Data augmentation is commonly applied to increase the amount of data for training. The previous study has shown the improvement of object detectors trained with additional augmented datasets in detecting larger oil objects. However, it performed worse in targeting smaller oil objects; thus, the study suggested further examinations before applying data augmentation. In this study, the trained object detector is applied to YODA; hence how data augmentation might help is also examined by
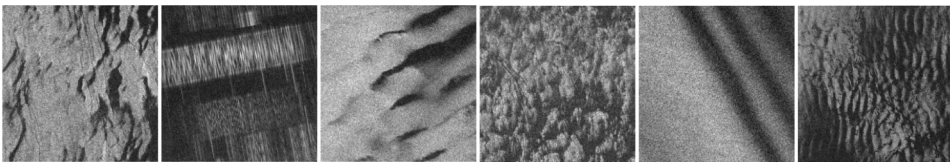
Figure 4. Examples of image patches from the *no-oil set* showing look-alikes and other remarkable features.

Table 1. Numbers of image patches collected in the *oil set* and *no-oil set*.

Dataset	# Image patches	
	oil set	*no-oil set*
training	4273	4369
validation	1218	1245
test	576	747

comparing the performance of YODA with using different object detectors. An additional augmented dataset was obtained from the rotation of image patches in *oil set* and *no-oil set* by 90º, 180º and 270º. To ensure the dataset for performance evaluation is not used in the training phase, mosaics in 2019 were inspected jointly by two experienced human interpreters, and the annotations are used as ground truth for comparison in Section 3.

According the previous study (Yang, Singha, and Mayerle 2022), object detectors have their limitation on targeting tiny objects, which were defined as:

$$
\begin{aligned}
&\text{if} \quad h_{obj} < h_{img} \cdot T_{tiny}/h_{model}[\text{px}]\cup \\
&\qquad w_{obj} < w_{img} \cdot T_{tiny}/w_{model}[\text{px}] \\
&\Rightarrow \quad \text{Obj} \in \text{Obj}_{tiny},
\end{aligned}
\tag{1}
$$

where h_{obj}, h_{img} and h_{model} respectively refer to the heights of the object, image and object detector model input; w refers to the corresponding width. T_{tiny} is a threshold for defining tiny objects, it was revealed that tiny objects with $T_{tiny} = 20\text{px}$ were difficult to be detected by the object detector. Therefore, the threshold T_{tiny} is also applied in this study.

2.3. Oil spill detection subsystem

In the oil spill detection subsystem, YODA examines mosaics with an object detector from the training phase. The strategies of applying such a detector to target oil objects inside mosaics automatically and the followed-by segmentation method are explained in the following.

2.3.1. YODA
Generally, the dataset used in the training phase should be representative of the actual tasks in operation. For instance, if oil objects mostly occupy areas of around $128 \times 128\text{px}$, however, the dimensions of images used for the training and operational phases are $640 \times 640\text{px}$ and $10240 \times 10240\text{px}$, respectively, then the detector is likely to perform poorly in operation. Thus, YODA first generates image patches from mosaic results by sliding windows with default dimensions of $640 \times 640\text{px}$ and $1280 \times 1280\text{px}$ and their corresponding sliding distances of 600px and 1200px, respectively. These image patches are rescaled to 8-bit with pixel values from 0 to 255 in order to eliminate the difference among backscatter values of different acquisitions and to optimize the processing time (Bayramov, Kada, and Buchroithner 2018); note that this procedure was executed in the training phase as well.

After the initial image patches are generated, the trained object detector goes through all these patches to find oil objects inside. The detections are examined again with regard to being too close to the border of the image patches or covering an area of more than 75% of the respective image patch. In the former case, another image patch will be generated with the detected object in the centre of the image; in the latter case, an enlarged image patch will be used. Several iterations are executed automatically until all the detections fit the criteria. The performance of YODA is evaluated by comparing the final detections with the manual inspections in 2019 and shown in Subsection 3.1.

2.3.2. Segmentation section

After obtaining detections from YODA, they are segmented into binary masks, defining the exact locations covered by oil. Figure 5 summarizes the procedures in YODA and illustrates the segmentation approach. As the detector defines slick boundaries inaccurately in some cases, the bounding boxes of the detections are extended in each direction by a certain ratio, and the extended areas are defined as the regions of interest (ROIs). Mosaics are rescaled to 8-bit images in YODA; however, the segmentation section utilizes the backscattering coefficient in order to segment the oil slick pixel by pixel as accurately as possible. The ROIs from the mosaics are loaded as image patches, and each pixel is substituted by the median of all the pixels in the kernel area for reducing speckle noise. This procedure is also known as median blur, the kernel areas is set to 5×5px to 10×10px depending on the size of ROI.

Discontinuity and similarity are the two basic properties used in segmentation algorithms (Gonzalez and Woods 2018). As oil slicks dampen the gravity-capillary waves and reduce radar backscatter, they usually appear as dark formations compared to the slick-free sea surface. To distinguish oil slicks and look-alikes, classifiers in previous studies learned different features such as statistical, geometrical, textural, contextual and SAR polarimetric characteristics (Al-Ruzouq et al. 2020). As one of the statistical features, the power-to-mean ratio is commonly applied for defining the homogeneity of either oil slicks or their surroundings (Singha, Bellerby, and Trieschmann 2013; Solberg et al. 1999). Therefore, it could also be employed for detecting the discontinuity of the oil slick and its surroundings. The power-to-mean ratio of each pixel in the ROI is calculated with its surrounding pixels and defined as σ/μ, where σ and μ refer to the standard deviation and mean of the backscatter coefficients, respectively. The edge between a slick and its surroundings is hence revealed by the discontinuity of σ/μ and extracted with a certain threshold.

However, in this study area, oil spills commonly appear near ships or at the coasts where the backscatter coefficients are higher compared to sea surface and slicks. To
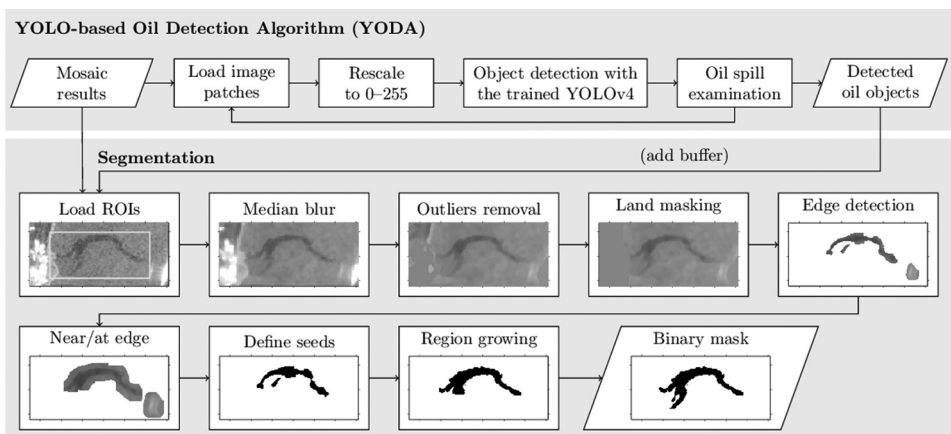
Figure 5. Detailed structure of the oil spill detection subsystem (see Figure 1 for the structure of the entire system). The explanations on YODA and the segmentation step are shown in Subsections 2.3.1 and 2.3.2, respectively.

prevent these outliers (i.e. brighter pixels) and pixels covering land areas from interfering with the edge detection, they are identified by the constant false alarm rate (CFAR) algorithm (Brusch et al. 2011) and a land mask, respectively, and replaced by the average value of the remaining pixels in the ROI prior to the edge detection. In CFAR algorithm, each pixel (target pixel) in the ROI is surrounded by a guard area and then further by a background area. The system uses 19×19px and 39×39px for the guard and background windows, respectively. The mean (μ_b) and standard deviation (σ_b) of the background pixels are used to calculate a threshold for finding bright pixels:

$$T = \mu_b + n \cdot \sigma_b, \tag{2}$$

where constant parameter n for scaling the threshold is set to 15.

Considering pixels lying on or near edge as a small subset, based on the distribution of their backscattering values, a threshold is defined for obtaining oil pixels inside the subset; these oil pixels are then regarded as seed points. Afterwards, a final oil slick binary mask can be generated by growing the seed points recursively based on similarity to their neighbouring pixels, which is defined by the mean and standard deviation of the backscatter coefficients among the seed points and the pixels passing the criteria in the previous iterations. Subsection 3.2 provides examples of oil binary masks generated by the proposed segmentation method.

2.4. Oil slick trajectory simulation subsystem

Oil binary masks from the segmentation section in the oil spill detection subsystem are delivered to the oil slick trajectory simulation subsystem, which uses the forecasts of synoptic conditions and applies a Lagrangian 3D oil numerical MEDSLIK model to emulate oil advection – diffusion and weathering processes (e.g. evaporation, dispersion and emulsification). The slick is divided into Lagrangian parcels as representations for the advection – diffusion processes and simulates their motion at the water surface (Zodiatis et al. 2012).

Advection of oil at the sea surface is driven by wind, sea currents, and waves. The MEDSLIK model uses the common 'wind factor' approach (Zodiatis et al. 2017), where the wind-induced drift is computed as a fraction of the wind speed (i.e. wind factor) and directed at a certain angle to the wind direction. In this study, the default simulations use 3.1% as the wind factor and set the wind direction as the direction of the wind-induced drift. The wind-induced drift is computed from the SKIRON atmospheric model provided by the University of Athens (Kallos et al. 1997; Papadopoulos, Katsafados, and Kallos 2001). The current-induced drift is computed from subsurface sea current velocity to avoid double counting the wind influence; the simulated modelled velocity is interpolated to 30 m depth. The forecast of subsurface ocean current, along with sea water temperature, is provided by the SELIPS high-resolution ocean model (Goldman et al. 2014, 2015), which uses the SKIRON model, the Copernicus sea surface temperature (SST) observations and the ALERMO ocean model (Korres and Lascaratos 2003). The wave-induced Stokes drift is derived from significant wave height, mean wave period and mean wave direction from the WAM wave model (Gertman et al. 2006), which uses wind forecast from SKIRON. Diffusion is emulated by a random walk model calculated with a given diffusion coefficient; the default coefficient is set to $2.0 \, \text{m}^2 \cdot \text{s}^{-1}$. Figure 1 illustrates the relations between

different models to support the explanations above. It should be noted that wind velocity and sea water temperature are used to compute weathering processes in MEDSLIK (Mackay 1980; Mackay et al. 1980).

Once the oil slick is detected, the information of time, location, slick volume and oil type with default parameters is provided to the oil slick trajectory simulation subsystem for simulating the trajectory. It checks the availability of daily forecast datasets from each model (i.e. SKIRON, SELIPS and WAM) to run a simulation using the most up-to-date forecasts at any time. For past events, the synoptic forcing is created from the concatination of the first day from each forecast datset. The SKIRON, SELIPS and WAM models provide hourly output at a horizontal resolution of 0.05°, 0.01° × 0.00833° and 0.125°, respectively.

The subsystem is connected to a web interface with a database and a processing backend. The web interface displays the oil spill simulation results, which can be downloaded by users. It also allows users to execute simulations with their own incident scenarios or by modifying the existing scenarios. The incident details and simulation parameters are stored in the database. The processing backend manages the simulation queue, converts the MEDSLIK output to georeferenced file formats, monitors the state of the forecast datasets and processes new SAR observations.

The uncertain oil properties, such as oil type, slick thickness and age, from SAR observations make simulating oil transport and fate challenging. The default slick parameters are thickness of 1 μm and medium crude oil type (API gravity is 33). The defaults are relevant to the automatic simulations, but users can use the web interface to prepare and execute another simulation with modified oil type and volume. Subsection 3.2 compares estimated oil trajectory with continuous SAR observations to evaluate the performance of the oil slick trajectory simulation subsystem.

3. Results and discussion

This section first explains the calculation of the performance assessment of YODA in Subsection 3.1. Subsections 3.1.1 and 3.1.2 are then compared to the performance of YODA applying object detectors trained with different datasets and data augmentation configurations, respectively. Afterwards, YODA is further tested on one-year data in Subsection 3.1.3; its advantages and disadvantages over different regions are also covered in the discussion. Subsection 3.2 compares SAR observations with oil slick trajectory simulation, supplied with several examples. In the end, Subsection 3.3 illustrates the whole oil slick detection and early warning system and provides a latency test.

3.1. Performance assessment of YODA

YODA goes through the input mosaics with sliding windows, targets oil objects inside with the custom-trained detector, and outputs the information of oil objects. Each detection is followed by a confidence score, which indicates the probability of an oil object appearing in the bounding box and how well the box fits the object (Redmon et al. 2016). The confidence score is defined as:

$$S_{\text{conf, class}_i} = P(\text{class}_i|\text{obj}) \cdot P(\text{obj}) \cdot IoU, \tag{3}$$

where $P(\text{obj})$ refers to the probability that the bounding box contains an object, and $P(class_i|\text{obj})$ denotes the conditional class probability showing that if an object is located at the grid cell, how likely it belongs to a specific class i. IoU represents the intersection over union, indicating how accurately the box predicts and defined as (Everingham et al. 2010):

$$IoU = \frac{\text{area}(B_p \cap B_g)}{\text{area}(B_p \cup B_g)}, \tag{4}$$

where $B_p \cap B_g$ and $B_p \cup B_g$ respectively refer to the intersection and union of the bounding boxes of the prediction (B_p) and the ground truth (B_g). In the testing, there should not be ground truth while running the object detector. Predictions from the object detector are mentioned as detections in the following paragraphs and abbreviated as *detn* in the tables. As there is no real ground truth data for oil slicks, ground truth (written as *gt* in the following tables) refers to manually inspected oil slicks described in Subsection 2.2.

Only detections with confidence scores greater than a certain threshold, $thres_{score}$, are regarded as valid detections and compared to manual inspections. Suppose a valid detection has a corresponding manual inspection such that their IoU is greater than the prescribed $thres_{IoU}$. In that case, it is considered a True Positive (TP) detection, and its corresponding manual inspection belongs to TP as well. As mentioned in Subsection 2.3, the detections are examined with several iterations in YODA, so that there might be several detections targeting the same slicks if they do not satisfy the criteria of combination. Therefore, the numbers of TP in manual inspections and detections might be different. Note that detections which have intersections with each other can be considered in one ROI in the segmentation section. False Positive (FP) refers to the detections where there is no matched manual inspection. On the other hand, the manual inspections that are not targeted by the detector belong to False Negative (FN). It is relatively important for an operational system to focus on how negative results influence the performance of the system. Thus, false discovery rate (FDR) and false negative rate (FNR) are used in the following subsections, they are defined as follows:

$$\begin{aligned} FDR &= \frac{FP}{(TP+FP)}, \\ FNR &= \frac{FN}{(TP+FN)}. \end{aligned} \tag{5}$$

Some manual inspections have intersected detections but belong to FN because their IoU do not pass $thres_{IoU}$ or the detections have confidence scores lower than $thres_{score}$. On the other hand, some detections are regarded as FP due to their poor IoU with the manual inspections. Figure 6 shows an oil slick outlined in blue and its corresponding detection in pink; the smaller slick in the figure is not annotated to avoid confusion. In this example, the extent of the slick was not defined precisely, which indicates the need for improvement in the precision of the object detector. However, this kind of FP does not target look-alikes, and the segmentation section can still define the exact coverage of the oil. On the contrary, it is especially important to reduce the number of FP originating from look-alikes, as they might lead to unnecessary cleanup operations. Therefore, the calibrated FDR and FNR are introduced and defined as:

oil: 78% (IoU: 49%)
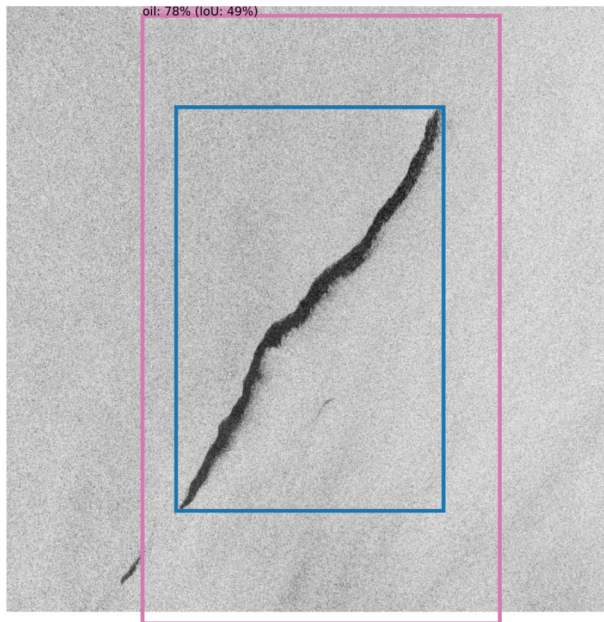
Figure 6. An example of FP, which does not lead to false alerts, is shown. The blue and pink bounding boxes mark the manual inspection and the detection, respectively. Note that this example focuses on the selected oil slick, so the smaller slick in the scene is not annotated to avoid confusion.

$$
\begin{aligned}
FDR_{cal} &= \frac{FP_{cal}}{(TP + FP)}, \\
FNR_{cal} &= \frac{FN_{cal}}{(TP + FN)},
\end{aligned}
\tag{6}
$$

where FP_{cal} shows the detections targeting at look-alikes and FN_{cal} shows the oil slicks with no detection (either valid or invalid detections) intersected. The difference between original and calibrated FDR shows the ratio of detections targeting oil slicks but are not precise enough. Similarly, the difference between original and calibrated FNR shows the ratio of manual inspections that are not detected or detected but with a lower confidence score. In the following discussion, $thres_{score} = 50\%$ and $thres_{IoU} = 50\%$ are applied.

3.1.1. Adding additional no-oil dataset

As mentioned in Subsection 2.3.1, considering *no-oil set* in training might help improve the performance of the object detector on avoiding FP. In order to verify this assumption, the performance of YODA applying detector trained with *oil set* is compared to applying an enhanced detector trained with both the *oil* and *no-oil sets*; in the following, *YODA-orig*

Table 2. Performance assessment of YODA using two custom-trained object detectors on one-month mosaics in June 2019 (see Subsection 3.1.1 for detailed comparison).

		# TP		# FP		# FN		FDR [%]		FNR [%]	
Version	# detn	gt	detn	orig	cal	orig	cal	orig	cal	orig	cal
YODA-orig	1707	191	370	1337	1182	75	9	78.3	69.2	28.2	3.4
YODA-enh	257	124	179	78	54	142	55	30.4	21.0	53.4	20.7

#gt = 266, $thres_{score} = 50\%$, $thres_{IoU} = 50\%$

and *YODA-enh* refer to YODA applying the two detectors, respectively. Table 2 shows the measures of *YODA-orig* and *YODA-enh* for testing on one-month mosaics in 2019. The calibrated FDR of *YODA-orig* was 69.2% but *YODA-enh* had its FDR decreased to 21.0%, indicating the substantial improvement of the detector on avoiding FP by applying additional *no-oil set* into training. As *YODA-orig* detected 1707 objects while there were only 266 manual inspections, the detector likely targeted most of the dark formations, which had similar patterns as oil slicks, and led to a low calibrated FNR of 3.4%. Regarding *YODA-enh*, the original and calibrated FNR were different by 32.7%, meaning that those detections were not defining the extent of the slicks well or not confident enough (i.e. confidence scores smaller than *thres*$_{score}$).

To further analyze if FNR associates with the sizes of oil objects, manual inspections were categorized into different size groups and shown in Table 3. Oil objects were sorted according to the area of the corresponding bounding boxes. If the area is smaller than 12500px (i.e. 5 km^2), the object belongs to the *small* group; on the other hand, in the *large* group the area is greater than or equal to 100000px (i.e. 40 km^2). The remaining objects belong to the *medium* group. According to the table, the detector had relatively high original FNR on *medium* and *large* groups, meaning it works worse on detecting larger oil objects. One possible reason is that *small* oil objects occupied around 72.3% of all manual inspections; thus, the detector is better at detecting smaller objects. Among all the *large* oil objects, 14.3% of them were not detected, and 64.3% of their corresponding detections should be improved to pass the thresholds. It is likely due to the lack of examples in the dataset and the complexity of the larger oil objects in shapes. Therefore, to increase the number of larger oil objects in training, data augmentation focusing on them could be beneficial.

3.1.2. Additional augmented dataset

The following compares the FDR and FNR of YODA applying detectors trained with different data augmentation configurations; the inspection time was from January to June 2019. As the previous subsection has proven that adding *no-oil* set for training is beneficial for lowering the FDR, the detectors applied to YODA in this subsection used the same dataset as detector in *YODA-enh* but included an additional augmented dataset. Image patches containing oil objects in *large* or *medium* groups were rotated by 90°, 180° and 270° and saved in the augmented dataset. Table 4 shows the numbers of image patches and data augmentation configurations of different object detectors applied to YODA. The detectors for *YODA-enh-aug1* and *YODA-enh-aug3* included the augmented dataset with the rotation of 90° and three different rotation angles, respectively. The detector for *YODA-enh-aug3b* used the same *oil set* as the one for *YODA-enh-aug3* but considered the augmentation of the *no-oil set* to balance the numbers of scenes in the *oil* and *no-oil sets*.

Table 3. Extended comparison from Table 2 focusing on the changes in FN and FNR of *YODA-enh* for different size groups.

Group	# gt	FN		FNR [%]	
		orig	cal	orig	cal
(1) Small	208	99	39	47.6	18.8
(2) Medium	44	32	14	72.7	31.8
(3) Large	14	11	2	78.6	14.3
all	266	142	55	53.4	20.7

*thres*$_{score}$ = 50%, *thres*$_{IoU}$ = 50%

Table 4. Overview of the selection of datasets and the numbers of oil objects used for different custom-trained detectors used in YODA described in Subsection 3.1.2. The application of rotation data augmentation with 90º, 180º and 270º are also noted.

	oil set			no-oil set		
Version	orig	aug (L+M)	#	orig	aug	#
YODA-enh	✓		6067	✓		6361
YODA-enh-aug1	✓	✓ (90)	8555	✓		6361
YODA-enh-aug3	✓	✓ (90, 180, 270)	13531	✓		6361
YODA-enh-aug3b	✓	✓ (90, 180, 270)	13531	✓	✓	12722

Table 5. Performance assessment of YODA using different custom-trained object detectors on six-month data from January to June 2019. Detailed comparisons are explained in Subsection 3.1.2.

		# TP		# FP		# FN		FDR [%]				FNR [%]			
								orig		cal		orig		cal	
Version	# detn	gt	detn	orig	cal	orig	cal	all	L+M	all	L+M	all	L+M	all	L+M
YODA-enh	2099	695	1400	699	453	965	431	33.3	52.6	21.6	43.2	58.1	74.3	26.0	33.9
YODA-enh-aug1	1798	638	1165	633	410	1022	538	35.2	49.3	22.8	38.6	61.6	64.4	32.4	32.9
YODA-enh-aug3	2024	586	1136	888	683	1074	422	43.9	69.3	33.7	63.8	64.7	66.6	25.4	25.5
YODA-enh-aug3b	831	339	635	196	92	1321	929	23.6	42.9	11.1	32.6	79.6	84.4	56.0	59.7

#gt = 1660, $thres_{score}$ = 50%, $thres_{IoU}$ = 50%
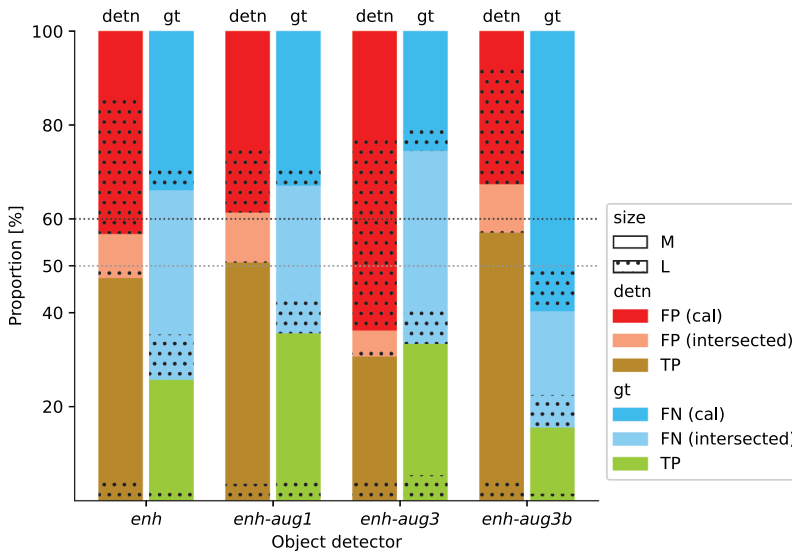
Figure 7. The proportion of positive and negative results on *large* and *medium* oil objects with different custom-trained detectors for YODA applied; this figure is used as an extended comparison from Table 5. The proportions of TP and FP among all the detections are displayed in brown and red, respectively. As for the manual inspections, TP and FN are displayed in green and blue, respectively. The lighter colours show negative results where manual inspections and detections have intersections. A detailed comparison can be found in Subsection 3.1.2.

Table 5 indicates the performance of YODA using different object detectors. Since it is better at detecting *small* oil objects according to Subsection 3.1.1, the performance on *large* and *medium* oil objects is emphasized; their FDR and FNR are listed in separate

columns, and the detection results on these objects are visualized in Figure 7. There seems to be a slight improvement of *YODA-enh-aug1* on FNR in comparison to *YODA-enh*; however, the ratio of calibrated FP (i.e. the calibrated FDR) on *large* oil objects has decreased. Regarding FNR, *YODA-enh-aug1* lowered the original FNR, but the calibrated one did not differ much. It seems that different detectors have similar abilities to detect *large* and *medium* oil objects, but there were more detections from *YODA-enh-aug1* passing the thresholds than the ones from *YODA-enh*.

To further investigate how different rotation augmentations might help with the performance, *YODA-enh-aug1* and *YODA-enh-aug3* are compared. As mentioned in Subsection 3.1.1, adding image patches with look-alikes or other remarkable features could help; however, with the imbalanced numbers of images in the *oil* and *no-oil sets* for training the detector for *YODA-enh-aug3*, FDR has increased. In other words, it shall be essential to have similar numbers of images with and without oil objects inside, and *YODA-enh-aug3b* supports this assumption as it had the lowest FDR among all. However, there were only 831 detections but around 1660 manual inspections; the detector seems conservative, which led to a relatively high FNR.

In summary, *YODA-enh* had the best overall performance; however, *YODA-enh-aug1* worked better on targeting larger oil objects. Considering the impact of large oil slicks on the environment, *YODA-enh-aug1* is the most suitable for the system. However, both of them have pros and cons; thus, they are compared further in the following subsection, which focuses on the performances on one-year mosaics and discusses their competence in different regions.

3.1.3. Performance of the YODA on one-year data

Since algae distribution, maritime traffic and wind condition might vary monthly or seasonally, testing YODA for an entire year is necessary. Figure 8 shows the spatial distribution of the manually inspected oil slicks collected in 2019 and marks the different zones for performance assessment in the following paragraphs. Oil could cause long-term biological and ecological consequences; thus, to raise the environmental awareness of the authorities, it is vital to highlight the pollution within the territorial waters, where it might damage the vulnerable coastal and marine ecosystems of its corresponding countries. The territorial waters are defined as 12 nautical miles (around 22 km) off the coast; the green line illustrates the region by extending the coastline for 22 km. Note that it is not definitive for the actual territorial waters boundaries.

Concerning sources of look-alikes, Figure 9 gives an example on 22 January 2023, illustrating possible land-sourced chemical or oil spills, ship-based discharges, wind-induced look-alikes and algae. Wind-induced look-alikes commonly appeared off the east coast and interfered with detecting oil slicks. As they were not commonly seen off the south coast, the red line in Figure 8 separates the coastal region into the *Eastcoast* and *Southcoast* zones. Usually, wind-induced look-alikes covered a larger area compared to the coastal spills; thus, the orange line, extended from the coastline by 5km, splits the *Eastcoast* zone into *Eastcoast-5* and *Eastcoast-22* zones.

This study focuses on the area where the ocean circulation model is applicable; the model area excluding the coastal zones is defined as *Sea-in* zone, 47.1% of the oil objects in 2019 located in this zone. Unlike the coastal area, which contained different sources of spills and look-alikes, oil spills away from the coast in the study area were mainly from
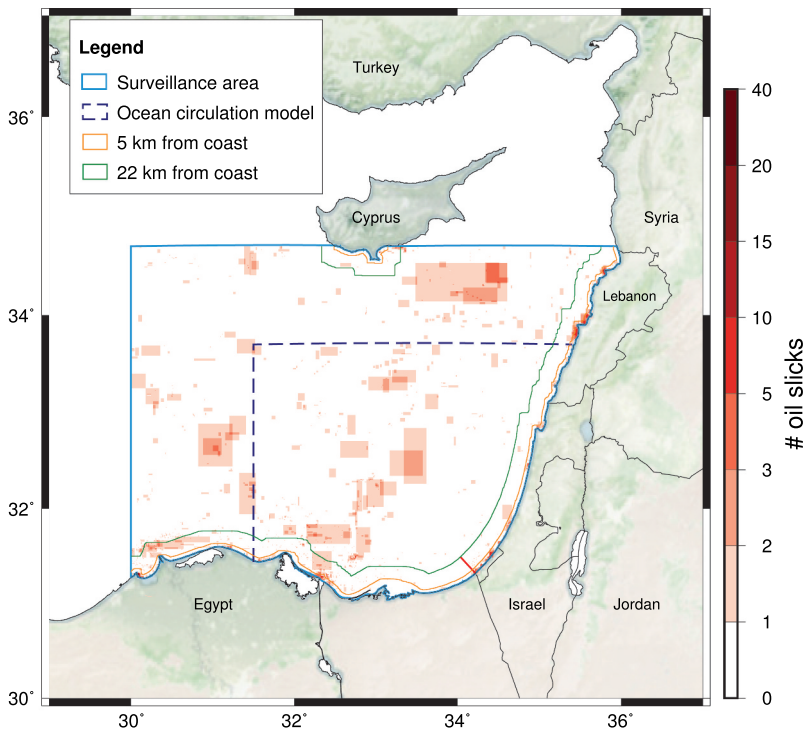
Figure 8. Based on the definition of the study area shown in Figure 2, different zones are marked for further discussion on the performance of YODA in different locations described in Subsection 3.1.3. The heatmap shows the number of oil slicks collected from 2019 for testing YODA. The red line separates the east and south coasts.

ships. For this reason, the non-coastal area outside the model area is included in the comparison and considered as *Sea-out* zone. Oil objects and detections are categorized into different zones regarding the location of their centre coordinates. Table 6 summarizes the definition of different zones and shows the number of manual inspections in each zone.

Table 7 shows the performance of *YODA-enh* and *YODA-enh-aug1* in different zones. Overall, they both had similar FDR, but *YODA-enh* had a lower FNR than *YODA-enh-aug1*, meaning that *YODA-enh* detected more oil objects; this aligns with the findings in Subsection 3.1.2. Similar results are also shown in the *Sea-in* and *Sea-out* zones but with lower calibrated FNR, possibly due to fewer wind shadows as lookalikes in these zones.

Previous studies suggested the ideal wind speed range should be between 3 and 7–10 m/s for visualizing oil slicks in SAR observations since there was a high possibility of lookalikes due to low wind and thinner slicks might be invisible with the strong wind due to a combination of oil dispersion (Brekke and Solberg 2005). To better understand how wind speed might influence performance, (Figures 10 and Figure 11) illustrate the monthly performance and the monthly mean wind speed from the SKIRON forecast, respectively, in 2019. Among all the monthly performances, both *YODA-enh* and *YODA-enh-aug1* seem to have the worst performance in December, where FDR and FNR were high. In December and
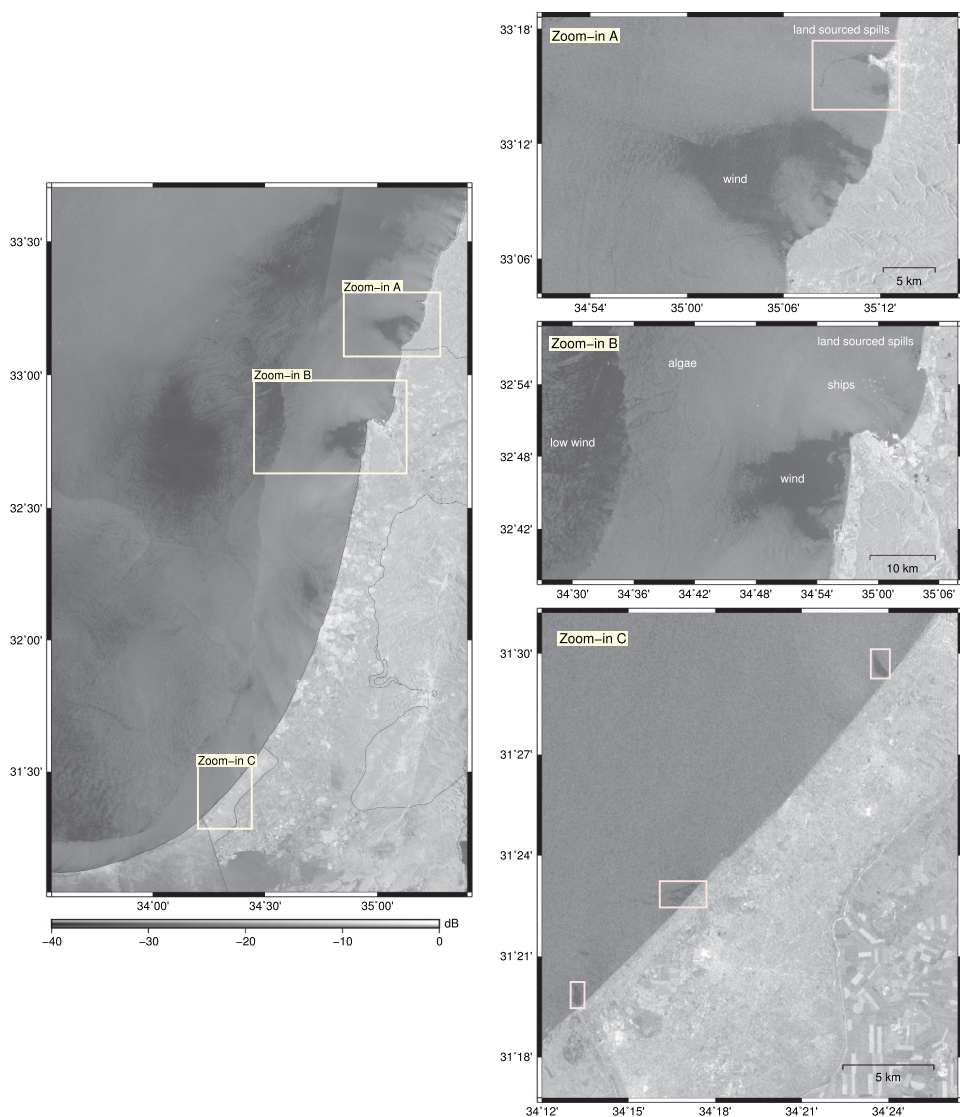
Figure 9. A mosaic on 22 January 2023 at 03:44 showing different sources of dark formations. Zoom-in areas a and B present some common wind-induced look-alikes, algae, land-sourced spills, and discharges from ships. Zoom-in area C shows a hotspot with regular land-sourced spills.

Table 6. Definition of different zones in the study area and numbers of manually inspected oil slicks are listed. The geographic locations are illustrated in Figure 8.

Zone	# gt	Descriptions
Eastcoast-5	287	Within 5km off the east coast in the model area
Eastcoast-22	40	Between 5km and 22km off the east coast in the model area
Southcoast	493	Within 22km off the south coast in the model area
Sea-in	731	Model area excluding the zones described above
Sea-out	747	Surveillance area excluding the coastal areas and the zones described above

Table 7. Performance assessment of YODA-enh and YODA-enh-aug*1* in different zones on one-year mosaics from January to December 2019.

			# TP		# FP		# FN		FDR [%]		FNR [%]	
Version	Zone	# detn	gt	detn	orig	cal	orig	cal	orig	cal	orig	cal
YODA-enh	Eastcoast-5	342	131	231	111	77	236	116	32.5	22.5	64.3	31.6
	Eastcoast-22	98	14	45	53	47	33	11	54.1	**48.0**	70.2	23.4
	Southcoast	559	224	382	177	86	324	151	31.7	**15.4**	59.1	**27.6**
	Sea-in	902	323	496	406	244	406	147	45.0	27.1	55.7	20.2
	Sea-out	1021	314	619	402	227	466	168	39.4	22.2	59.7	21.5
	overall	2922	1006	1773	1149	681	1465	593	39.3	23.3	59.3	24.0
YODA-enh-aug1	Eastcoast-5	281	127	194	87	65	240	152	31.0	23.1	65.4	41.4
	Eastcoast-22	55	15	27	28	23	32	11	50.9	**41.8**	68.1	23.4
	Southcoast	407	188	314	93	38	360	225	22.9	**9.3**	65.7	**41.1**
	Sea-in	849	305	455	394	225	424	188	46.4	26.5	58.2	25.8
	Sea-out	1052	303	555	497	280	477	214	47.2	26.6	61.2	27.4
	overall	2644	938	1545	1099	631	1533	790	41.6	23.9	62.0	32.0

$thres_{score} = 50\%$, $thres_{IoU} = 50\%$
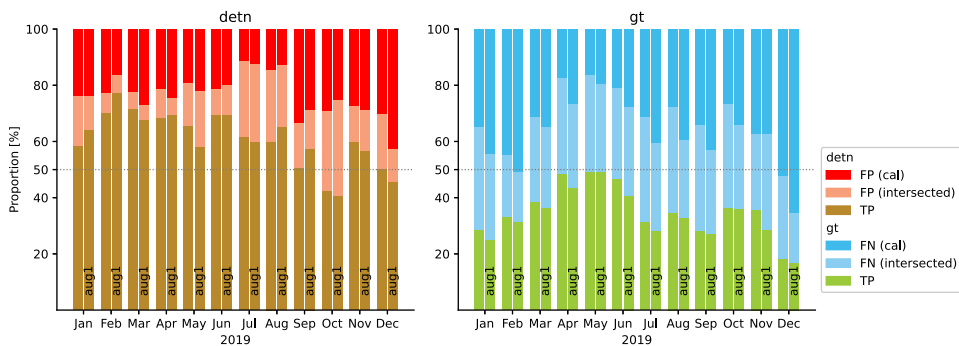
Figure 10. The monthly proportion of positive and negative results of *YODA-enh* and *YODA-enh-aug1* illustrating the possible influence of strong wind and algae seasons explained in Subsection 3.1.3. The bars without annotation and annotated with *aug1* refer to *YODA-enh* and *YODA-enh-aug1*, respectively.

January, the strong wind might lead to poor visibility on slimmer oil slicks and result in high calibrated FNR in these months. On the contrary, October was the month with the overall lowest wind speeds ranging from 2 to 5 m/s, especially in the regions closer to the coast where the mean wind speed was lower than 3 m/s. The FDR seems to be high as there might be more SAR observations with wind speeds lower than the recommended speed. Therefore, wind speed should be considered in the system to provide more reliable results.

A previous study highlighted the operational oil spills with sizes of about 1–10 tonnes released by ships that occurred almost daily in the Mediterranean Sea (Kostianoy, Kostianaia, and Soloviev 2020). Heavy shipping activities, mainly due to the Suez Canal, have caused the *Southcoast* zone to become an oil pollution hotspot, especially near the Port Said according to Figures 2 and Figure 8. Wind-induced look-alikes were less common in this zone than the *EastCoast*; however, the intersection of ships and the water surface might cause ship wakes and result in lower radar backscatter compared to their surroundings (Tings, Pleskachevsky, and Wiehle 2023). In other words, ship wake was one possible source of look-alikes in the heavy ship traffic *SourthCoast*. Figure 12 shows an
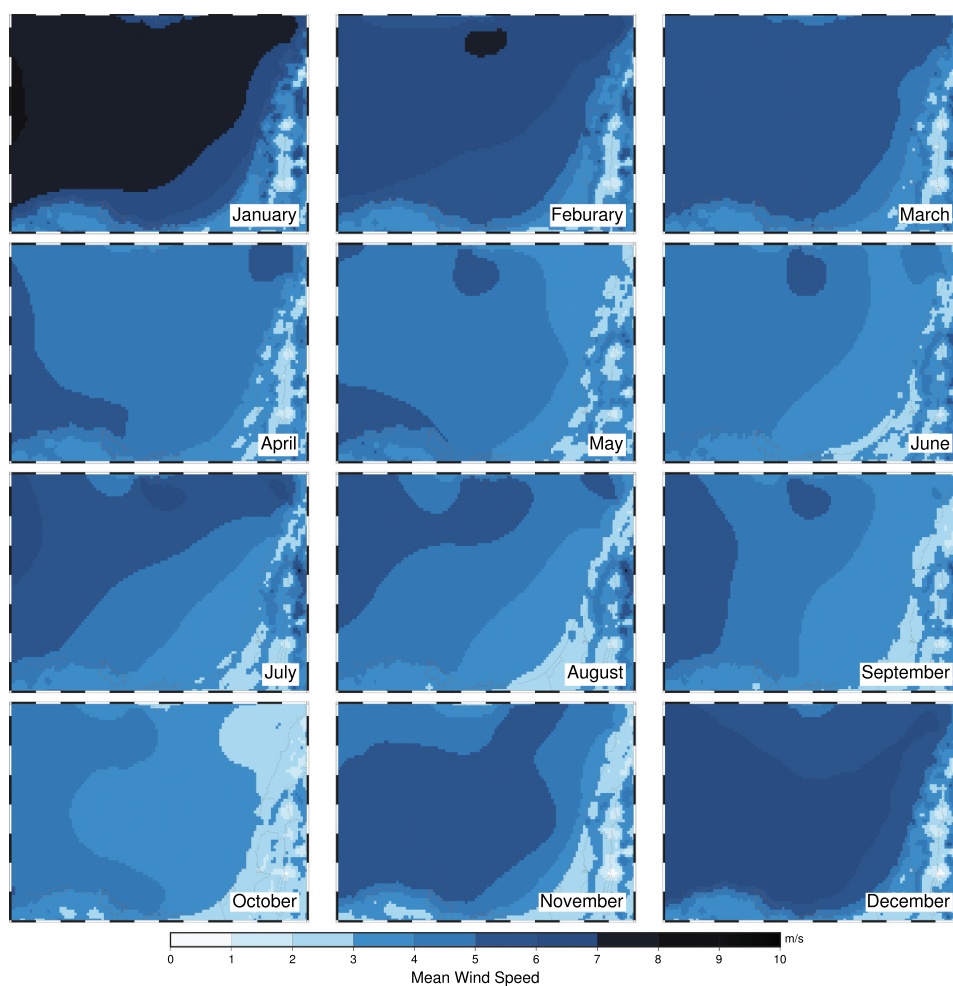
Figure 11. Monthly average of the wind speed in 2019 from SKIRON.

example of oil slicks and ship wakes in the same SAR observation where the backscatter values of slicks were lower than those of ship wakes.

The calibrated FDR of *YODA-enh* and *YODA-enh-aug1* were 15.4% and 9.3%, respectively, which were relatively low compared in other zones. It indicated that most of the detections in this zone were correct; however, the calibrated FNR of *YODA-enh* and *YODA-enh-aug1* were 27.6% and 41.1%, respectively, which were higher than overall. As thin slicks or slicks propagated and evaporated after a few days on the water might lead to similar low back-scatter values as ship wakes do, the distinction between them is challenging. Therefore, the detectors were more conservative with sending the detections of oil, which led to low FNR. Considering FDR, *YODA-enh-aug1* seems suitable in this zone for reducing the possibility of sending false alarms. However, including ship and wake information might help avoid possible FP; in this case, *YODA-enh* with lower FNR could also be applied.

Similar results are also shown in the *Eastcoast-5* zone, where the calibrated FNR were relatively high compared to other zones, especially for *YODA-enh-aug1*. There were ship discharges near the ports in both zones but more land-sourced spills in this zone, according
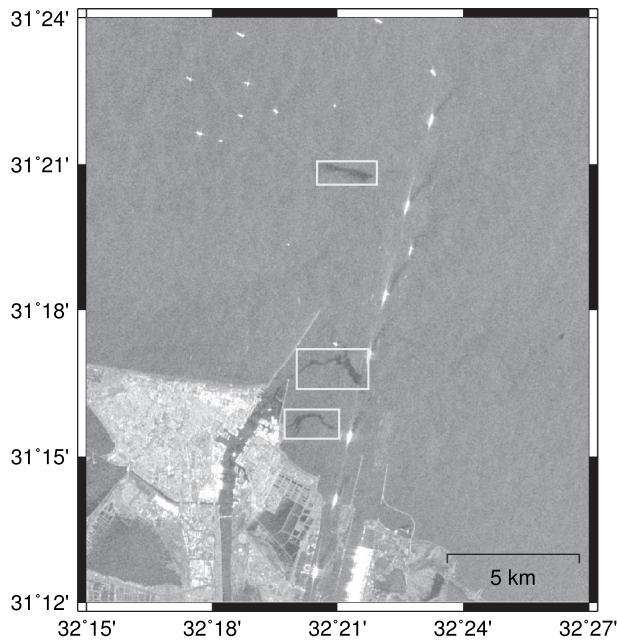
Figure 12. An example on 7 July 2022 at 03:51 showing oil slicks along with ship wakes near Port Said off the Egyptian coast. The bounding boxes define slicks, and dark formations near ships (bright pixels) are likely to be ship wakes.

to the manual inspections in this study. On top of that, these land-sourced spills were commonly interfered with by wind-induced dark formations or active transponders; the latter are commonly used for security reasons in the Eastern Mediterranean Sea. Figure 13 shows an example SAR scene with radio frequency interference.

Concerning the *Eastcoast-22* zone, the FDR seems to be relatively high for both *YODA-enh* and *YODA-enh-aug1*. As the common wind-induced look-alikes in this zone might confuse the interpretation of spills, dark formations near such look-alikes were not annotated as oil objects in the training stage unless there was clear evidence. In addition, image patches with these unsure dark formations were not included in either the *oil* or the *no-oil sets* mentioned in Table 1 for training the detectors. However, those unsure dark formations were not avoidable for testing the system performance on one-year data. In
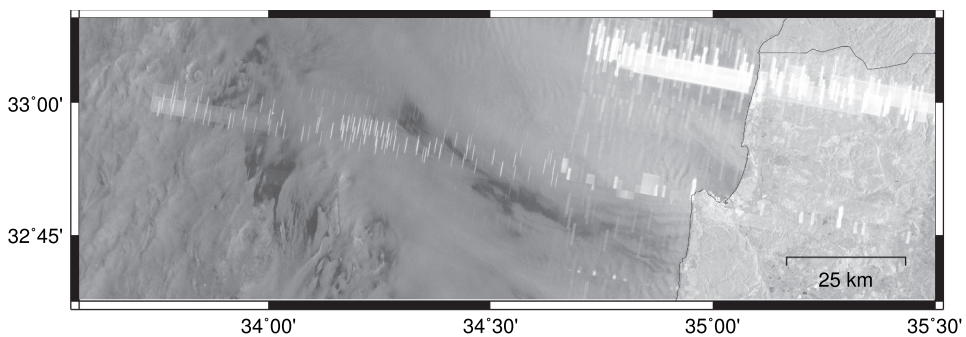

Figure 13. An example on 28 February 2017 at 03:42 shows radio frequency interference on SAR, which constantly appears in certain regions.

other words, some detections might target actual spills but be defined as FP because those spills were not annotated as oil objects. Therefore, considering synoptic conditions at the corresponding time of SAR observations might help increase the quality of manual inspections and the confidence of detections.
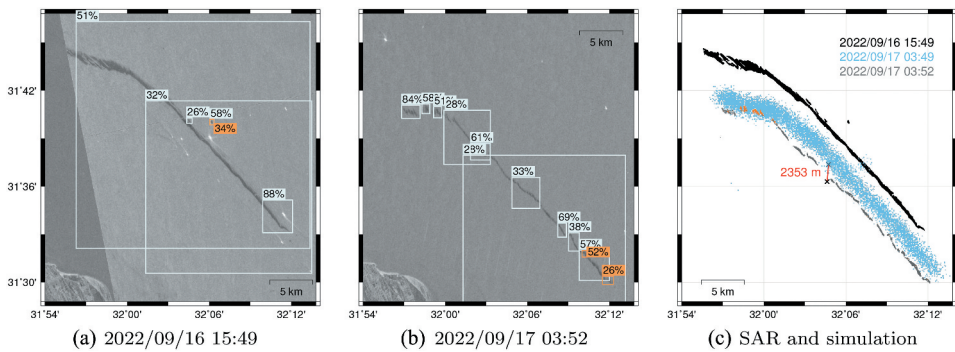
3.2. Comparison between SAR observations and slick trajectory simulation

In Subsection 3.1, the performance of YODA was evaluated. After detecting oil slicks, they are segmented into binary masks and delivered to the MEDSLIK model for hourly estimation of the slick trajectory. In some cases, the same oil slicks appeared in SAR observations from different Sentinel-1 orbits at different time, which could help understand how SAR observations and simulations are similar or different from each other. Several oil slicks in continuous SAR observations are presented in the following, the continuous observations are all around 12 hours apart. The distance between centroids of SAR observation and trajectory simulation are computed and used for comparison. Both *YODA-enh* and *YODA-enh-aug1* were applied to the selected cases as illustrations on how different detectors target oil slicks.

The annotations of figures in this subsection follow the same format. Detections annotated in mosaics from *YODA-enh* and *YODA-enh-aug1* are displayed as blue (light) and orange (dark) bounding boxes, respectively; their confidence scores are attached to the bounding boxes (see Equation 3 for the definition of confidence scores). In Subsection 3.1, only valid detections (i.e. detections passing the confidence score threshold, $thres_{score}$) are discussed; however, all detections in the selected cases are displayed in this subsection. As for comparing SAR observations and simulations, the black binary masks display SAR observations sent to the MEDSLIK model for trajectory simulation. The estimated positions are marked in light blue and compared to the corresponding binary masks (in grey) acquired from SAR, and the overlapped areas are shown in orange; their centroids are shown as black and blue crosses for SAR observation and simulation,

(a) 2022/09/16 15:49　　　　　　　　　(b) 2022/09/17 03:52
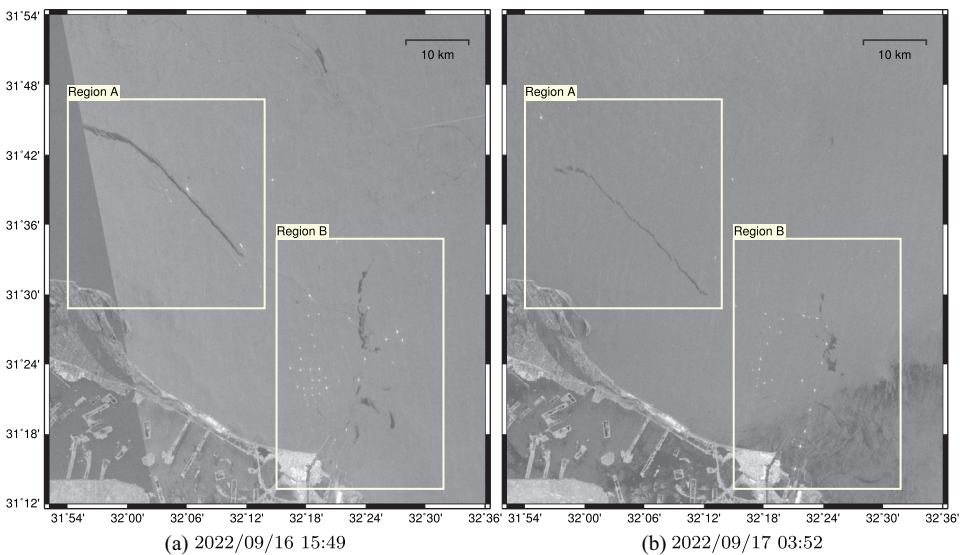
Figure 14. An example of several oil slicks near Port Said off the Egyptian coast continuously observed by SAR on 16 and 17 September 2022. The detections and simulations are compared in different zoom-in maps Figures 15 and Figure 16 for regions a and B, respectively.

Figure 15. Zoom-in maps of region a in Figure 14 showing detections on two continuous SAR acquisitions and a comparison between SAR observations and trajectory simulation.
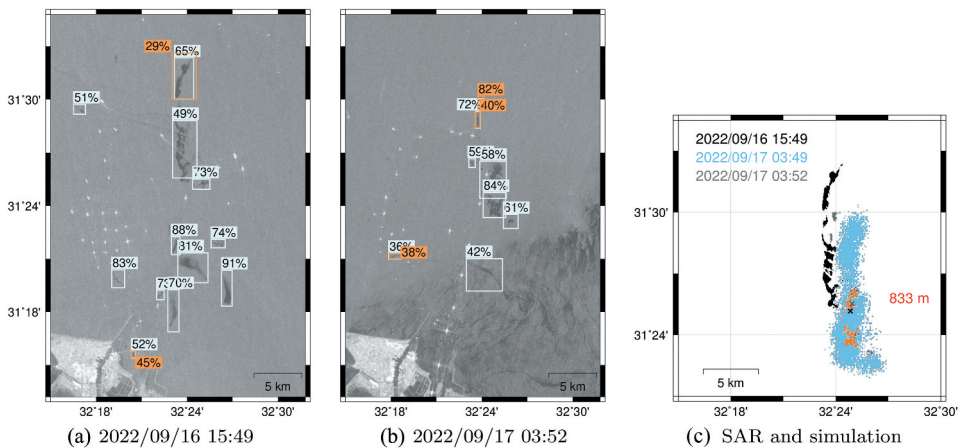

Figure 16. Zoom-in maps of region B in Figure 14 showing detections on two continuous SAR acquisitions and a comparison between SAR observations and trajectory simulation focusing on the northern slicks.

respectively. The red line marks the distance between the two centroids. Note that simulations are calculated not only for forward but also for backward estimations.

Figure 14 shows oil slicks on 16 and 17 September 2022 at an oil spill hotspot off the Egyptian coast according to the heatmap in Figure 2. Most of the slicks were located in the zoom-in regions A and B and presented in zoom-in maps in Figures 15 and Figure 16, respectively. As there were look-alikes covering areas closer to the coast on 17 September, the comparison between simulation and SAR observation only focuses on the northern slicks. Overall, most of the observed slicks covered smaller areas after 12 hours; however, the simulations show that they might influence larger regions. The oil is assumed to be composed of volatile component and non-volatile residue in the model. The evaporation of each parcel is an exponential decay process. At the same time, there is eddy diffusivity and spreading processes which redistribute the concentration of oil and of the oil parcels.
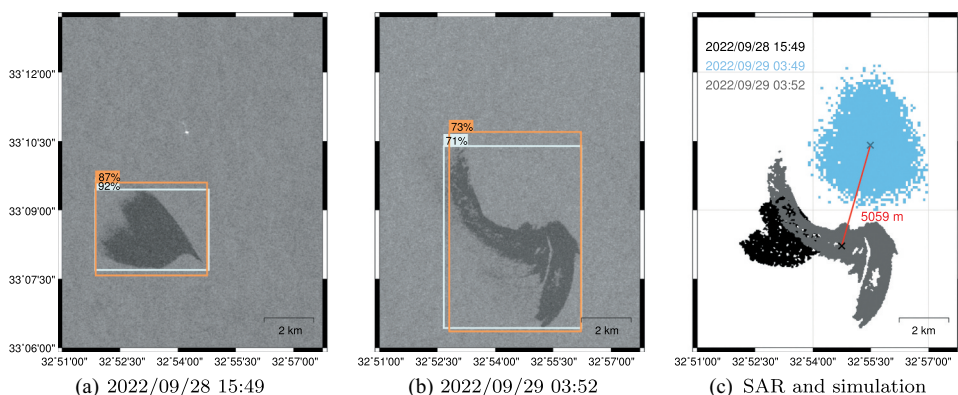
Figure 17. An example of an oil slick with its shape changed significantly observed by SAR on 28 and 29 September 2022, showing the detections on two SAR scenes and a comparison between SAR observations and trajectory simulation.

It is possible that the diffusivity is overestimated; the actual slick evaporated, dissolved into the water, or had low concentrations so they were not able to be detected.

However, in some cases, SAR observations and trajectory simulations did not align well, Figure 17 shows such an example on 28 and 29 September 2022. The shape of the slick has changed by the wind and currents according to SAR observations. The most likely reason for the misfit between forecast and detection is the uncertainty in the numerical forecasts of current and wind. By assimilation of velocity observations, the forecast may be improved. In that respect, the misfit between observation and forecast could potentially be used to improve the forecast quality at oil spill hot spots. Moreover, Li et al. (2022) have presented a machine learning approach using reanalysis and forecast to train a system that corrects the wind field and improves the error of oil forecast. This case is notable because the slick occurred near the centre of an eddy, so changes in circulation direction and the resulting deformations of the slick are more pronounced.
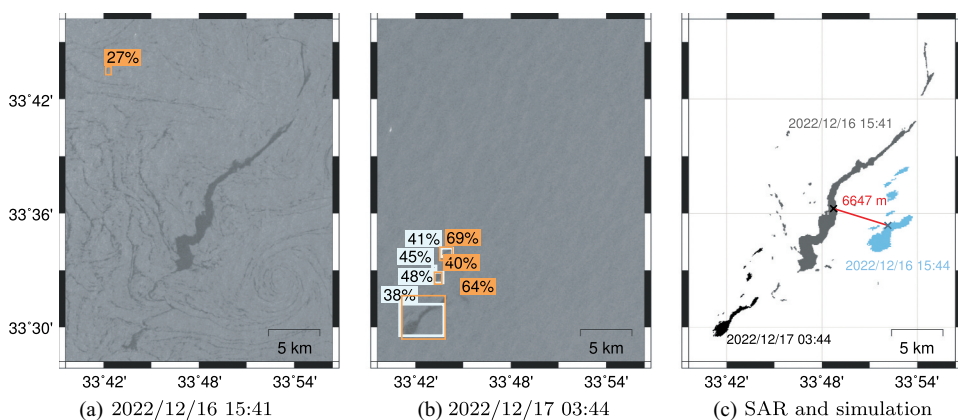
Figure 18. An example of a possible oil slick appeared near look-alikes on 16 December 2022; the observation on the next day and its backward simulation from the following day helped confirm the slick.

Another example demonstrates the advantage of oil trajectory simulation, which helps indicate oil slicks surrounded by look-alikes. Figure 18 shows oil slicks on 16–17 December 2022. There were some look-alikes on 16 December, which might be algal blooms driven by currents and made it hard to distinguish whether there was a slick. SAR observation on 17 December shows a slick having similar shape to the possible slick on 16 December and its location fits with the travelling direction from trajectory simulation. Though the size of the possible slick on 16 December covered much larger area than the one on 17 December, which is likely due to the evaporation of the oil, it is reasonable to consider those two slicks as one slick at different time. Currently, both *YODA-enh* and *YODA-enh-aug1* could not find the slick on 16 December as it was surrounded by look-alikes. However, for continuous surveillance on a specific study area, simulation of slicks is beneficial for helping the system to focus on certain regions likely to contain possible oil slicks which are derived from the trajectory estimation of oil from previous scenes.

To compare *YODA-enh* and *YODA-enh-aug1*, the following discussion focuses on their detections in the above examples. In the first example (see Figures 15 and Figure 16) , *YODA-enh* performed better than *YODA-enh-aug1*. Most of the slicks were detected by *YODA-enh* though the continuous slicks were considered as many detections. As explained in Subsection 2.3, sliding windows are applied to generate image patches for feeding the detector. If the detections do not pass the defined criteria, they are examined again in the new image patches. In some situations, the size of image patches would be larger than the default sizes of the sliding windows. There should be an upper limit for the size of image patches; otherwise, the slim and long oil slicks could be invisible to the object detector. Therefore, it is reasonable to return several detections for one single long and slim slick. In addition, the segmentation method combines the overlapping detections and generates one binary mask. As one long oil slick could be regarded as several detections, it might lead to the poor system performance shown in Subsection 3.1.

Oil slicks in the second example (see Figure 17) appeared at the wind condition that was suitable for SAR observation, both oil slicks were wide and distinct from their surrounding. Thus, both detectors targeted the slicks correctly with high confidence scores. However, in the last example (see Figure 18), both detectors did not find oil slicks on 16 December; *YODA-enh-aug1* even had a false detection on look-alikes with a low confidence score of 27%. It is likely that the patterns of such look-alikes are similar to some oil slicks the detector learned during training. Such results align well with the comparison in Subsection 3.1.2 (see Table 5) that *YODA-enh-aug1* has better performance on larger oil slicks, but *YODA-enh* performs better overall. Considering 17 December (see Figure 18(b)), both detectors targeted the oil slicks, but none of the detections from *YODA-enh* passed $thres_{score}$. *YODA-enh-aug1* showed higher confidence scores on detecting those oil slicks.

3.3. Oil spill detection and early warning system

Figure 19 provides the approximate time stamps for the operational system and illustrates the system on 28 September 2022. The Sentinel-1 SAR descending and ascending tracks pass the study area at around 03:30–04:05 and 15:30–16:05 UTC, respectively. Usually, they are available on Copernicus Open Access Hub at around 22:00 UTC. Therefore, the forecasts of the
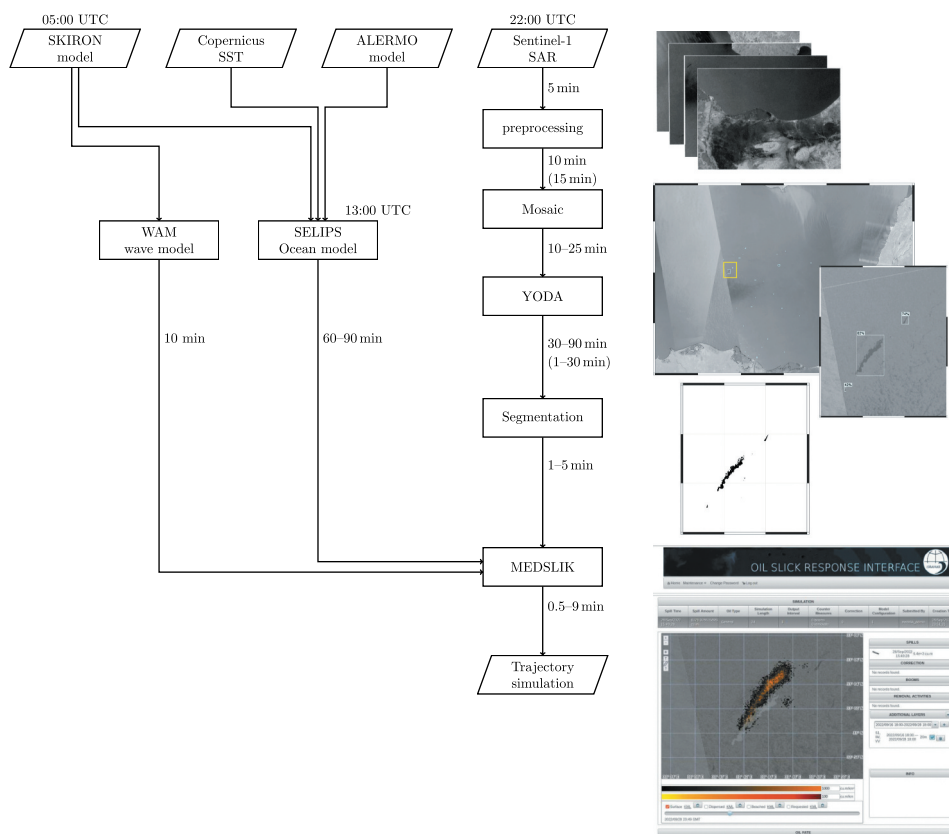
Figure 19. The operational chain of the proposed oil spill detection and early warning system, along with the time stamps for each process. The CPU-only and GPU machines are compared from downloading sentinel-1 data to delivering oil binary masks to the MEDSLIK model. The time stamps in brackets show the different latency from the GPU machine. See Subsection 3.3 for further explanation.

synoptic conditions should be ready for the trajectory simulation using the MEDSLIK model. In any case, it should be possible to obtain the forecasts from the day before. From downloading Sentinel-1 data to delivering oil binary masks to the MEDSLIK model, two different machines are compared in the latency test. One is a CPU-only machine which uses Intel Xeon E7-8857v2 CPU and 256 GB RAM; another has Nvidia GeForce RTX 3080 GPU with 10 GB VRAM, AMD Ryzen 7 3700X CPU and 32 GB RAM. The latency of YODA highly depends on the complexity of SAR scenes, as the complex scenes would increase the number of iterations. The major differences are the time intervals for the preprocessing step and YODA; time stamps for the GPU machine are noted in brackets. Because of the limited RAM, the GPU machine is slower in the preprocessing section as fewer scenes can be processed in parallel; however, using GPU increases the efficiency of detecting oil objects. Therefore, deploying on a machine with multi-core CPU, sufficient RAM and a GPU for deep learning inference can achieve optimal latency. The trajectory simulation is set to 24 durations in the automatic system; increasing the duration will result in more time costs. The system also offers tools for users to define their polygons (see Figure 20).
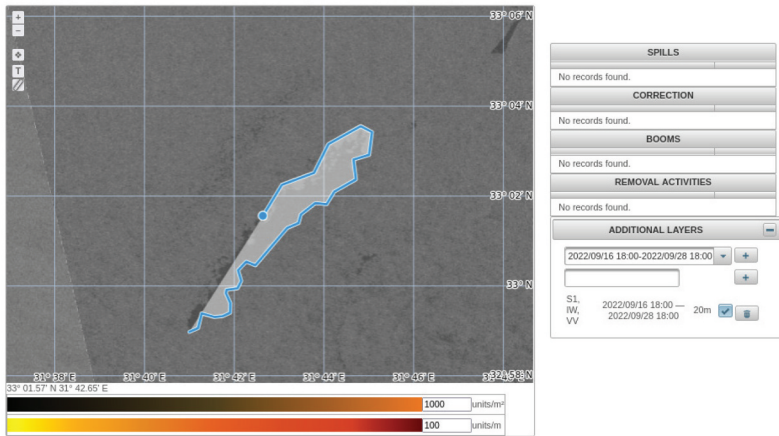
Figure 20. A screenshot of the web interface showing how users can define their own oil binary masks.

However, skipping the mosaic step and working on preprocessed scenes in SAR geometry is possible and should accelerate the system. From preprocessing SAR scenes to delivery of trajectory simulation, it takes around 1.5 h for the current system. As shown in Figure 19, it takes 10–25 min to generate a mosaic covering the study area, which could be a helpful step in the pre-operational stage, as mentioned in Subsection 2.1. However, skipping the mosaic step and working on preprocessed scenes in SAR geometry is possible and should accelerate the system.

4. Conclusion

This study developed an automated system which provides regular surveillance over the Southeastern Mediterranean Sea and sends detections to human operators for confirmation. The performance of applying different custom-trained object detectors for YODA was evaluated in Subsection 3.1; *YODA-enh* and *YODA-enh-aug1* stood out as performing better than the others and had similar overall calibrated FDR of 23.3% and 23.9%, respectively. However, *YODA-enh* had a lower calibrated FNR of 24.0% than 32.0% from *YODA-enh-aug*, meaning that *YODA-enh-aug1* missed more slicks. To further compare them and evaluate the slick trajectory simulation from the MEDSLIK model, several oil slicks in consecutive SAR observations were carried out in Subsection 3.2. Both *YODA-enh* and *YODA-enh-aug1* worked well in some cases but did not find oil slicks in other cases; thus, it might be important to analyse their superiority over different criteria.

Various sources of look-alikes appeared regularly in the study area and influenced the performance, especially for algae, wind-induced dark formations and ship wakes. Algae occur seasonally, and their density could be provided by chlorophyll-like pigment concentration (*chl*) data, which might help with filtering out the possible FP. In the *EastCoast* zone where wind-induced look-alikes appeared regularly, both detectors had high calibrated FDR of 48% and 41.8% for *YODA-enh* and *YODA-enh-aug1*, respectively. On the other hand, ship wakes commonly appeared in the *SouthCoast* zone with heavy ship traffic also led to high calibrated FNR of 41.1% for *YODA-enh-aug1*. Therefore, including additional information, such

as *chl*, synoptic conditions and ship and wake information, is suggested for providing more reliable results and determination of detection confidence and alert levels.

The proposed automated early warning oil spill surveillance system provides execution results and reports within around 1.5 h after downloading SAR observations with a GPU machine. To accelerate the system, future work could include wind speed and *chl* density information in YODA to focus on subareas of the given surveillance area, which have ideal wind speed range and low *chl* density. Alternatively, changing the system structure to work on preprocessed data in SAR geometry can also speed up the system. Oil slicks appear differently under different circumstances, such as wave height, oil volume, wind speed, oil type, sea surface temperature and currents (Bayramov, Kada, and Buchroithner 2018). Therefore, the proposed system might be limited to the types and sources of oil slicks commonly occurring in the Southeastern Mediterranean Sea. However, regarding the system as a prototype combining deep learning-based object detection algorithm and segmentation method to apply in other oil pollution hotspots is feasible.

Acknowledgements

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

ORCID

Yi-Jie Yang http://orcid.org/0000-0002-4098-8119
Suman Singha http://orcid.org/0000-0002-1880-6868

Data Availability statement

The data that support the findings of this study are available from the corresponding author, Y.-J. Yang, upon reasonable request.

References

Al-Ruzouq, R., M. Barakat A Gibril, A. Shanableh, A. Kais, O. Hamed, S. Al-Mansoori, and M. Ali Khalil. 2020. "Sensors, Features, and Machine Learning for Oil Spill Detection and Monitoring: A Review." *Remote Sensing* 12 (20): 3338. https://doi.org/10.3390/rs12203338.

Bayramov, E., M. Kada, and M. Buchroithner. 2018. "Monitoring Oil Spill Hotspots, Contamination Probability Modelling and Assessment of Coastal Impacts in the Caspian Sea Using SENTINEL-1, LANDSAT-8, RADARSAT, ENVISAT and ERS Satellite Sensors." *Journal of Operational Oceanography* 11 (1): 27–43. https://doi.org/10.1080/1755876X.2018.1438343.

Bochkovskiy, A., C.-Y. Wang, and H.-Y. M. Liao. 2020. "YOLOv4: Optimal Speed and Accuracy of Object Detection." *ArXiv Preprint*. https://doi.org/10.48550/arXiv.2004.10934.

Brekke, C., and A. H. S. Solberg. 2005. "Oil spill detection by satellite remote sensing." *Remote Sensing of Environment* 95 (1): 1–13. https://doi.org/10.1016/j.rse.2004.11.015.

Brusch, S., S. Lehner, T. Fritz, M. Soccorsi, A. Soloviev, and B. van Schie. 2011. "Ship Surveillance with TerraSAR-X." *IEEE Transactions on Geoscience and Remote Sensing* 49 (3): 1092–1103. https://doi.org/10.1109/TGRS.2010.2071879.

Carpenter, A. 2016. "European Maritime Safety Agency Activities in the Mediterranean Sea." In *The Handbook of Environmental Chemistry*, edited A. Carpenter and A. G. Kostianoy, 191–213, Vol. 83. Cham: Springer. https://doi.org/10.1007/698_2016_18.

Dhavalikar, A. S., and P. C. Choudhari. 2022. "Detection and Quantification of Daily Marine Oil Pollution Using Remote Sensing." *Water, Air, & Soil Pollution* 233 (8): 336. https://doi.org/10.1007/s11270-022-05752-0.

El-Magd, I. A., M. Zakzouk, A. M. Abdulaziz, and E. M. Ali. 2020. "The Potentiality of Operational Mapping of Oil Pollution in the Mediterranean Sea Near the Entrance of the Suez Canal Using Sentinel-1 SAR Data." *Remote Sensing* 12 (8): 1352. https://doi.org/10.3390/rs12081352.

El-Magd, I. A., M. Zakzouk, E. M. Ali, and A. M. Abdulaziz. 2021. "An Open Source Approach for Near-Real Time Mapping of Oil Spills Along the Mediterranean Coast of Egypt." *Remote Sensing* 13 (14): 2733. https://doi.org/10.3390/rs13142733.

European Maritime Safety Agency. 2017. "The CleanSeanet Service: Taking Measurements to Detect and Deter Marine Pollution." Brochure. Accessed May 31, 2023. http://www.emsa.europa.eu/csn-menu/download/2913/2123/23.html.

European Space Agency. 2013. "Data Distribution Schedule." Accessed April 19, 2023. https://sentinel.esa.int/web/sentinel/missions/sentinel-1/data-distribution-schedule .

European Space Agency. 2020. "SNAP - ESA Sentinel Application Platform V8.0." http://step.esa.int .

European Space Agency. 2022a. "End of Mission of the Copernicus Sentinel-1B Satellite." Accessed August 3, 2022. https://sentinels.copernicus.eu/web/sentinel/-/end-of-mission-of-the-copernicus-sentinel-1b-satellite/.

European Space Agency. 2022b. "Launches Secured for five Sentinel Satellites." Accessed December 7, 2022. https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Launches_secured_for_five_Sentinel_satellites.

Everingham, M., L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman. 2010. "The PASCAL Visual Object Classes (VOC) Challenge." *International Journal of Computer Vision* 88 (2): 303–338. https://doi.org/10.1007/s11263-009-0275-4.

Fan, J., and C. Liu. 2023. "Multitask GANs for Oil Spill Classification and Semantic Segmentation Based on SAR Images." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 16:2532–2546. https://doi.org/10.1109/JSTARS.2023.3249680.

Gertman, I., A. Murashkovsky, V. Levin, G. Kallos, and D. S. Rosen 2006. "Wave Monitoring and Wind Input as Key Issues in Operational Wave Forecasting Systems." In *Proceedings of the Fourth International Conference on EuroGOOS 'European Operational Oceanography: Present and Future'*, edited by H. Dahlin, N. C. Flemming, P. Marchand, and S. E. Petersson, 743–749. Brest, France: EuroGOOS Office and European Commission.

Goldman, R., E. Biton, E. Brokovich, S. Kark, and N. Levin. 2015. "Oil Spill Contamination Probability in the Southeastern Levantine Basin." *Marine Pollution Bulletin* 91 (1): 347–356. https://doi.org/10.1016/j.marpolbul.2014.10.050.

Goldman, R., S. Efrati, Y. Lehahn, I. Gertman, and E. Heifetz. 2014. "Comparison of Remotely Sensed Chlorophyll and Lagrangian Coherent Structures for Velocity Field Validation." In *EGU General Assembly Conference Abstracts*, 10332. Vienna, Austria.

Gonzalez, R. C., and R. E. Woods. 2018. *Digital Image Processing*. New York, USA: Pearson Education.

Hovland, H. A., J. A. Johannessen, and G. Digranes. 1994. "Slick Detection in SAR Images." In *Proceedings of 1994 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Vol. 4, 2038–2040. Pasadena, CA, USA.

Huang, X., B. Zhang, W. Perrie, L. Yingcheng, and C. Wang. 2022. "A Novel Deep Learning Method for Marine Oil Spill Detection from Satellite Synthetic Aperture Radar Imagery." *Marine Pollution Bulletin* 179:113666. https://doi.org/10.1016/j.marpolbul.2022.113666.

Kallos, G., S. Nickovic, A. Papadopoulos, D. Jovic, O. Kakaliagou, N. Misirlis, L. Boukas, et al. 1997. "The Regional Weather Forecasting System SKIRON: An Overview." In *Proceedings of the Symposium on Regional Weather Prediction on Parallel Computer Environments*, Vol. 15, 17. Athens, Greece: University of Athens, Greece.

Korres, G., and A. Lascaratos. 2003. "A One-Way Nested Eddy Resolving Model of the Aegean and Levantine Basins: Implementation and Climatological Runs." *Annales Geophysicae* 21 (1): 205–220. https://doi.org/10.5194/angeo-21-205-2003.

Kostianoy, A., E. A. Kostianaia, and D. M. Soloviev. 2020. "Oil Pollution in the Mediterranean Waters of Egypt" Edited by Elbeih, S. F., Negm, A. M., Kostianoy, A. *Environmental Remote Sensing in Egypt*. Springer Geophysics. Cham: Springer. 305–328. https://doi.org/10.1007/978-3-030-39593-3_11.

Krestenitis, M., G. Orfanidis, K. Ioannidis, K. Avgerinakis, S. Vrochidis, and I. Kompatsiaris. 2019. "Oil Spill Identification from Satellite Images using Deep Neural Networks." *Remote Sensing* 11 (15): 1762. https://doi.org/10.3390/rs11151762.

Li, Y., W. Huang, X. Lyu, S. Liu, Z. Zhao, and P. Ren. 2022. "An Adversarial Learning Approach to Forecasted Wind Field Correction with an Application to Oil Spill Drift Prediction." *International Journal of Applied Earth Observation and Geoinformation* 112:102924. https://doi.org/10.1016/j.jag.2022.102924.

Mackay, D. 1980. "Calculation of the Evaporation Rate of Volatile Liquids." In *Proceedings of 1980 National Conference on Control of Hazardous Material Spills*, 361–368. Louisville, KY, USA.

Mackay, D., I. Buist, R. Mascarenhas, and S. Paterson. 1980. *Oil Spill Processes and Models*. En40-980/1980E-PDF. Ottawa, Ontario, Canada: Environmental Impact Control Directorate.

Mera, D., J. M. Cotos Veronica Bolon-Canedo, and A. Alonso-Betanzos. 2017. "On the Use of Feature Selection to Improve the Detection of Sea Oil Spills in SAR Images." *Computers & Geosciences* 100:166–178. https://doi.org/10.1016/j.cageo.2016.12.013.

Migliaccio, M., F. Nunziata, and A. Buono. 2015. "SAR polarimetry for sea oil slick observation." *International Journal of Remote Sensing* 36 (12): 3243–3273. https://doi.org/10.1080/01431161.2015.1057301.

Papadopoulos, A., P. Katsafados, and G. Kallos. 2001. "Regional Weather Forecasting for Marine Application." *Global Atmosphere Ocean System* 8 (2–3): 219–237. https://doi.org/10.1080/1023673029000003543.

Polinov, S., R. Bookman, and N. Levin. 2021. "Spatial and Temporal Assessment of Oil Spills in the Mediterranean Sea." *Marine Pollution Bulletin* 167:112338. https://doi.org/10.1016/j.marpolbul.2021.112338.

Redmon, J., S. Divvala, R. Girshick, and A. Farhadi. 2016. "You Only Look Once: Unified, Real-Time Object Detection." In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 779–788. Las Vegas, NV, USA.

Shaban, M., R. Salim, H. Abu Khalifeh, A. Khelifi, A. Shalaby, S. El-Mashad, A. Mahmoud, M. Ghazal, and A. El-Baz. 2021. "A Deep-Learning Framework for the Detection of Oil Spills from SAR Data." *Sensors* 21 (7): 7. https://doi.org/10.3390/s21072351.

Singha, S., T. J. Bellerby, and O. Trieschmann. 2013. "Satellite Oil Spill Detection Using Artificial Neural Networks." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 6 (6): 2355–2363. https://doi.org/10.1109/JSTARS.2013.2251864.

Singha, S., R. Ressel, D. Velotto, and S. Lehner. 2016. "A Combination of Traditional and Polarimetric Features for Oil Spill Detection Using TerraSAR-X." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 9 (11): 4979–4990. https://doi.org/10.1109/JSTARS.2016.2559946.

Singha, S., M. Vespe, and O. Trieschmann. 2013. "Automatic Synthetic Aperture Radar Based Oil Spill Detection and Performance Estimation via a Semi-Automatic Operational Service Benchmark." *Marine Pollution Bulletin* 73 (1): 199–209. https://doi.org/10.1016/j.marpolbul.2013.05.022.

Solberg, A. H. S., and R. Solberg. 1996. "A Large-Scale Evaluation of Features for Automatic Detection of Oil Spills in ERS SAR Images." In *Proceedings of 1996 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Vol. 3, 1484–1486. Lincoln, NE, USA.

Solberg, A. H. S., G. Storvik, R. Solberg, and E. Volden. 1999. "Automatic Detection of Oil Spills in ERS SAR Images." *IEEE Transactions on Geoscience and Remote Sensing* 37 (4): 1916–1924. https://doi.org/10.1109/36.774704.

Stathakis, D., K. Topouzelis, and V. Karathanassi. 2006. "Large-Scale Feature Selection Using Evolved Neural Networks." In *Image and Signal Processing for Remote Sensing XII*, edited by L. Bruzzone, 636513. Vol. 6365. Stockholm, Sweden: International Society for Optics and Photonics, SPIE. https://doi.org/10.1117/12.688149.

Stevens, J. 2020. "NASA Earth Observatory Map." Accessed September 8, 2020. https://visibleearth.nasa.gov/images/147190/explorer-base-map.

Tang, D., J. Sun, L. Zhou, S. Wang, R. P. Singh, and G. Pan. 2019. "Ecological Response of Phytoplankton to the Oil Spills in the Oceans." *Geomatics, Natural Hazards and Risk* 10 (1): 853–872. https://doi.org/10.1080/19475705.2018.1549110.

Tings, B., A. Pleskachevsky, and S. Wiehle. 2023. "Comparison of Detectability of Ship Wake Components Between C-Band and X-Band Synthetic Aperture Radar Sensors Operating Under Different Slant Ranges." *ISPRS Journal of Photogrammetry and Remote Sensing* 196:306–324. https://doi.org/10.1016/j.isprsjprs.2022.12.008.

Topouzelis, K. N. 2008. "Oil Spill Detection by SAR Images: Dark Formation Detection, Feature Extraction and Classification Algorithms." *Sensors* 8 (10): 6642–6659. https://doi.org/10.3390/s8106642.

Topouzelis, K., V. Karathanassi, P. Pavlakis, and D. Rokos. 2007. "Detection and Discrimination Between Oil Spills and Look-Alike Phenomena Through Neural Networks." *ISPRS Journal of Photogrammetry and Remote Sensing* 62 (4): 264–270. https://doi.org/10.1016/j.isprsjprs.2007.05.003.

Wessel, P., and W. H. F. Smith. 1996. "A Global, Self-Consistent, Hierarchical, High-Resolution Shoreline Database." *Journal of Geophysical Research: Solid Earth* 101 (B4): 8741–8743. https://doi.org/10.1029/96JB00104.

Yang, Y.-J., S. Singha, and R. Mayerle. 2022. "A Deep Learning Based Oil Spill Detector Using Sentinel-1 SAR Imagery." *International Journal of Remote Sensing* 43 (11): 4287–4314. https://doi.org/10.1080/01431161.2022.2109445.

Zodiatis, G., G. Coppini, L. Perivoliotis, R. Lardner, T. Alves, N. Pinardi, S. Liubartseva, M. De Dominicis, E. Bourma, and A. Augusto Sepp Neves. 2017. "Numerical Modeling of Oil Pollution in the Eastern Mediterranean Sea." In *The Handbook of Environmental Chemistry*, edited by A., Carpenter and A. Kostianoy, 215–254. Vol. 83. Cham: Springer. https://doi.org/10.1007/698_2017_131.

Zodiatis, G., M. De Dominicis, L. Perivoliotis, H. Radhakrishnan, E. Georgoudis, M. Sotillo, R. W. Lardner, et al. 2016. "The Mediterranean Decision Support System for Marine Safety Dedicated to Oil Slicks Predictions." *Deep Sea Research Part II: Topical Studies in Oceanography* 133:4–20. https://doi.org/10.1016/j.dsr2.2016.07.014.

Zodiatis, G., R. Lardner, T. M. Alves, Y. Krestenitis, L. Perivoliotis, S. Sofianos, and K. Spanoudaki. 2017. "Oil Spill Forecasting (Prediction)." *Journal of Marine Research* 75 (6): 923–953. https://doi.org/10.1357/002224017823523982.

Zodiatis, G., R. Lardner, D. Solovyov, X. Panayidou, and M. De Dominicis. 2012. "Predictions for Oil Slicks Detected from Satellite Images Using MyOcean Forecasting Data." *Ocean Science* 8 (6): 1105–1115. https://doi.org/10.5194/os-8-1105-2012.