Stacked Autoencoders for Outlier Detection in Over-The-Horizon Radar Signals

Eftychios Protopapadakis[1], Athanasios Voulodimos[1], Anastasios Doulamis[1], Nikolaos Doulamis[1], Dimitrios Dres[2], Matthaios Bimpas[2]

[1] *National Technical University of Athens, Athens 15780, Greece*
[2] *Telesto Technologies, Cholargos 15561, Greece*

Abstract: Detection of outliers in radar signals is a considerable challenge in maritime surveillance applications. High-Frequency Surface-Wave (HFSW) radars have attracted significant interest as potential tools for long-range target identification and outlier detection at Over-The-Horizon (OTH) distances. However, a number of disadvantages, such as their low spatial resolution and presence of clutter, have a negative impact on their accuracy. In this paper, we explore the applicability of deep learning techniques for detecting deviations from the norm in behavioral patterns of vessels (outliers) as they are tracked from an OTH radar. The proposed methodology exploits the non-linear mapping capabilities of deep stacked autoencoders in combination with density based clustering. A comparative experimental evaluation of the approach shows promising results in terms of the proposed methodology's performance.

Keywords: Outlier detection, over-the-horizon (OTH) radar, stacked autoencoders, OPTICS algorithm.

1 Introduction

Detection of targets and outliers in radar signals is a research issue that has gained significant attention in the academic and industrial research community, mainly because of the important associated impact of relevant applications in surveying of large areas. High-Frequency Surface-Wave (HFSW) radars is a category of radars that operate at the frequency band 3-30 MHz and, in contrast with other radars, use ground wave or sky wave propagation and ionospheric reflections of the electromagnetic waves for target detection, which allows for achieving longer ranges, where microwave radars cannot perform [1], but to the detriment of the attained accuracy. For many years, HFSW radars, or Over-The-Horizon (OTH) radars, as they are commonly known, have been used to remotely measure oceanographic parameters, providing information about surface currents, wave spectra, wind direction and intensity, etc. [2]. Their extraordinary range (up to 200 nautical miles) combined with their continuous mode of operation make for an ideal

candidate tool for long-range oceanic surveillance. However, many associated weaknesses, e.g. low spatial resolution, high non-linearity, and important presence of clutter, negatively impact their performance as early-warning tools for detection, tracking and identification of vessels.

The promising capabilities of OTH radars have attracted significant interest from the research community and have already resulted to various approaches (e.g. [3], [4]). Nevertheless, related research issues continue to present significant challenges, which can be attributed to few reasons, briefly described below:

(i) different targets may present similar dielectric and frequency properties thus making it hard to make a clear distinction among them.
(ii) given multipath propagation effects of rough surfaces, scattering from some objects tends to overwhelm the weak backscattering of targets.
(iii) due to the changes in atmosphere and ground conditions, noise is added which can confuse the analysis of a radar signal.
(iv) ocean and ionospheric clutter generate noise especially for HFSW radars.

On a different note, the surge of deep learning and the great results it has produced in other signal analysis domains, such as computer vision, speech recognition and natural language processing, creates certain expectations regarding its potential efficacy in radar signal analysis applications. Deep learning allows computational models of multiple processing layers to learn and represent data with multiple levels of abstraction mimicking how the brain perceives and processes multimodal information, thereby implicitly capturing intricate structures of large-scale data. Complex abstractions are learnt at a given level based on relatively simpler abstractions formulated in the preceding layer in the hierarchy.

The goal of this paper is to present a framework for detecting deviations from the norm in behavioral patterns of vessels (henceforth called *outliers*), as they are tracked from an OTH radar. The proposed methodology exploits the non-linear mapping capabilities of deep stacked autoencoders (SAs) [5] in combination with density based clustering. Stacked Autoencoders are used in an unsupervised way to map the track history of any vessel into a compact and informative feature vector. Then, at any moment all tracked ships are projected into a new feature space and clustered using density based algorithms, such as OPTICS [6]. The outcome of the clustering stage then indicates possible outliers.

The remainder of this paper is structured as follows: Section 2 presents an overview of the related work. In Section 3 we describe in detail the proposed methodology for outlier detection in OTH radar signals, which is followed by the experimental evaluation of the methodology in Section 4. Finally, Section 5 concludes the paper.

2 Related work

In the literature, several signal processing and machine learning methods have been investigated and proposed to acquire more reliable data with lower noise and extract semantic information

from radar signals. Kouemou and Opitz [7] introduced a wavelet-based feature analysis combined with Hidden Markov Models (HMM) to classify real radar signals into predefined categories. Spectral analysis [8] is used by Garbanzo-Salas and Hocking [9] for detecting small objects from harmonic pulse radar data. The use of on-line bootstrapping machine learning tools to improve target detection rate of radar signals is also one major research area [10]. Radar data can be analyzed using the concepts of transfer learning since often we have available only a small number of labelled data while the majority of signals captured are unlabelled (non-annotated) [11]. Other works focus on modeling of ionospheric disturbances on spaceborne interferometric Synthetic Aperture Radar (SAR) via Echo-State Networks [12], [13] or ensemble classifiers [14].

De-noising techniques for radar signals include low level processing such as the median filter or other non-linear convolution schemes [15]. Other approaches spatially or temporally decompose radar signals by wavelet transforms [16], [17]. This way, we can find patterns distributed on space and time domain to improve targets detection efficiency. These methodologies can be extended to the analysis of Synthetic Aperture Radar (SAR) images [11], or by incorporating sparsity-based signal analysis [18]. A neural network based scheme for detecting salient objects in SAR images is recently presented [19]. The goal is to identify changes in SAR content. A similar approach for detecting changes using non-linear stacked restricted Boltzmann machines is given in the work of Lie et al. [20], while multi-layered feature learning to improve detection accuracy of SAR images is described by Xie et al. [21]. Furthermore, low-power HF surface-wave (HFSW) radars have demonstrated to be a cost-effective long-range early-warning sensor for ship detection and tracking [22], [23]. A detailed description of various ways in which HFSW radar technology can be used for maritime surveillance is provided by Braca et al. [24].

Regarding deep vs. "shallow" learning schemes, traditional machine learning techniques exploit shallow architectures, i.e. they use a single layer for data/feature transformation, even in a highly non-linear space. Shallowness refers here to the simplicity of these architectures that use only one (or few) layer(s) of processing, responsible for transforming the raw input signals or features into the problem-specific feature space. Instead, in a deep learning paradigm, the architectures are composed of many (deep) non-linear processing stages [25]. Deep learning has been extensively applied in many fields, such as computer vision [26] (e.g. behavior recognition [27] and human tracking [28]) and speech recognition [29]. However, its applicability in radar signal processing had not being investigated until very recently [30]. Even so, most of the proposed works pertain to object detection in SAR image data [31], essentially resembling visual analysis approaches.

3 The proposed methodology

The proposed methodology exploits the non-linear mapping abilities of stacked autoencoders (SAs) [5] in combination with density based clustering, to identify irregular occurrences, using over-the-horizon radar data. Such an approach is based on two main assumptions:

1. *The history of a naval vessel, in terms of speed, position, course, signal frequency or other related data, provided by a ground radar, suffices to extract meaningful features.*

2. *Unexpected deviation from the norm is observed for a few ships, denoted henceforth as outliers.*

The approach is relatively straightforward: Given a set of OTH data entries, SAs are used in an unsupervised way to map the track history of any vessel into a compact and informative feature vector. Then, at any moment all tracked ships are projected into a new feature space and clustered using OPTICS [6], a widely used density based algorithm. The clustering outcome informs about possible outliers. In the following subsections, the different stages of the methodology are presented, after a brief description of the data involved. Figure 1 provides a high-level view of the proposed approach.

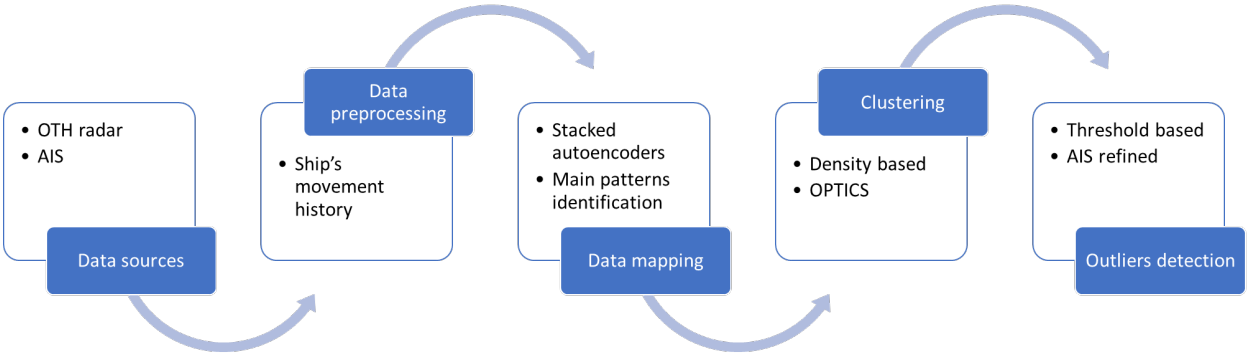


Figure 1. Proposed approach flowchart.

3.1 OTH and AIS data

Heterogeneous data, such as automatic identification system (AIS) data, high-frequency surface wave (HFSW) radar data, and synthetic aperture radar (SAR) data, have been exploited in research for maritime surveillance purposes [32]. In our case, two sources of information were fused to support the outlier detection process: OTH radar and AIS data.

The OTH radar data used for the setting and evaluation of the presented work was acquired by the HFSW STRADIVARIUS radar by Diginext [33]. OTH radar detection (plot) and tracking (track) data are the output of the OTH radar for a given period. The plot and track data provided include estimated position coordinates, velocity, course, Doppler frequency, global and local noise levels, azimuth, and other parameters, appropriately timestamped.

On a different note, AIS is an automatic tracking system used for collision avoidance on ships and by vessel traffic services. AIS information supplements marine radar, which continues to be the primary method of collision avoidance for water transport. Vessels equipped with AIS transceivers can be tracked by AIS base stations located along coast lines. The International Maritime Organization's International Convention for the Safety of Life at Sea requires AIS to be present aboard international voyaging ships with gross tonnage of 300 or more, and all passenger ships regardless of size [34]. AIS reports contain both dynamic information (e.g., latitude, longitude, course over ground, speed over ground, and time) and static information (e.g., vessel type and dimension information).

3.2 Density based clustering as a basis for outlier detection

Clustering refers to the task of identifying groups or clusters in a data set. In density-based clustering, a cluster is a set of data objects spread in the data space over a contiguous region of high density of objects. Density-based clusters are separated from each other by contiguous regions of low density of objects. Data objects located in low-density regions are typically considered noise or outliers [35]. OPTICS algorithm [6], as one among various approaches for hierarchical density based clustering, includes ordering points to identify the clustering structure. OPTICS is based on DBSCAN [36] and the work of Stuetzle [37].

OPTICS computes a Minimum Spanning Tree (MST) of the data, where edge weights represent pairwise distances. These distances are smoothed by a density estimator, called core distance. The core distance of a point $\boldsymbol{x}_i$ is the smallest threshold r such that $\boldsymbol{x}_i$ is still considered a core object by the DBSCAN algorithm, i.e., $\boldsymbol{x}_i$ has at least k objects in its neighborhood within radius r. The resulting distance, which is used to construct the MST, is called reachability distance (RD). Taking k as input parameter for smoothing the density estimation, the reachability distance of point $\boldsymbol{x}_i$ is defined relative to a reference object $\boldsymbol{y}$ as the minimum of the core distance of $\boldsymbol{y}$ and the actual distance between $\boldsymbol{x}_i$ and $\boldsymbol{y}$. The outcome of the algorithm can provide us information about the clustering of the objects (see sec. 3.4).

3.3 Using stacked autoencoders for data representation

Density based algorithms, traditionally, use the Euclidian distance metric [38]. Such distance metrics are prone to high dimensionality related problems. If we have a feature space of many dimensions, i.e. the tracked course of a ship, clustering performance decreases.

Let $\boldsymbol{n}$ and $\boldsymbol{m}$ be points drawn from a d-dimensional Gaussian distribution, so that $\boldsymbol{n}\sim N(\mu_1, \sigma_1^2 \cdot \boldsymbol{I})$ and $\boldsymbol{m}\sim N(\mu_2, \sigma_2^2 \cdot \boldsymbol{I})$. Then their expected distance satisfies [39]:

$$\begin{aligned} E\{\|\boldsymbol{n}-\boldsymbol{m}\|^2\} &= E\left\{\sum_{i=1}^{d} |n_i - m_i|^2\right\} \\ &= \sum_{i=1}^{d} \{Var(n_i - m_i) + E\{n_i - m_i\}^2\} \\ &= d \cdot (\sigma_1^2 + \sigma_2^2) + \|\mu_1 - \mu_2\|^2 \end{aligned} \tag{3.1}$$

Thus, the term $d \cdot (\sigma_1^2 + \sigma_2^2)$, where d is a scalar denoting the dimensions of the Gaussian distribution, overshadows the informative term $\|\mu_1 - \mu_2\|^2$. At this point, the need of robust low-dimension features becomes apparent. In such cases the use of autoencoders is advised [5].

An autoencoder is a neural network that is trained to attempt to copy its input to its output. Internally, it has a hidden layer h that describes a code used to represent the input. The network may be viewed as consisting of two parts: an encoder function $h = f(x)$ and a decoder that produces a reconstruction $r = g(h)$. Autoencoders are designed to be unable to learn to copy perfectly, since they are trained such that $g(f(\boldsymbol{x})) \approx \boldsymbol{x}$ instead of $(f(\boldsymbol{x})) = \boldsymbol{x}$. The model often

learns useful properties of the data, because it is forced to prioritize which aspects of the input should be copied.

Usually, training the autoencoder to perform the input copying task will result in h taking on useful properties, constraining h to have smaller dimension than x. An autoencoder whose code dimension is less than the input dimension is called undercomplete. Learning an undercomplete representation forces the autoencoder to capture the most salient features of the training data

The learning process is described simply as minimizing a loss function, e.g. $L\left(\boldsymbol{x}, g\left(f(\boldsymbol{x})\right)\right)$, where L is a loss function penalizing $g\left(f(\boldsymbol{x})\right)$ or being dissimilar from $\boldsymbol{x}$, such as the mean squared error. When the decoder is linear and L is the mean squared error, an undercomplete autoencoder learns to span the same subspace as PCA. In this case, an autoencoder trained to perform the copying task has learned the principal subspace of the training data as a side-effect

A sparse autoencoder is simply an autoencoder whose training criterion involves a sparsity penalty $\Omega(\boldsymbol{h})$ on the code layer $\boldsymbol{h}$, in addition to the reconstruction error, i.e. $L\left(\boldsymbol{x}, g\left(f(\boldsymbol{x})\right)\right) + \Omega(\boldsymbol{h})$. Sparse autoencoders are typically used to learn features for another task such as classification. An autoencoder that has been regularized to be sparse must respond to unique statistical features of the dataset it has been trained on, rather than simply acting as an identity function.

The core idea of our work lies in using stacked autoencoders to capture a representation of the main patterns present in the data. By doing so, any outlier in data samples will not be explained well using that representation. In other words, outliers will have significant variations from the rest of the data.

3.4 Identifying outliers

The outlier detection is a combinatory threshold-based approach built on the interquartile range rule, as in [40], OPTICS output (see sec. 3.2), and AIS/OTH matched data (see sec. 3.5).

OPTICS outputs (i.e. reachability distances of the ordered ships) is treated as a continuous signal, over which we identify the peaks. Peaks correspond to significant changes between the closest compared vehicles. As such, anything that varies from the norm, has a peak, allowing us the easy identification of a possible outlier. Then, we calculate a threshold value $ths^{(t)}$ defined as $ths^{(t)} = \frac{1}{m}\sum_{i}^{m} RD_o(i)$, $m = \lceil 0.1 \cdot n_t \rceil$, where n_t denotes the number of ships at a time t and $\boldsymbol{RD}_o$ is the reachability distances vector, in a descending order.

In case that an outlier provides AIS data, the detection regarding that ship is ignored. At first, for a specific time instance, ships are ordered in a density-reachable way (Figure 2). Points close to each other should belong to the same cluster, unless there is a significant change in RD value. Then, the outlier RD value threshold is defined over 10% of highest RDs.

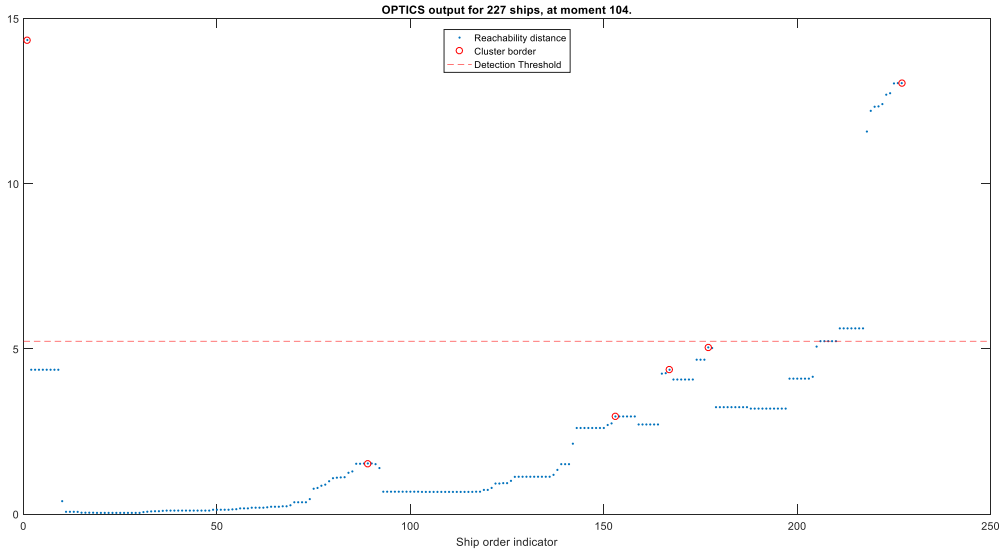


Figure 2. (Best viewed in color.) Illustration of an instance of the outlier detection mechanism at a specific time moment.

3.5 Matching OTH data to AIS

As explained in sec. 3.1, AIS data contain, among other, ships' trajectory points. These coordinates are compared to the radar ones, to identify the similarity among the trajectories. Let us denote as $\boldsymbol{T}_R^{(v_i)} = [t_1, \dots, t_p]$ the available discrete time instances, created from the ground radar for ship v_i, $i = 1, \dots, n$. The equivalent case for AIS data is $\boldsymbol{T}_{AIS}^{(v_j)} = [t_1, \dots, t_q]$ for any ship v_j, $j = 1, \dots, l$, that provides AIS data.

Figure 3 illustrates the available trajectories over a specified area for both radar and AIS data. At this point, we should note that trajectories are calculated for various time intervals, which do not, usually, coincide among the two systems. Typically, for the same ship $p > q$, in a ratio of four radar time instances to one AIS time instance. Also, note that $l < n$, so that a 1-to-1 match among radar and AIS tracked ships is not feasible. Therefore, we should consider both the temporal and the spatial information, to find the matches. The following algorithm (presented in pseudocode in Table 1) performs the vessel matching (Figure 4), given OTH and AIS information recorded at the same time (for further details about used data see Sec. 4.2).

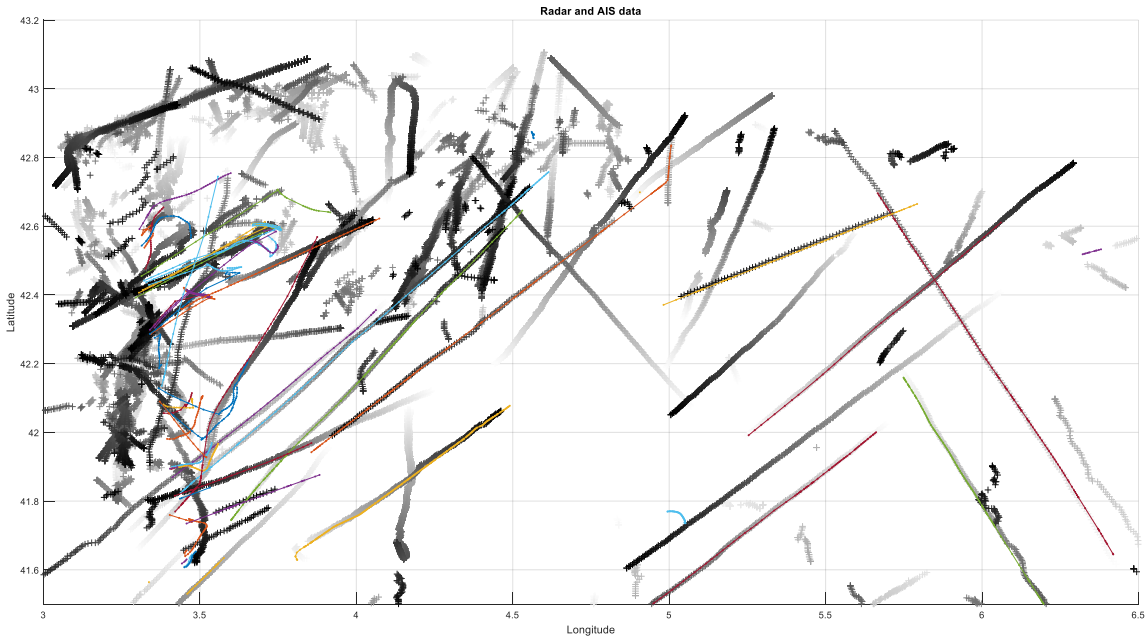


Figure 3. (Best viewed in color.) An illustration of the investigated ship trajectories. Ground radar trajectories are plotted in grayscale. The fading colors correspond to past times.

The matching process is based on a voting mechanism. For each of the radar tracked ship v_i, at a time instance t_w, $w = 1, \dots, p$, we calculate the k closest ships $\mathcal{v}_i$, according to their AIS position at the specific time. In order to identify the corresponding (closest) AIS time instance of ship $\mathcal{v}_i$ to radar entry v_i, at a time t_w^R, we calculate the time difference $\boldsymbol{T}_{diff}^{\mathcal{v}_i} = \left[t_1^{AIS} - t_w^R, \dots, t_q^{AIS} - t_w^R\right]$, then the corresponding time instance is given as $c_t = \arg \min_{\boldsymbol{T}_{diff}(m) \geq 0} \left\{\boldsymbol{T}_{diff}^{\mathcal{v}_i}(1), \dots, \boldsymbol{T}_{diff}^{\mathcal{v}_i}(q)\right\}$. In case that $c_t = \emptyset$, AIS entry $\mathcal{v}_j$ is not matched to v_i at time t_w. Then, once we have a set of matched ship instances $M_{v_i}^{t_j} = \{\mathcal{v}_r\}_{r=1}^{R}$,R<q, we find k closest entries to v_i according to their position (i.e. longitude, latitude), so that $M_{v_i}^{t_w} = \{\mathcal{v}_r\}_{r=1}^{k}$, $k \ll q$.

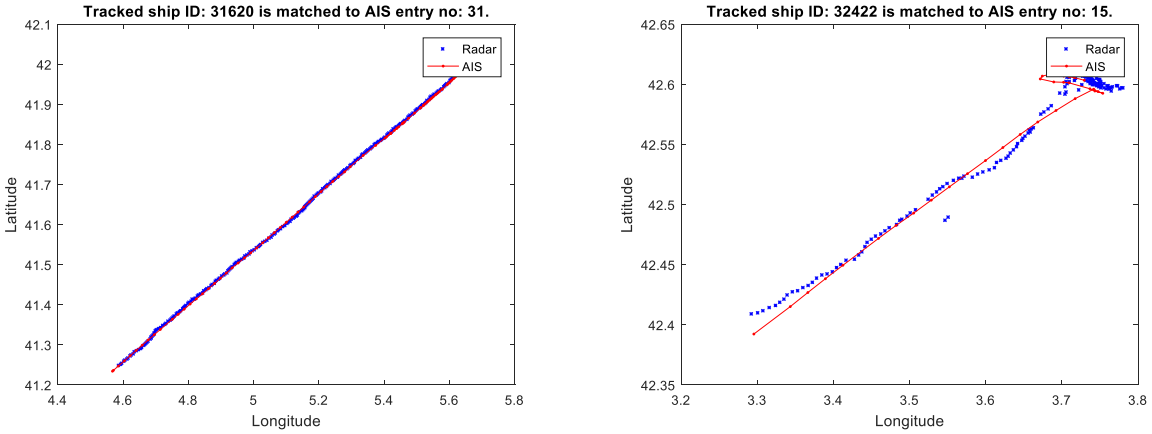
Table 1. The proposed algorithm in pseudocode.

Input: OTH and AIS trajectories and other provided information for a set of past time instances $\{t\}$ **Output**: $n \times m \times t$ matrix of the closest AIS entries to each OTH entry for a set of past time instances $\{t\}$
FOR each time instance t FOR each tracked ship v_i, $i = 1, \dots, n_t$ FOR each AIS transmitting ship $\mathcal{v}_j$ Check difference in track time and AIS transmission time $\boldsymbol{T}_{diff}^{\mathcal{v}_i}$ Find corresponding time instances c_t IF $c_t = \emptyset$

No AIS entry is matched to v_i
ELSE
Run $k$nn search using coordinates among v_i and $\{v_j\}_{c_t}$
Maintain 4 closest entries
END
END
END
END

Input: $n \times m \times t$ matrix of the closest AIS entries to each OTH entry for a set of past time instances $\{t\}$
Output: $n \times 1$ array with indices of the matched AIS to each one of the n OTH tracked vehicles

Initialize $n \times m$ matrix FOR the votes V_m
FOR each OTH tracked ship v_i
FOR each time instance t
IF ship n_i had ship m_j in the vicinity
$V_m[i][j] = V_m[i][j] + 1$
END
END
END
WHILE AIS vessels remain unmatched
$MatchIdx_{AIS2OTH} = arg\max_{\text{row}}(V_m)$
IF $length(MatchIdx_{AIS2OTH}) > 1$
KEEP AIS entry closest to OTH vehicle
END
Update V_m
END



(a) (b)

Figure 4. (Best viewed in color.) Illustration of matched trajectories between ground radar and AIS data (a) and matched trajectories despite the noise, due to minor course deviations (b).

4 Experimental results

In the following subsections, we describe the dataset utilized for the experiments, the performance evaluation metrics employed and the system setup details, before presenting the experimental evaluation of the proposed framework.

4.1 Computational complexity

Data preprocessing creates a set of k OTH data related entries, for a predefined set of past moments, for each one of the n_t tracked ships, at a moment t. Since both k and m are constants defined by the user, the required runtime is $O(n)$. The mapping process of a trained SA is $O(1)$ per datum, since SAs are neural networks with a defined number of neurons. OPTICS processes each point once, and performs one ϵ-neighborhood query during this processing. Given a spatial index that grants a neighborhood query in $O(log(n))$ runtime, an overall runtime of $O(n \cdot log(n))$ is obtained. The matching process between AIS and OTH entries require an overall runtime of $O(n \cdot m)$, $m_t \ll n_t$, since we compare each of the n_t OTH tracked ships to each of the m_t ships equipped with AIS. Table 2 displays the computational complexity of the different processing steps.

Table 2. Computational complexity of the different processing steps.

Processing step	Data preprocessing	Data mapping	Data clustering	OTH and AIS matching
Complexity	$O(n)$	$O(n)$	$O(n \cdot \log(n))$	$O(n \cdot m)$

4.2 Utilized dataset

The utilized dataset pertains to approximately 6 hours of data captured from the Mediterranean coast of France by DIGINEXT in July 2016 in the context of the RANGER EU Horizon 2020 project. AIS data for the same period were also obtained for use as ground truth.

A total of 556 ship entries were in this 6-hour dataset. The following data provided entries are used:

1. Longitude and latitude: position values provided in degrees. The typical range is $[-180,180]$ and $[-90,90]$ respectively.
2. Course and speed: course is calculated in degrees, typically in the range $[-180,180]$ and speed in m/s.
3. Doppler frequency: is calculated in Hz, typically in the range $[-0.5, 0.5]$.
4. Raw Rx azimuth: azimuth angle from the Rx site in the raw spatial grid (equivalent to the reception beam), typically in the range $[110,230]$.
5. local noise: noise level in the surrounding of the plot. It is calculated in (dBm), in the range $[-120, -40]$.
6. global noise: background noise level of all range-Doppler map. It is calculated in (dBm), in the range $[-120, -80]$.

4.3 Performance metrics

Formally, a cluster analysis can be described as the partitioning a number of N classification objects in K groups or clusters $\{C_k\}$, $k = 1, \ldots, K$. Given N objects $\boldsymbol{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$, where x_{ij} denotes the j-th element of $\boldsymbol{x}_i$. The grouping of all objects $\boldsymbol{x}_i$, $i = 1, \ldots, N$ in K clusters can be defined as follows:

$$w_{ki} = \begin{cases} 1, & iff\ \boldsymbol{x}_i \in C_k \\ 0, & otherwise \end{cases} \tag{4.1}$$

The above formulation ensures that the association of each object to a cluster is unique.; A unique association is a valid case for both hierarchical and partitioning cluster analysis. Given matrix $\boldsymbol{W}$, various internal quality indices have been calculated, to determine an optimal clustering.

4.3.1 Calinski Harabasz index

The Calinski Harabasz index (CHI) [41] is defined according to the following equation:

$$CHI(k) = \frac{T_{\boldsymbol{B}}/K - 1}{T_{\boldsymbol{W}}/N - K} \tag{4.2}$$

where $T_{\boldsymbol{B}}$ is defined as:

$$T_{\boldsymbol{B}} = \sum_{k=1}^{k} |\overline{\boldsymbol{C}}_k| \|\boldsymbol{C}_k - \overline{\boldsymbol{x}}\| \tag{4.3}$$

and $T_{\boldsymbol{W}}$ is defined as:

$$T_{\boldsymbol{W}} = \sum_{k=1}^{k} \sum_{i=1}^{N} w_{ki} \|\boldsymbol{x}_i - \overline{\boldsymbol{C}}_k\|^2 \tag{4.4}$$

$T_{\boldsymbol{W}}$ starts at a comparably large value. With increasing number of clusters k, approaching the optimal clustering solution in K^* groups, the value should significantly decrease due to an increasing compactness of each cluster. As soon as the optimal solution is exceeded an increase in compactness and thereby a decrease in value might still occur. However, any decrease in value should be notably smaller.

Calculated for each possible cluster solution, the maximum CHI value indicates the best cluster partitioning of the data.

4.3.2 Davies–Bouldin index

The Davies–Bouldin index (DBI) [42] is an internal evaluation scheme, where the validation of how well the clustering has been done is made using quantities and features inherent to the dataset. DBI is defined as follows:

$$DB(k) = \frac{1}{K} \sum_{k=1}^{K} R_k \tag{4.5}$$

where R_k is defined as:

$$R_k = max\left(\frac{\mathcal{S}_k + \mathcal{S}_j}{d_{kj}}\right), j = 1, \dots, K \;\; j \neq k \tag{4.6}$$

d_{kj} is a distance function, defined as $d_{kj} = \left\|\overline{\boldsymbol{x}}_k - \overline{\boldsymbol{x}}_j\right\|$ and $\mathcal{S}_k$ is defined as:

$$\mathcal{S}_k = \frac{1}{\sum_{i=1}^{N} w_{ki}} \sum_{i=1}^{N} w_{ki} \|\boldsymbol{x}_i - \overline{\boldsymbol{x}}_k\| \tag{4.7}$$

All the above equations assume that $k \in [1, K]$.

For each cluster C_k an utmost similar cluster—regarding their intra-cluster error sum of squares—is searched, leading to R_k. The index then defines the average over these values. In this case, the minimum index value corresponds to the best cluster solution.

4.3.3 Silhouette

The silhouette value is a measure of how similar an object is to its own cluster (cohesion) compared to other clusters (separation). The silhouette ranges from -1 to 1, where a high value indicates that the object is well matched to its own cluster and poorly matched to neighboring clusters. If most objects have a high value, then the clustering configuration is appropriate. If many points have a low or negative value, then the clustering configuration may have too many or too few clusters.

For each datum $\boldsymbol{x}_i$, let $\alpha(\boldsymbol{x}_i)$ be the average dissimilarity (distance) of $\boldsymbol{x}_i$ i with all other data within the same cluster C_k. Let $b(\boldsymbol{x}_i)$ be the lowest average dissimilarity of $\boldsymbol{x}_i$ to any other cluster $C_l, l \neq k$, of which $\boldsymbol{x}_i$ is not a member. We now define a silhouette as:

$$s(\boldsymbol{x}_i) = \frac{b(\boldsymbol{x}_i) - \alpha(\boldsymbol{x}_i)}{\max\{\alpha(\boldsymbol{x}_i), b(\boldsymbol{x}_i)\}} \tag{4.8}$$

thus, $s(\boldsymbol{x}_i) \in [-1,1]$. Values close to one indicate that the datum $\boldsymbol{x}_i$ is appropriately clustered at C_k. The average silhouette value over all data, i.e. $\bar{s} = \frac{1}{n}\sum_{i=1}^{n} s(\boldsymbol{x}_i)$, is another measurement for the quality of the generated clusters.

4.4 Experimental setup

The first step should be the definition of the feature space on which radar data are mapped. As a starting point, we investigated the dimensional space provided by PCA, maintaining 99.1% of the original variation. The adopted stacked autoencoder approach consists of three layers or four layers, depending on the PCA outcome. The loss function was the well-known mean square error [43] with L2 and Sparsity regularizers [44].

Ships track history is composed of 9 consecutive frames, each containing all data as described in sec. 4.2. Data are normalized using minmax approach, prior to mapping or clustering approach. The system ignores ships with a narrow appearance span. Any ship that has not enough sufficient entries, i.e. ¾ of past moments tracks is not taken under consideration.

4.5 Evaluation of results

OPTICS algorithm outcomes depend on the selection of minimum cluster size. We have investigated the clustering outputs assuming at least 2, 5, 8, 11, 14, 17, 20, 23 and 26 members in each cluster. Clustering over SA mapped data performed better than using raw or PCA mapped data, for most of investigated cases.

According to CHI (Figure 5), highest scores are achieved when using 26 ships per cluster. It is intriguing that cluster performance scores over raw data outperforms PCA mapped data scores. There is an increasing trend on the CHI as the minimum cluster size increases. The trend is clearly illustrated for SAs, less for raw data and slightly for PCA projected data.

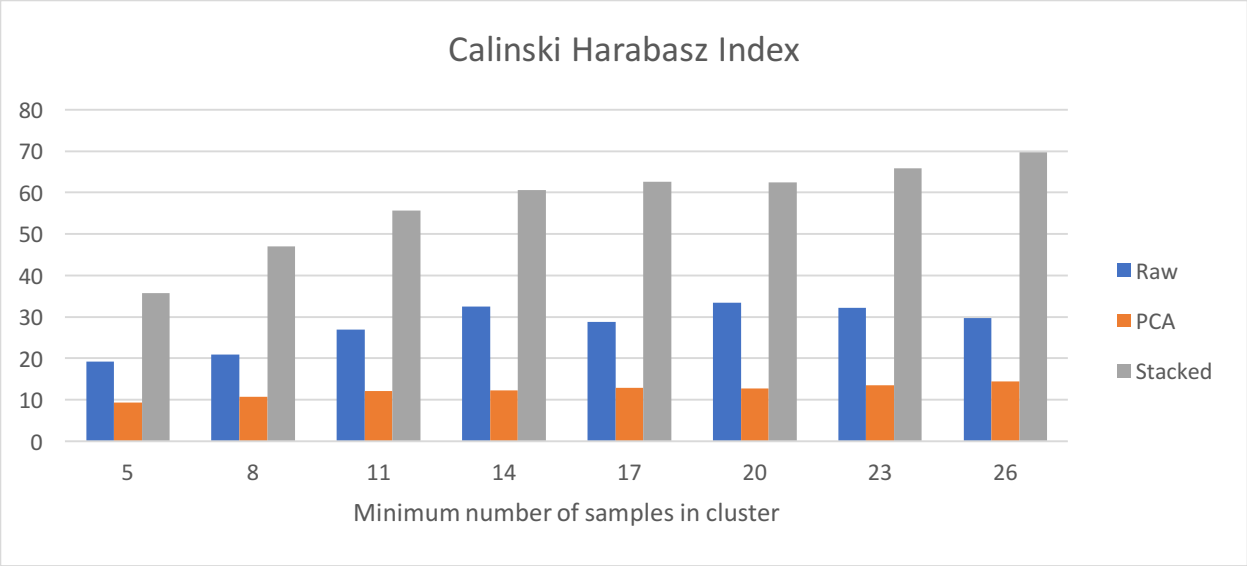


Figure 5. The impact of minimum cluster size (OPTICS input parameter) on Calinski Harabasz index average score. Stacked Autoencoders CHI scores are better in all the investigated cases, compared to PCA and raw data based clusters.

The next step was the investigation of DBI scores for the same minimum cluster size setup (Figure 6). This time, the best scores are achieved using 14 or 20 as the cluster size. SA mapping provides with better clustering scores in five out of seven investigated cases. Regardless the mapping method, CHI scores, over SA mapped data, improve as the number of clusters rises, but not in a monotonic way.

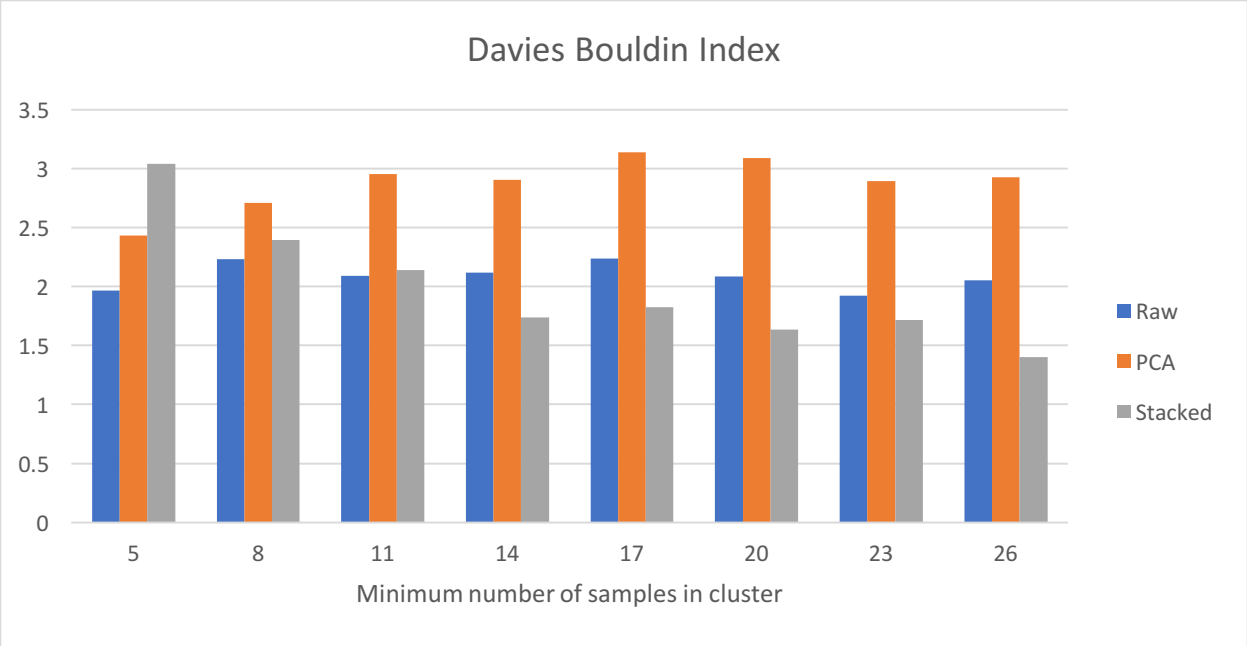


Figure 6. The impact of minimum cluster size (OPTICS input parameter) on Davies-Bouldin index average score. Stacked Autoencoders CHI scores are better in six out of eight investigated cases, compared to PCA based clusters, and five out of eight cases compared to raw data.

The last cluster performance metric was the average silhouette distance (Figure 7). Results suggest that accepting two ships as minimum cluster size is the best possible setup, for PCA mapped data. On the other hand, if we use SA for data mapping, the minimum cluster size should be set as 20.

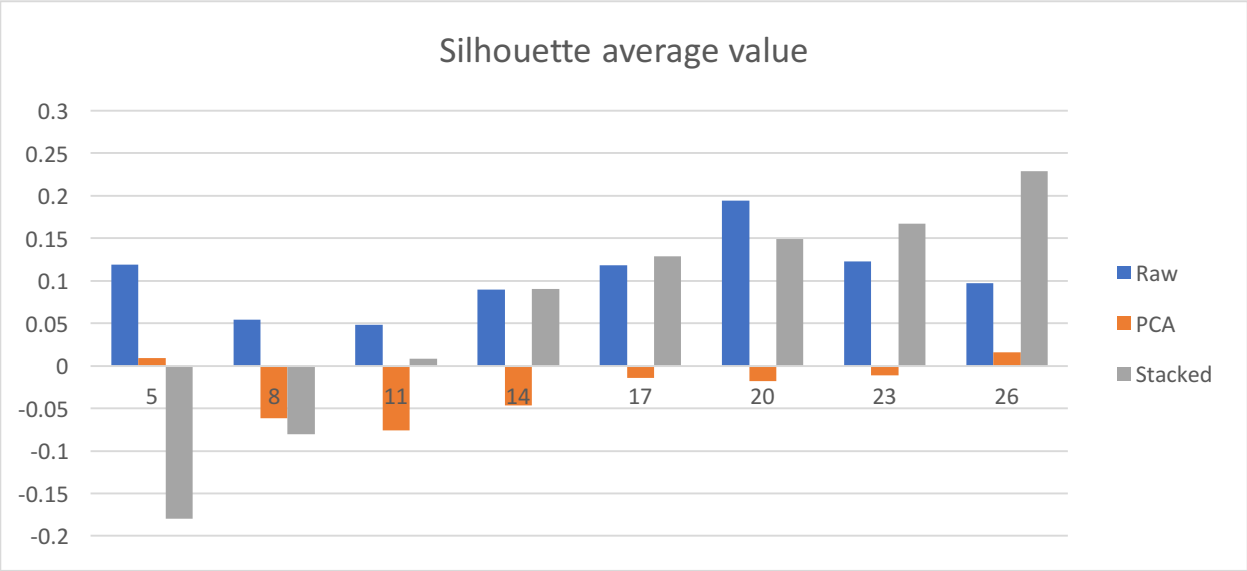


Figure 7. Impact of minimum cluster size (OPTICS input parameter) on silhouette average values. Stacked Autoencoders silhouette scores are better in five out of eight investigated cases, compared to Raw based clusters.

Another significant performance metric is the average reachability distance itself. The smaller the reachability distance of a point is, the higher is the density around it. The core idea of the proposed approach is that only outliers should vary significantly from the norm, on the projected feature space. Thus, all the ships, minus the outliers, should have similar feature values, which results in reduced reachability distances.

Providing more training data allows SA to adjust the mapping process to the norm. As illustrated in Figure 8 the average reachability distance tends to one, at a slow pace, while increasing the number of training samples. The variance of the RD is, also, reduced when using more time instances for training, as shown in Figure 9. Furthermore, SA mapping allows for the creation of more clusters compared to PCA or raw data clustering (Figure 10).

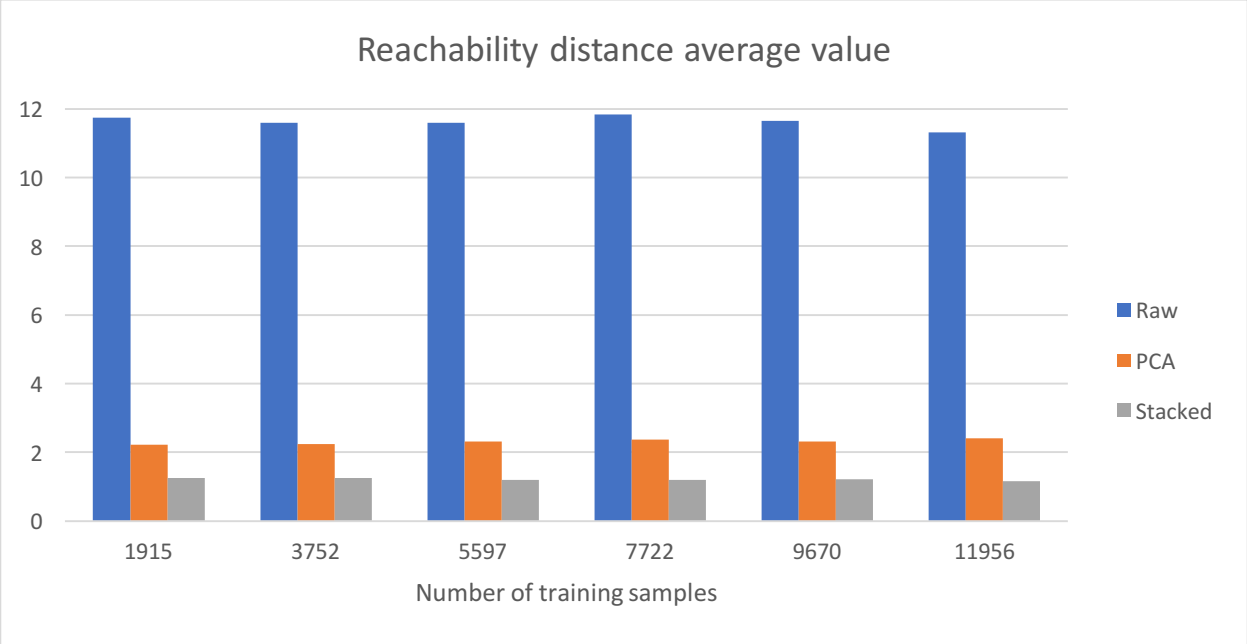


Figure 8. An illustration of how the number of training paradigms affects the average reachability distances (OPTICS outputs). Raw data average RD value exceeds 10, in each of the cases.

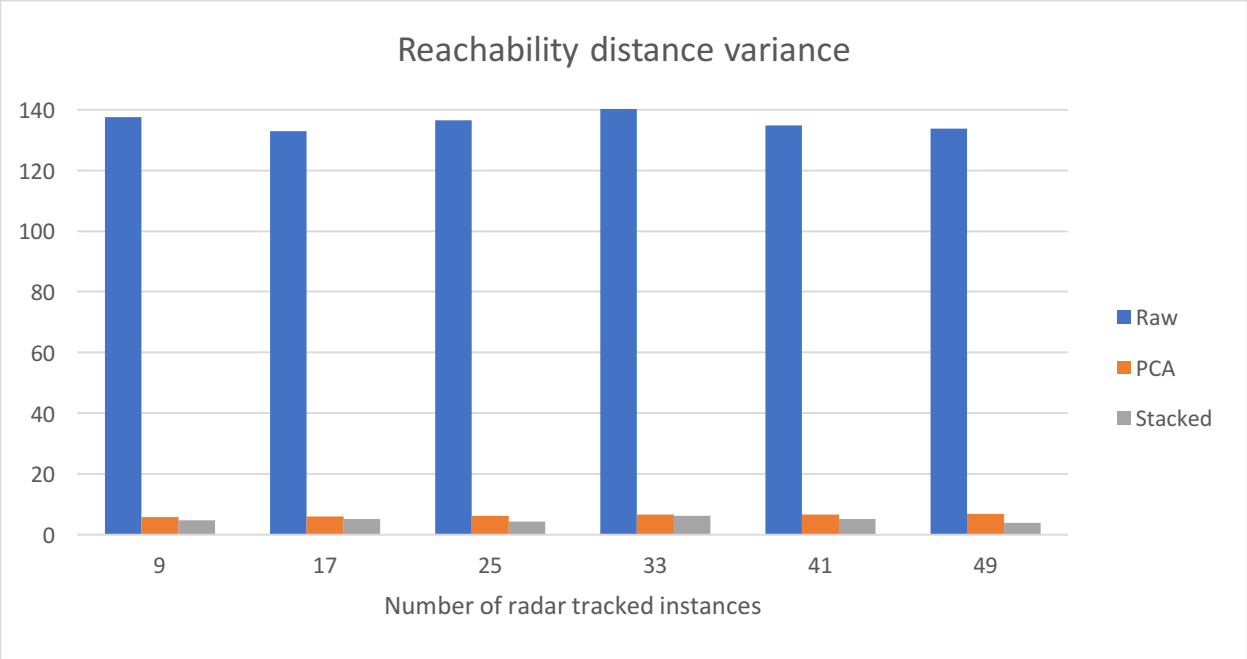


Figure 9. Illustration of the training period span effect on the variance in reachability distances. Raw data RD variance exceeds 40, in each case.

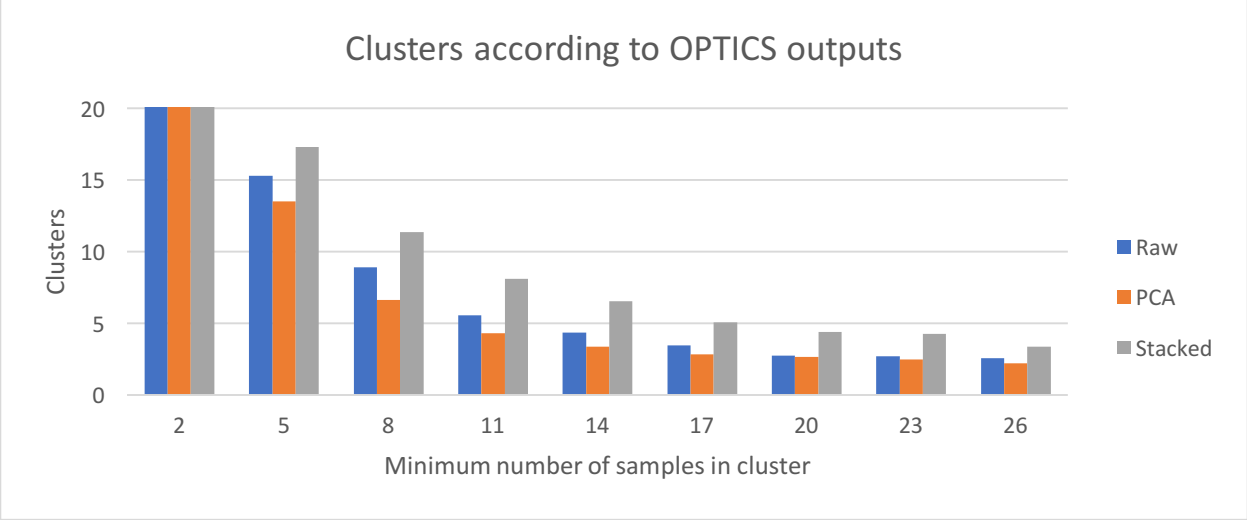
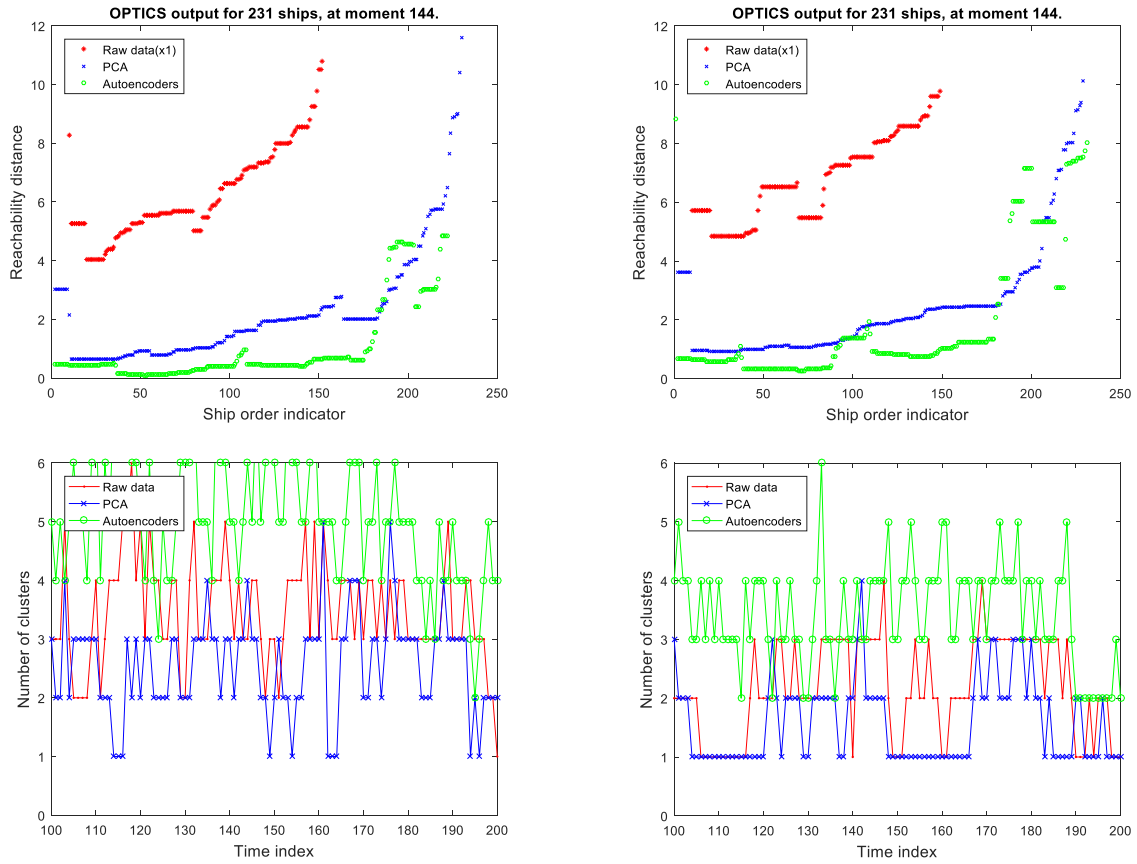


Figure 10. Average number of generated clusters given various mapping approaches. In all of the investigated cases (i.e. different minimum cluster size), SAs provide more clusters.

Regardless of the adopted feature mapping approach, OPTICS outputs are at least four times less in value, compared to calculated RDs using raw data (see Figure 11 (a), (b) top). Additionally, SAs result in more clusters, in most of the cases (see Figure 11 (a), (b) bottom). Increasing the number of minimum ships per cluster, close objects have almost identical reachability distances, resulting in almost linear sub regions, within RD curve.



(a) (b)

Figure 11. (Best viewed in color.) Comparison of OPTICS outputs over the same time instance, setting as minimum cluster size (a) 20 and (b) 26 ships. Stacked autoencoders result in more clusters than PCA or raw data; that implies more peaks in the signal, which leads to more outliers' detection.

The last step of the performance analysis provides with empirical findings. In most of the cases, SAs mapped data results in detection of more outliers compared to the other approaches (Figure 12). The maximum number of detected outliers was three. PCA resulted in no detections at any time.

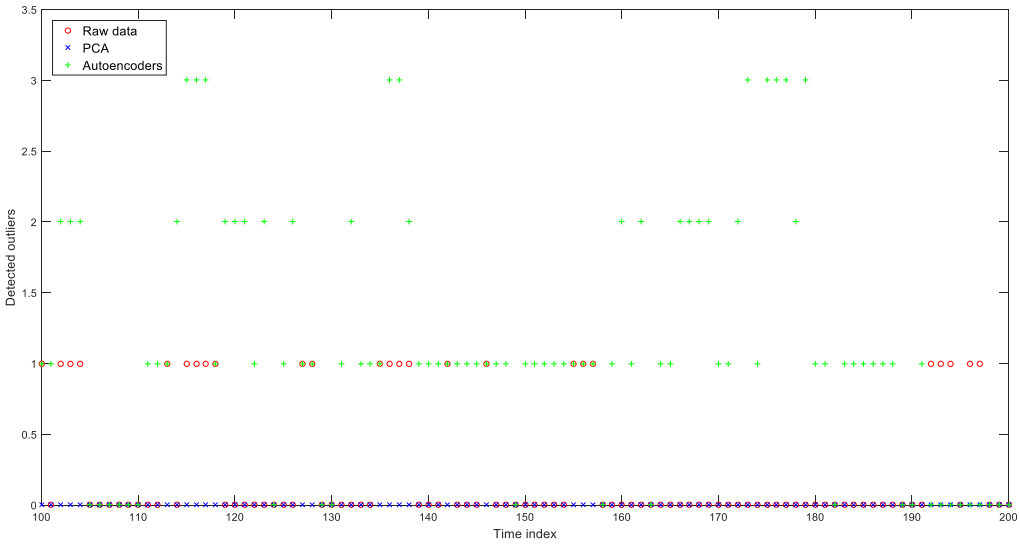


Figure 12. (Best viewed in color.) Illustration of the detected outliers through time. Using SAs' mapped data results in more outliers compared to the other approaches. Some of the selected outliers correspond to ships equipped with AIS transmitters.

There was the possibility of unwanted outlier identification. In particular, ships providing AIS data were considered, few times, possible outliers. Figure 13 illustrates the case. Typically, using SAs resulted in few possible outliers, which however were not accepted as valid detection, as explained in sec. 3.4.

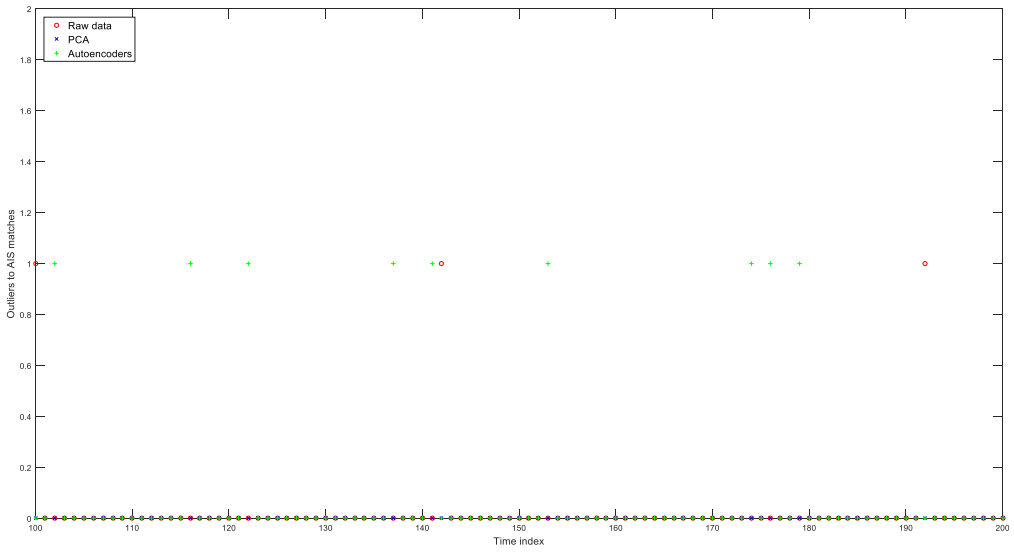


Figure 13. (Best viewed in color.) Illustration of the ships identified as possible outliers, while providing AIS data. Such cases are not considered as outliers.

5 Conclusions

In our article, a novel approach that identifies unexpected behavior in ship plot and track patterns, as captured by an OTH radar, has been presented. The core idea is the unsupervised development of a mapping process, which can project the raw data in a compact, lower feature space. Outliers projected to the same space should have significantly different values. Stacked autoencoders and PCA were used for the mapping process and compared against the exploitation of raw data, for the identification of unusual ship behavior. Density-based clustering algorithms (OPTICS) were employed for clustering-based outlier detection. Experimental results suggest that the approach based on SAs outperforms the other approaches in both generated cluster quality and outliers' identification.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

The research leading to these results has received funding from the European Commission's H2020 research and innovation programme, under grant agreement no 700478 (RANGER project). The authors would like to thank all project partners for their collaboration, and especially our partners from Diginext for the provision of the OTH and AIS data and documentation.

References

[1] L. Sevgi, A. Ponsford, and H. C. Chan, "An integrated maritime surveillance system based on high-frequency surface-wave radars. 1. Theoretical background and numerical simulations," *IEEE Antennas Propag. Mag.*, vol. 43, no. 4, pp. 28–43, Aug. 2001.

[2] S. Grosdidier, A. Baussard, and A. Khenchaf, "HFSW Radar Model: Simulation and Measurement," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 9, pp. 3539–3549, Sep. 2010.

[3] D. Nikolić *et al.*, "Multi-radar multi-target tracking algorithm for maritime surveillance at OTH distances," in *2016 17th International Radar Symposium (IRS)*, 2016, pp. 1–6.

[4] G. Vivone, P. Braca, and J. Horstmann, "Knowledge-Based Multitarget Ship Tracking for HF Surface Wave Radar Systems," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3931–3949, Jul. 2015.

[5] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press.

[6] M. Ankerst, M. M. Breunig, H.-P. Kriegel, and J. Sander, "OPTICS: Ordering Points to Identify the Clustering Structure," in *Proceedings of the 1999 ACM SIGMOD International Conference on Management of Data*, New York, NY, USA, 1999, pp. 49–60.

[7] G. Kouemou and F. Opitz, "Impact of Wavelet based signal processing methods in radar classification systems using Hidden Markov Models," in *2008 International Radar Symposium*, 2008, pp. 1–4.

[8] M. Garbanzo-Salas and W. K. Hocking, "Spectral analysis comparisons of Fourier-theory-based methods and minimum variance (Capon) methods," *J. Atmospheric Sol.-Terr. Phys.*, vol. 132, pp. 92–100, Sep. 2015.

[9] P.-H. Jau *et al.*, "Signal processing for harmonic pulse radar based on spread spectrum technology," *IET Radar Sonar Amp Navig.*, vol. 8, no. 3, pp. 242–250, Aug. 2013.

[10] L. Liu, D. Fu, Y. Zhu, D. Su, and M. Diao, "Applications of Bootstrap in Radar Signal Processing," in *Intelligent Computation in Big Data Era*, 2015, pp. 474–479.
[11] D. Xu *et al.*, "Compressive Sensing of Stepped-Frequency Radar Based on Transfer Learning," *IEEE Trans. Signal Process.*, vol. 63, no. 12, pp. 3076–3087, Jun. 2015.
[12] B. A. Massinas, A. Doulamis, N. Doulamis, and D. Paradissis, "An Echo State Network for Ionospheric Disturbances Behavior Modeling on Spaceborne Interferometric Synthetic Aperture Radar," in *AIAA SPACE 2013 Conference and Exposition*, American Institute of Aeronautics and Astronautics.
[13] B. A. Massinas, A. Doulamis, N. Doulamis, and D. Paradissis, "Applied Optimal Estimation for Ionospheric Disturbances Behavior on Spaceborne Interferometric Synthetic Aperture Radar Systems," in *AIAA SPACE 2015 Conference and Exposition*, American Institute of Aeronautics and Astronautics.
[14] B. A. Massinas, A. Doulamis, N. Doulamis, and D. Paradissis, "Ensemble Classifiers in Optimal Estimation for Ionospheric Disturbances Behavior on Spaceborne Interferometric SAR Systems," in *AIAA SPACE 2016*, American Institute of Aeronautics and Astronautics.
[15] M. Cao, S. Li, R. Wang, and N. Li, "Interferometric Phase Denoising by Median Patch-Based Locally Optimal Wiener Filter," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 8, pp. 1730–1734, Aug. 2015.
[16] X. Meng, Z. He, G. Feng, and B. Xiao, "An Improved Wavelet Denoising Algorithm for Wideband Radar Targets Detection," *Circuits Syst. Signal Process.*, vol. 32, no. 4, pp. 2003–2026, Aug. 2013.
[17] R. Ahmed, N. Maheshwari, and P. Lalla, "Wavelet based iterative thresholding for denoising of remotely sensed optical and synthetic aperture radar images," in *2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies*, 2014, pp. 1331–1335.
[18] M. C. Shastry, R. M. Narayanan, and M. Rangaswamy, "Sparsity-based signal processing for noise radar imaging," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 51, no. 1, pp. 314–325, Jan. 2015.
[19] M. Gong, J. Zhao, J. Liu, Q. Miao, and L. Jiao, "Change Detection in Synthetic Aperture Radar Images Based on Deep Neural Networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 1, pp. 125–138, Jan. 2016.
[20] J. Liu, M. Gong, J. Zhao, H. Li, and L. Jiao, "Difference representation learning using stacked restricted Boltzmann machines for change detection in SAR images," *Soft Comput.*, vol. 20, no. 12, pp. 4645–4657, Dec. 2016.
[21] H. Xie, S. Wang, K. Liu, S. Lin, and B. Hou, "Multilayer feature learning for polarimetric synthetic radar data classification," in *2014 IEEE Geoscience and Remote Sensing Symposium*, 2014, pp. 2818–2821.
[22] S. Maresca, P. Braca, J. Horstmann, and R. Grasso, "A network of HF surface wave radars for maritime surveillance: Preliminary results in the German Bight," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 6077–6081.
[23] S. Maresca, P. Braca, J. Horstmann, and R. Grasso, "Maritime Surveillance Using Multiple High-Frequency Surface-Wave Radars," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 5056–5071, Aug. 2014.
[24] P. Braca, S. Maresca, R. Grasso, K. Bryan, and J. Horstmann, "Maritime surveillance with multiple over-the-horizon HFSW radars: An overview of recent experimentation," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 30, no. 12, pp. 4–18, Dec. 2015.

[25] G. E. Hinton and R. R. Salakhutdinov, "Reducing the Dimensionality of Data with Neural Networks," *Science*, vol. 313, no. 5786, pp. 504–507, Jul. 2006.
[26] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, "Deep learning for visual understanding: A review," *Neurocomputing*, vol. 187, pp. 27–48, Apr. 2016.
[27] K. Makantasis, A. Doulamis, N. Doulamis, and K. Psychas, "Deep learning based human behavior recognition in industrial workflows," in *2016 IEEE International Conference on Image Processing (ICIP)*, 2016, pp. 1609–1613.
[28] N. Doulamis and A. Voulodimos, "FAST-MDL: Fast Adaptive Supervised Training of multi-layered deep learning models for consistent object tracking and classification," in *2016 IEEE International Conference on Imaging Systems and Techniques (IST)*, 2016, pp. 318–323.
[29] G. Hinton *et al.*, "Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 82–97, Nov. 2012.
[30] E. Mason, B. Yonel, and B. Yazici, "Deep learning for radar," in *2017 IEEE Radar Conference (RadarConf)*, 2017, pp. 1703–1708.
[31] S. Deng, L. Du, C. Li, J. Ding, and H. Liu, "SAR Automatic Target Recognition Based on Euclidean Distance Restricted Autoencoder," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. PP, no. 99, pp. 1–11, 2017.
[32] G. Vivone, L. M. Millefiori, P. Braca, and P. Willett, "Performance Assessment of Vessel Dynamic Models for Long-Term Prediction Using Heterogeneous Data," *IEEE Trans. Geosci. Remote Sens.*, vol. PP, no. 99, pp. 1–14, 2017.
[33] "HF Surface Wave Radar." [Online]. Available: http://www.diginext.fr/en/offer/critical-operation-support-systems/hf-surface-wave-radar. [Accessed: 16-Jun-2017].
[34] "Automatic Identification Systems (AIS)." [Online]. Available: http://www.imo.org/en/OurWork/safety/navigation/pages/ais.aspx. [Accessed: 04-Jun-2017].
[35] H.-P. Kriegel, P. Kröger, J. Sander, and A. Zimek, "Density-based clustering," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 1, no. 3, pp. 231–240, May 2011.
[36] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, and others, "A density-based algorithm for discovering clusters in large spatial databases with noise.," in *Kdd*, 1996, vol. 96, pp. 226–231.
[37] W. Stuetzle, "Estimating the Cluster Tree of a Density by Analyzing the Minimal Spanning Tree of a Sample," *J. Classif.*, vol. 20, no. 1, pp. 025–047, May 2003.
[38] E. Protopapadakis and A. Doulamis, "Semi-Supervised Image Meta-Filtering Using Relevance Feedback in Cultural Heritage Applications," *Int. J. Herit. Digit. Era*, vol. 3, no. 4, pp. 613–627, Dec. 2014.
[39] M. Hein, J.-Y. Audibert, and U. von Luxburg, "From Graphs to Manifolds – Weak and Strong Pointwise Consistency of Graph Laplacians," in *Learning Theory*, P. Auer and R. Meir, Eds. Springer Berlin Heidelberg, 2005, pp. 470–485.
[40] F. Xiao and C. Fan, "Data mining in building automation system for improving building operational performance," *Energy Build.*, vol. 75, pp. 109–118, Jun. 2014.
[41] T. Caliński and J. Harabasz, "A dendrite method for cluster analysis," *Commun. Stat.*, vol. 3, no. 1, pp. 1–27, Jan. 1974.
[42] D. L. Davies and D. W. Bouldin, "A Cluster Separation Measure," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-1, no. 2, pp. 224–227, Apr. 1979.
[43] E. Protopapadakis *et al.*, "A genetically optimized neural classifier applied to numerical pile integrity tests considering concrete piles," *Comput. Struct.*, vol. 162, pp. 68–79, Jan. 2016.

[44] D. W. Winters, B. D. V. Veen, and S. C. Hagness, "A Sparsity Regularization Approach to the Electromagnetic Inverse Scattering Problem," *IEEE Trans. Antennas Propag.*, vol. 58, no. 1, pp. 145–154, Jan. 2010.