A supervised learning approach for rainfall detection from underwater noise analysis

A. Trucco[1], R. Bozzano[2], E. Fava[1], S. Pensieri[2], A. Verri[3], and A. Barla[3]

[1] Dept. of Electrical, Electronic, Telecommunications Engineering, and Naval Architecture (DITEN), University of Genoa, Italy
[2] National Research Council of Italy, Institute for the Study of Anthropic Impact and Sustainability in the Marine Environment (CNR-IAS),
[3] Dept. of Informatics, Bioengineering, Robotics, and System Engineering (DIBRIS), University of Genoa, Italy

Abstract – Underwater noise analysis allows estimation of parameters of meteorological interest, difficult to monitor with in-situ devices, especially in very harsh environments such as polar waters. Rainfall detection is a fundamental step of acoustical meteorology toward quantifying precipitation and, indirectly, wind. To date, this task has been conducted with some success by using a few frequency bins of the noise spectrum, and combining their absolute values and slopes into some inequalities. Unfortunately, these algorithms do not perform well when applied to spectra obtained by averaging multiple noise recordings made over the course of an hour. Supervised, machine learning models allow the use of all the frequency bins in the spectrum, exploiting relationships that are difficult for a human observer to identify. Among the different models tested, a binary classifier based on random forest performed well with moderate computational load. Using a data set consisting of over 18,000 hourly-averaged spectra (approximately 25 months of in-situ recordings) and comparing the results with measurements from a surface-mounted rain gauge, the proposed system detects precipitations greater than 1 mm/h with 90% probability, keeping the false alarm probability below 0.5%. This system has demonstrated remarkable robustness as performance is achieved without excluding spectra corrupted by sounds produced by other sources, such as naval traffic and wind blowing over the sea surface.

Keywords – underwater acoustics, acoustical meteorology, rainfall detection, noise analysis, supervised learning, machine learning.

I. INTRODUCTION

The measurement of rain and wind in the marine environment is an essential operation in understanding and monitoring natural phenomena, especially in relation to climate change and risk prevention [1-4]. Satellites for meteorological observation provide a valuable contribution, although the spatial and temporal resolutions they provide do not always meet monitoring requirements. This problem is particularly felt in the polar environment due to the reduced coverage that these satellites offer at higher latitudes [3,5]. Weather surveillance radars, operating along the coast, and surface rain gauges and anemometers, installed on oceanographic fixed or mobile platforms, also present critical issues that make it difficult to deploy these devices on a large scale [3,6,7]. For these reasons, estimating wind speed and rainfall intensity using underwater acoustic noise is considered a crucial technique for a better understanding of the oceans, either as an alternative to or in support of satellites, coastal radar systems, and meteorological buoy networks [6,8].

In recent years, acoustical meteorology has received considerable attention, demonstrating that wind and rain can be measured with satisfactory accuracy using low-cost underwater devices installed on fixed moorings or moving platforms, in a variety of ocean environments [1,3,7,9-17]. However, several problematic issues are still present and need convincing answers to achieve proper operation of these devices in the field [2,18,19]. These issues include the possibility of performing wind and rain estimates when only acoustic data averaged over a significant period of time are available. Indeed, deploying a network of measurement devices in polar waters that operate completely autonomously, continuously operating for one or more years, could force a dramatic reduction in processing and storage resources. In mobile platform installations, it may be necessary to minimize the transmission resources to be committed during the rare and brief periods of surfacing. These savings requirements led Vagle, Large and Farmer, in their pioneering work [9], to propose a method for estimating wind speed using the average of the acoustic data acquired, at various times, over a period of one hour. Recently, a similar proposal has also been formulated and tested for rain monitoring [20].

In the literature concerning the acoustic measurement of rain, the intensity of precipitation is estimated through two distinct steps [3,7,10-13,15,18,20,21,22]: the detection of rainfall and, if any, the estimation of its intensity. With the exception of [20], for both operations the input data are derived from acoustic signals gathered over a few seconds or, at most, a few minutes. This paper investigates the possibility of detecting precipitation over a one-hour period by exploiting only the average of consecutive acoustic spectra acquired at intervals of a few minutes during that hour, without performing any processing to filter out measurements potentially affected by noise sources other than rain. The detection also aims to reveal intermittent rain falling for a period shorter than the hour under examination. Furthermore, the adopted dataset consists of spectra acquired during different deployments of the acoustic sensor, seasons, and environmental conditions, covering about 25 months of operation, using the same platform. Over such an extended time interval, while precipitation detection is an essential step in quantifying rainfall, it can also be useful in estimating wind speed, due to the combined effect of these phenomena on underwater noise, and in monitoring of oceanographic parameters, such as sea surface salinity.

The methods mentioned above [1,7,9,10,13,14,21] that aim to perform rainfall detection using short-term acoustic data (short-term being the term adopted to indicate data gathered over some seconds or a few minutes) do not provide satisfactory performance when hourly-averaged acoustic data are used as input. In addition, the detection is performed by decision rules that exploit only the values and slopes of the acoustic spectrum at a few predetermined frequencies (for this reason, these methods will be referred to as rule-based). To overcome this restriction, a recent paper [20] proposed machine learning methods to estimate rainfall intensity and wind speed using all the frequency bins of the underwater noise spectrum as input data, in order to exploit implicit relationships that are not evident to the human observer. In [20] machine learning methods are also applied for rainfall detection, using hourly-averaged spectra as input data. Unfortunately, detection is limited to precipitation intensities greater than 1 mm/h and the performance obtained over a one-year period is worse than that reported in [15], where a rule-based estimation method [21] was fed with short-term data collected by the same equipment and over the same time period used in [20].

This paper is aimed at improving the performance of rule-based methods in rainfall detection, exploiting all the frequency bins of the spectrum within a scheme based on supervised machine learning models, already successfully applied to other detection problems of the underwater acoustic domain [23-25]. The new knowledge that this work introduces is twofold: the demonstration that hourly-averaged spectra can be used to detect rainfall with better performance than that achieved by rule-based methods fed by short-term data; and, secondly, an in-depth analysis of the potential and limitations of the machine learning models adopted, made possible by experimentation on real data collected at sea over a period of more than two years. In addition, the detection scheme proposed here is capable of operating even with extremely light precipitation, being able to detect rainfall intensity of 0.1 mm/h, a value that represents the instrumental limit of most commercial rain gauges.

This paper is organized as follows. Section II presents a brief state-of-the-art in rainfall detection from underwater noise. Section III describes the experimental set-up and data used in this study, the detection algorithms from the prior literature, and the proposed approach, based on machine learning models. Section IV reports and compares the detection results obtained from rule-based methods, the proposed approach, and a weather radar system operating simultaneously in the area of interest. Finally, some conclusions are drawn in Section V.

II. STATE-OF-THE-ART OVERVIEW

The methods proposed in past decades for rainfall detection through underwater noise analysis are based on spectral values and slopes at given frequencies, compared among them or against fixed thresholds. In [9], the spectral slopes between 3 and 8 kHz and between 3 and 19.5 kHz are compared to specific thresholds to achieve an indication of the precipitation presence. In [1], rainfall is classified in several categories depending on the difference between the average spectral levels in two bands: from 4 to 10 kHz and from 10 to 30 kHz. In [10], rainfall is detected when a set of inequalities, in which the spectral levels at 5, 8 and 25 kHz are linearly combined and compared against specific thresholds, are satisfied. Similarly, in [7], the spectral levels at 5.4, 8.3 and 21 kHz are adopted with different coefficients and thresholds. In addition, spectra corrupted by transient noise or by high wind are discarded, and a continuity check is applied to reduce the false detection rate: if no new rainfall detections occur within 10 minutes of the first detection, then such a detection is assumed to be false. At a later time, one of the authors of [7] proposed a new version of the detection algorithm [21], in which a higher number of inequalities combine the spectral levels at 5, 8, and 20 kHz (as is and squared) and the slopes between 2 and 8 kHz and between 8 and 15 kHz. In [13] this algorithm is further refined by updating a couple of threshold levels. Finally, in [14], a detection scheme is introduced in which the minimum and maximum spectral levels between 10 and 20 kHz are exploited. All these algorithms are set through the authors' observations of acoustic spectra collected in rainy and non-rainy conditions. Moreover, they are designed to process input data derived from short-term acoustic signals.

A statistical assessment of the detection results is provided only in [3,7,15], whereas the other papers cited above evaluate the proposed algorithms only on a few selected cases. In [7], thanks to the removal of noisy samples and the introduction of a continuity check, the probability of false alarm (P_{fa}, i.e., the probability of detecting rain in the absence of precipitation) is 0.004. The probability of detection (P_d, i.e., the probability of detecting rain when precipitation occurs) is 0.6 for a rainfall intensity greater than 5 mm/h and 0.8 for a rainfall intensity greater than 10 mm/h. In [15] the authors applied the detection algorithm described in [21], including noisy sample removal and a continuity check, obtaining P_{fa} = 0.0052 and P_d = 0.584 for a rainfall intensity greater than 0.1 mm/h. P_d increases to 0.839 when only samples with rainfall intensity greater than 1 mm/h are considered. Finally, in [3] the authors reported a P_d = 0.7 for a rainfall rate greater than 3 mm/h using the acoustic device described in [13].

The machine learning approach proposed in [20] for rainfall monitoring applies supervised models to hourly-averaged acoustic spectra, extending the analysis to all the frequency bins instead of only a few frequencies and slopes. For the detection task, a binary classifier is built through the CatBoost algorithm, setting the lower bound for rainfall intensity equal to 1 mm/h. When the detector is applied to the available one-year dataset (through the cross-validation scheme), a P_{fa} = 0.0332 and a P_d = 0.811 are obtained: a poorer performance than that obtained in [15] using the same dataset but exploiting short-term data in place of hourly-averaged data.

III. MATERIAL AND METHODS

A. Experimental measurements

The acoustic underwater noise and the rainfall intensity at sea surface were collected from 17 June 2011 to 6 September 2013 (with a few breaks, approximately 1.5 months overall) by apposite sensors installed on the meteo-oceanographic observatory W1M3A, moored on a deep-sea bed of 1,200 m, about 80 km off the Ligurian coast, in the northwestern part of the Mediterranean Sea, as detailed in [15,26,27].

The rainfall intensity was measured with a Vaisala RAINCAP Sensor, comprised in a Vaisala Weather Transmitter WXT520, placed on the upper part of the buoy trellis, at about 10 m above sea level. Precipitation measurements were acquired at high temporal resolution (5 seconds) and contribute to the measurements of the hourly cumulative rainfall intensity [15]. During the hours when the cumulative precipitation was acquired, the hourly average wind speed was also computed using measurements from a WindSonic 2D anemometer installed on the same trellis on the observatory at 10 m above sea level.

The underwater acoustic noise was acquired by a dedicated oceanic recorder, based on Passive Aquatic Listener (PAL) technology [13,28,29], clamped to the body of the platform at a depth of 36 m. This device is designed to operate unattended at sea for a long period of time powered by an internal battery, and to acquire an average of seven acoustic noise snapshots per hour. Each snapshot consists of a time series of 4.5 s, sampled at 100 kHz, which is processed on board to obtain a spectrum composed of 64 frequency bins, with a resolution of 0.2 kHz from 0.1 to 3 kHz and 1 kHz from 3 to 50 kHz. The spectra of the snapshots acquired in one hour (at an average interval of about 9 minutes from each other) were averaged, producing a mean spectrum that is included in the acoustic dataset used in this paper for rainfall detection.

In the entire period of operation, 18,193 hourly-averaged acoustic spectra were collected and are available for processing, amounting to about two and a half times those considered in [20]. The concurrent measurements of the hourly rainfall intensity are also available and are assumed, in this study, as the ground truth. The rain gauge measured a precipitation greater than 0.1 mm/h in 876 of the 18,193 hours considered. The maximum rainfall intensity measured was 51.5 mm/h, and the distribution of the observed intensities is shown in Fig. 1. The average wind speed ranged between 0.4 and 20.7 m/s with the distribution shown in Fig. 2. Finally, the tracks of the Automatic Identification System (AIS) used on ships reveal how many of them transited near the buoy in the period of data acquisition. Considering a circle with a radius of 5 km, centered at the position of the buoy, the number of hours in which at least one ship crossed the circle is 1,999, of which 78 are characterized by the presence of rain.

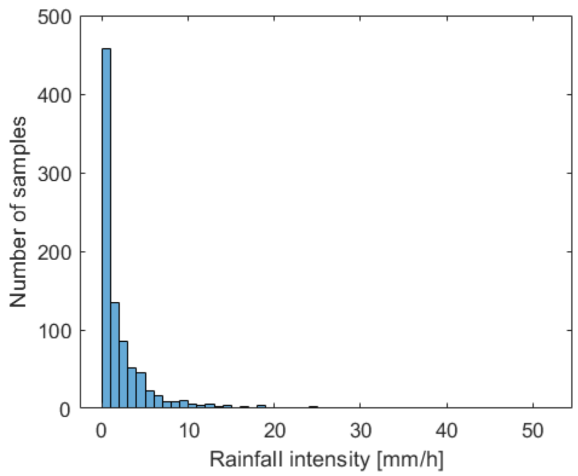


Fig. 1. Rainfall intensity distribution in the dataset samples.

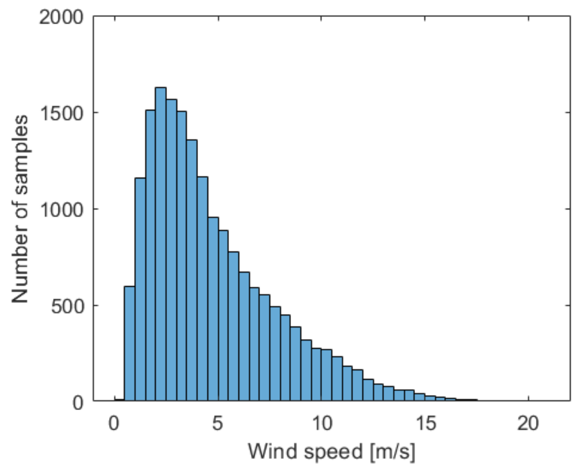


Fig. 2. Wind speed distribution in the dataset samples.

The W1M3A observation system is permanently moored in the operation area of the weather radar on Mount Settepani, located at about 1,400 m above sea level, about 87 km away from the buoy, jointly operated by the Regional Agency for environmental protection of Piedmont and Liguria, as detailed in [15,29]. Assuming the measurements provided by the rain gauge deployed on the W1M3A observatory as ground truth, the performance of this radar system in detecting rainfall at the buoy position was reported in [15], limited to the period from 17 June 2011 to 9 May 2012. This figure is used in this paper as a further comparison for the proposed underwater detection system.

B. Rule-based algorithms for rainfall detection

The notation $S(f_k)$ is introduced to indicate the sound spectral level of underwater noise, measured in dB re 1 μPa^2 Hz^{-1}, at the frequency f_k expressed in kHz.

According to [9], rainfall is present when:

$$S(19.5) - S(3) > -13.25 \quad \mathrm{OR} \quad S(8) - S(3) > -6.82 \tag{1}$$

In [7] rainfall is detected if at least one of the following three conditions is verified, the third condition being specific for drizzle:

$$S(21) + 2.35\, S(5.4) > 194 \tag{2}$$

$$S(21) > 48 \quad \mathrm{AND} \quad S(5.4) > 53 \tag{3}$$

$$S(21) > 44 \quad \mathrm{AND} \quad S(21) - 0.7\, S(8.3) > 14 \tag{4}$$

In addition, the removal of spectra corrupted by noise and the temporal continuity check are applied [7], as described in Section II. These operations lose their meaning when this algorithm is applied to hourly-averaged spectra.

In [13] and [21] rainfall is detected if at least one of the following four conditions is verified, the third condition being specific for drizzle and the fourth for rain with high wind:

$$S(20) - 0.75\, S(5) > 5 \quad AND \quad S(5) \leq 70 \tag{5}$$

$$S(8) > 60 \quad AND \quad Q(2{,}8) > \theta \quad AND \quad S(20) > 45 \tag{6}$$

$$S(8) < 50 \; AND \;\; Q(8{,}15) > -5 \;\; AND \;\; S(20) > 35 \;\; AND \;\; S(20) > 0.9\, S(8) \tag{7}$$

$$\begin{cases} S(20) + 0.1144\, S^2(8) - 12.728\, S(8) > -307 \quad AND \quad Q(2{,}8) > \theta \\ AND \;\; S(20) + 0.1\, S^2(8) - 11.5\, S(8) < -281 \quad AND \quad 51 < S(8) < 64 \end{cases} \tag{8}$$

where $Q(f_1, f_2)$ is the spectral slope, in dB/decade, between the frequencies f_1 and f_2 (expressed in kHz):

$$Q(f_1, f_2) = \frac{S(f_1) - S(f_2)}{\log_{10}(f_1) - \log_{10}(f_2)} \tag{9}$$

The difference between the algorithms in [13] and [21] is only the value assigned to the constant θ: θ = –18 dB/decade in [21] and θ = –13 dB/decade in [13].

The algorithms proposed in [1], [10] and [14] are not included in this collection because: [1] is mainly dedicated to the classification of rain, downstream of a detection carried out by other means; [10] represents a preliminary version of the algorithm proposed in [7]; [14] presents a general idea based on the maximum spectral slope observed between 10 and 20 kHz, with no specific detection algorithm.

C. Rainfall detection by supervised learning models

An alternative to the rule-based detection algorithms described above is to exploit all the information available, looking for relationships between the spectrum frequency bins (none excluded) and the rainfall presence through machine learning models driven by experimental observations. A supervised binary classifier (the classes of which are rainy and non-rainy) that receives as input the frequency bins of an hourly-averaged spectrum can successfully perform precipitation detection. In this study, standard supervised learning methods well suited to address binary classification in the presence of a large number of features are adopted: two linear classification techniques (Linear Discriminant Analysis and Logistic Regression), a kernel-based method (Support Vector Machine) and an ensemble learning method (Random Forest). Given the type and size of the available data, the spectrum of the methods adopted is considered sufficient to assess the advantage of employing a data driven approach to effectively detect rainfall.

1. Notation for data and performance

To discuss the characteristics of the models mentioned, the following notation will be used. A sample is an hourly-averaged spectrum and is indicated by the vector $\boldsymbol{x}$, $\boldsymbol{x} \in \mathbb{R}^d$. The vector is composed of d frequency bins, called features. In this study d is equal to 64 and the dataset contains 18,193 samples, a fraction of which are used for the training of the statistical models. The training set is indicated by $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^{L}$, where L is the number of samples used for the training phase; $\boldsymbol{x}_i$ is the ith training sample, associated with either rainy or non-rainy classes; and y_i is equal to ± 1 depending on membership of $\boldsymbol{x}_i$ to one of the two classes: $+1$ for the rainy case and -1 for the non-rainy case. The samples not belonging to the training set constitute the test set. The application of the trained model to a sample $\boldsymbol{x}$ taken from the test set allows us to assign such a sample to the $+1$ or -1 class. Since the actual rain condition is also known for the samples of the test set, P_d, P_{fa}, the overall accuracy (OA, i.e., the probability of correct classification) and the receiver operating characteristic (ROC) curve can easily be estimated. The ROC curve shows the possible tradeoffs between P_d and P_{fa} and can be traced by varying the threshold used to decide membership of the test sample $\boldsymbol{x}$ on the basis of the real-valued score produced by the trained model, when the model is applied to $\boldsymbol{x}$. The area under the ROC curve (AUC) is commonly used to quantitatively evaluate the detector positioning between the detector choosing at random (AUC = 0.5) and the ideal detector (AUC = 1.0).

2. Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) is a traditional method [30], based on decision theory and Bayes theorem, in which the probability density functions for the samples belonging to the $+1$ and -1 classes are assumed to be multivariate Gaussian with mean vectors $\boldsymbol{\mu}_{+1}$ and $\boldsymbol{\mu}_{-1}$, respectively, and the same covariance matrix Σ. The knowledge of these class-conditional densities, $f_{x|+1}(\boldsymbol{x})$ and $f_{x|-1}(\boldsymbol{x})$, together with the prior probabilities for the two classes, $P(-1)$ and $P(+1)$, makes computation of the class posterior probabilities possible for a given sample $\boldsymbol{x}$, $P(+1 \mid \boldsymbol{x})$ and $P(-1 \mid \boldsymbol{x})$. Specifically, the probability for the $+1$ class given sample $\boldsymbol{x}$ is:

$$P(+1 \mid \boldsymbol{x}) = \frac{f_{x|+1}(\boldsymbol{x})P(+1)}{f_{x|+1}(\boldsymbol{x})P(+1) + f_{x|-1}(\boldsymbol{x})P(-1)} \tag{10}$$

Sample $\boldsymbol{x}$ is assigned to the $+1$ class if this probability exceeds 0.5, to the -1 class otherwise. Although the 0.5-threshold is optimum in terms of overall classification accuracy, a different value can be set to modify the balance between P_d and P_{fa}, and, therefore, the tuning of the threshold allows the tracing of the detector's ROC curve. The samples belonging to the training set are used in LDA to estimate the mean vectors, the covariance matrix and the prior probabilities mentioned above. LDA is a linear classification method because membership of sample $\boldsymbol{x}$ can be equivalently assigned working on the log-odds function (i.e., the logarithm of the ratio between $P(+1 \mid \boldsymbol{x})$ and $P(-1 \mid \boldsymbol{x})$), which is a linear equation in $\boldsymbol{x}$.

3. Logistic Regression

The Logistic Regression (LR) model assumes the log-odds function to be a linear function in $\boldsymbol{x}$ and derives the equations for the class posterior probabilities without introducing any assumption about the class-conditional density functions [30]. In the binary case, the probability for the $+1$ class given the sample $\boldsymbol{x}$ results:

$$P(+1 \mid \boldsymbol{x}) = \frac{1}{1 + \exp(\beta_0 + \boldsymbol{\beta}^T \boldsymbol{x})} \tag{11}$$

i.e., a sigmoid function whose parameters β_0 and $\boldsymbol{\beta}$ can computed by maximizing a conditional log-likelihood function. The maximization is achieved through an iterative procedure in which the training set samples are exploited and the Newton-Raphson algorithm is typically applied to find the root of the first derivative [30]. As in LDA, sample $\boldsymbol{x}$ is assigned to the $+1$ class if $P(+1 \mid \boldsymbol{x})$ is greater than 0.5, to the -1 class otherwise, but different threshold values can be used to trace the detector's ROC curve and change the balance between P_d and P_{fa}.

4. Support Vector Machine

A Support Vector Machine (SVM) assigns sample $\boldsymbol{x}$ to one of the two classes based on the score of the discriminant function:

$$h(\boldsymbol{x}) = \sum_{i=1}^{L} \alpha_i y_i K(\boldsymbol{x}_i, \boldsymbol{x}) + b \tag{12}$$

where $K(\cdot,\cdot)$ is a kernel function and the coefficients α_i and b are optimized by solving a quadratic programming problem [30]. This optimization problem exploits the samples of the training set and includes a parameter C that bounds the range for α_i: $0 < \alpha_i < C,\ i = 1,2,\dots,L$. Sample $\boldsymbol{x}$ is assigned to the $+1$ class if $h(\boldsymbol{x})$ is positive, to the -1 class otherwise. As for previous methods, different threshold values can be used to trace the detector's ROC curve and change the balance between P_d and P_{fa}. In the SVM literature the most commonly adopted kernels are the linear function, polynomial function of order q and Gaussian radial basis function (RBF), defined, respectively, as:

$$K(\boldsymbol{x}_i, \boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{x}_{\boldsymbol{i}} \tag{13}$$

$$K(\boldsymbol{x}_i, \boldsymbol{x}) = (1 + \boldsymbol{x}^T \boldsymbol{x}_{\boldsymbol{i}})^q \tag{14}$$

$$K(\boldsymbol{x}_i, \boldsymbol{x}) = \exp\left({-\|\boldsymbol{x} - \boldsymbol{x}_i\|^2}/{2\sigma^2} \right) \tag{15}$$

where σ^2 is a specific parameter of the RBF kernel.

The choice of C and σ^2 (if the case) requires specific attention and, possibly, an optimization stage [30]. In addition, although not strictly necessary, all features of the dataset samples are often preliminarily standardized, so that each of them has a zero mean and a unitary variance. This operation makes features insensitive to the scales on which they are measured and favors numerical stability in the solution of the quadratic programming problem mentioned above.

5. Random Forest

Random Forest (RF) is an ensemble model that aggregates the predictions individually achieved by many decision trees, separately trained on a subset of samples randomly chosen from the training set [30]. A decision tree is an acyclic connected graph where each node represents a decision rule (called split) related to a single feature that leads to the partition of data in two groups. To automatically set the structure and splits of a decision tree, Classification and Regression Trees (CART) is a widely adopted algorithm in which a new node is created by identifying the feature that yields the best split in terms of a pre-selected metric. In an RF model B trees are generated and trained in an independent and identically distributed way by performing, for each tree $\mathcal{T}_b$, b = 1, …, B, the following steps [30]: (a) a subset of L samples is drawn randomly from the training set, uniformly and with replacement (this means that some samples are taken more than once, others are not chosen at all); (b) such a subset is used to grow the tree $\mathcal{T}_b$, for each node of which a pool of m features is selected (at random and uniformly from the d features) and used to identify the best feature and the best decision rule to split the node into two daughter nodes; (c) the previous step is repeated until at least one of the predefined stopping criteria is satisfied. When all the B trees are generated, an unknown sample $\boldsymbol{x}$ is classified as follows: the sequence of decision rules of the bth tree is applied to $\boldsymbol{x}$ in such a way that the corresponding class prediction $\hat{y}^b(\boldsymbol{x})$ is reached (namely, $+1$ or -1); the predictions from all the trees of the RF are used to compute a score

$$g(\boldsymbol{x}) = \frac{1}{B}\sum_{b=1}^{B} \hat{y}^b(\boldsymbol{x})\,; \tag{16}$$

sample $\boldsymbol{x}$ is assigned to the $+1$ class if $g(\boldsymbol{x})$ is positive, to the -1 class otherwise. Threshold values different from zero can be used to trace the detector's ROC curve and tune the balance between P_d and P_{fa}. Although the setting of B and m does not critically affect performance, it deserves some investigation, recalling that these two parameters affect the computational burden.

6. *Cross-validation*

The assessment of a trained statistical model performance is a crucial task for which K-fold cross-validation represents an easy and extensively applied option [30]. To exploit the available data for both training and testing a machine learning model, the dataset is split in K subsets (called folds), non-overlapped and of approximately equal-size. Taking the kth subset aside, the model is trained using the other K–1 subsets of data, and the test is performed using the data of the kth subset. This operation is repeated for k ranging from 1 to K, in such a way that every sample is used, in turn, to train and test the supervised model. By combining the predictions performed at each step k on the data subset kept aside, a prediction for each sample of the entire dataset is finally available. Because the prediction consists of the probability (or score) for membership of the sample of a given class, after setting a threshold value, estimation of OA, P_d, and P_{fa} is possible. In addition, the tuning of such a threshold allows the generation of the cross-validated ROC curve [30].

To cope with the different cardinality of the two classes in the dataset, the dataset partition in K subsets can be performed by a stratified scheme according to which each subset maintains approximately the same class proportions as the original dataset.

IV. RESULTS AND DISCUSSION

To delineate the desired performance of the rain detector, it is necessary to recall that rainfall is present in 5% of the one-hour periods included in the dataset and that the precipitation limit which distinguishes between rainy and non-rainy hourly-averaged spectra is particularly low (i.e., 0.1 mm/h). In this scenario, it is strictly necessary that the false alarm probability be very low, while a detection probability not too close to one may be acceptable. Consequently, the performance of a detector cannot be considered acceptable if P_{fa} exceeds 0.01.

A. Performance of rule-based algorithms

The application of the algorithms introduced in Section III.B to the dataset described in Section III.A provides the results summarized in Table I. It is important to recall that these algorithms were designed to detect rainfall using short-term acoustic spectra, whereas in this study they are applied to hourly-averaged spectra.

Table I.
Probabilities of detection, P_d, and false alarm, P_{fa}, for the rule-based algorithms applied to hourly-averaged spectra.

Algorithm	P_d	P_{fa}
Vagle *et al.*, 1990 [9]	0.880	0.570
Ma *et al.*, 2005 [7]	0.300	0.001
Nystuen, 2011 [21]	0.671	0.116
Nystuen *et al.*, 2015 [13]	0.586	0.094

The algorithm in [9] achieves high P_d, but this is accompanied by excessive P_{fa}. A bias in hydrophone sensitivity cannot be the cause of the problem, because the quantities compared with thresholds in (1) are subtractions between measurements. One option to make the algorithm more selective is to arbitrarily increase the two threshold values, as follows:

$$S(19.5) - S(3) > -13.25 + \delta \quad \text{OR} \quad S(8) - S(3) > -6.82 + \delta \tag{17}$$

where $\delta > 0$. Varying the value of δ between 0 and 5, the ROC curve in Fig. 3 is obtained. The P_{fa} reduction is obtained but, unfortunately, it is accompanied by a significant P_d decrease.

What is more, the algorithms in [7], [13] and [21] do not provide satisfactory performance, because P_d is too low, as for [7], or P_{fa} is too high, as for [13] and [21]. As discussed in [12], the performance of these algorithms can be optimized by considering potential errors in hydrophone sensitivity. To do this, the values $S(f_\mathrm{k})$ in equations from (2) to (8), for whatever f_k, are replaced by $S(f_\mathrm{k}) + \varepsilon$, where ε is intended to compensate a sensitivity bias. Varying ε between –10 and 10 dB re 1 $\mu Pa^2\ Hz^{-1}$, the ROC curves shown in Fig. 3 are obtained. This comparison clearly evinces that the rule-based algorithm achieving the best performance (with the discussed correction) is the one proposed in [7]. In particular, for ε = 2 dB re 1 $\mu Pa^2\ Hz^{-1}$, a detection probability $P_d = 0.521$ is accompanied by $P_{fa} = 0.010$.

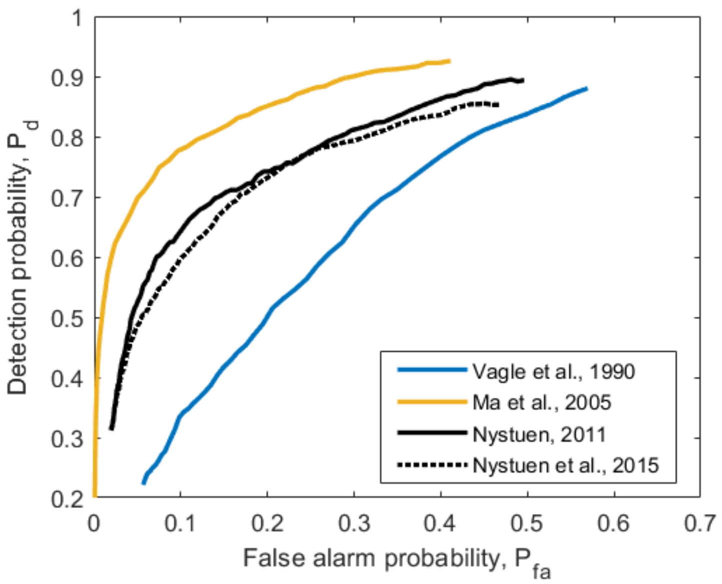


Fig. 3. ROC curves obtained by varying threshold values [9] and hydrophone sensitivity [7,13,21] in the rule-based algorithms listed in Table I.

B. Performance of the supervised learning models

To assess and compare the detection performance of the statistical models, 10-fold cross-validation with stratification in dataset partitioning is adopted. In addition, for the SVM approach feature standardization is applied, the constant C is set equal to 1.0, according to common practice, and the variance σ^2 of the Gaussian RBF kernel, after some tests, is set equal to 8.0. For the parameters of the RF, B = 100 trees and m = 22 features are used, although a change of these values in even rather broad ranges does not significantly affect the performance obtained.

The ROC curves obtained from the trained models are shown in Fig. 4 and their performance is summarized in Table II, where the P_d value for which P_{fa} = 0.01 is reported. More precisely, Table II shows the averages of the probabilities P_d and P_{fa} computed for the 10 folds in which the dataset is divided, together with the relative standard deviations.

The linear classifiers (i.e., LDA and LR) perform moderately better than the best rule-based algorithm, increasing the probability of detection to about 0.6. A further advantage is offered by the SVM and RF classifiers for which probability of detection exceeds 0.7. The change of the kernel function for the SVM classifier does not significantly alter the performance, although the linear case shows a lower detection ability and the Gaussian case reports the worst AUC figure. The OA values are all greater than 0.97, but this finding has little relevance because it is strongly influenced by the correct classification of non-rainy samples (probability 0.99, P_{fa} being 0.01) which are by far the most numerous. Overall, the best option among the models considered is the RF classifier because it achieves the best performance figures, shows a stability better than that of SVM classifiers with polynomial or RBF kernels, requires a computational load lower than that of such SVM classifiers, and is not appreciably affected by changes in the parameter setting. Accordingly, in the remainder of this section, further analysis and performance comparisons will be carried out with reference to the RF-based classifier. Indeed, the goal is not to identify the best model, but rather to demonstrate that the machine learning approach is well suited for rainfall detection also in case of drizzle phenomena, characterized by low rainfall intensity.

A portable computer with an Intel Core i7 CPU of 1.9 GHz and 16 GB of RAM memory trains the RF classifier, using a Matlab routine and the entire dataset, in about 20 seconds. The execution of the detection task on the more than 18,000 samples of the dataset requires less than 2 seconds.

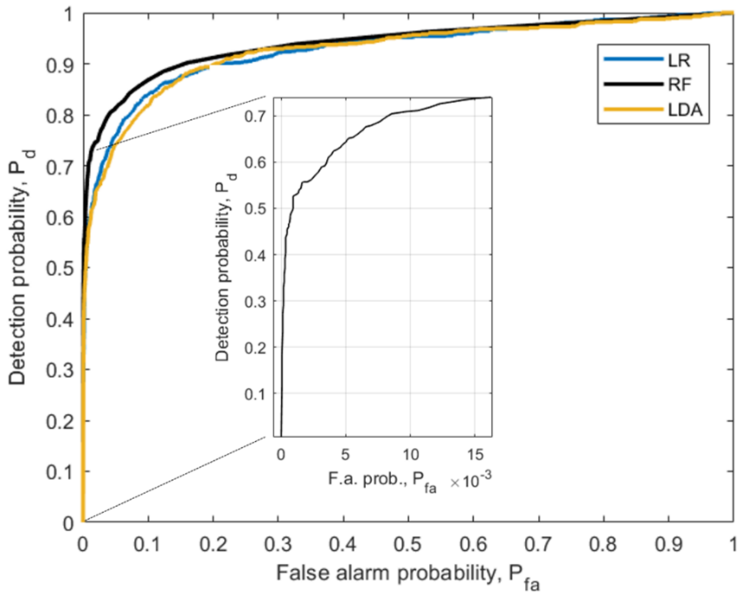
Table II.
Detection probability, P_d, false alarm probability, P_{fa}, overall accuracy, OA, and area under the ROC curve, AUC, for the supervised learning models. For the probabilities, the average ± the standard deviation is reported.

Classifier	P_d	P_{fa}	OA	AUC
LDA	0.583 ± 0.053	0.010 ± 0.003	0.970	0.926
LR	0.597 ± 0.058	0.010 ± 0.003	0.971	0.928
SVM, linear	0.667 ± 0.041	0.010 ± 0.003	0.974	0.931
SVM, polynomial, q=2	0.702 ± 0.060	0.010 ± 0.002	0.976	0.936
SVM, Gaussian RBF	0.703 ± 0.062	0.010 ± 0.003	0.976	0.897
RF	0.708 ± 0.054	0.010 ± 0.002	0.977	0.941

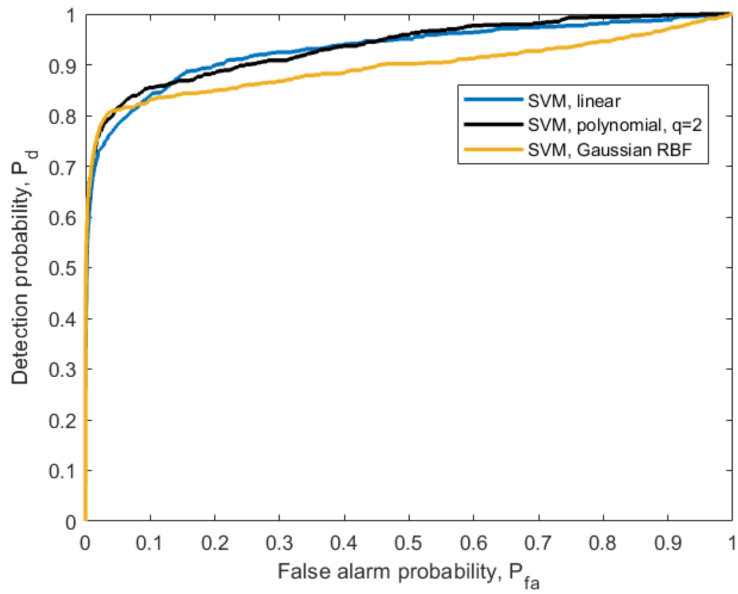
C. In-depth analysis and comparisons

In Fig. 4(a), the zoom of the ROC curve for the RF model demonstrates that P_d remains greater than 0.6 even if P_{fa} is reduced by as much as 0.0035. More precisely, the following probability pairs, {P_d, P_{fa}}, lie on that curve: {0.661, 0.006}, {0.644, 0.005}, {0.623, 0.004}, {0.588, 0.003}.

The ability of the classifier to detect the precipitation can be analyzed as a function of the rainfall rate [7,15], as shown in Fig. 5. In this case, P_d is estimated using the hourly samples in which the cumulated rainfall, measured by the rain gauge on the platform in one hour, is equal to or greater than a value G. The P_d curves shown in Fig. 5 are related to three choices of the threshold value, leading to different P_{fa}: 0.010, 0.005 and 0.003. P_d increases rapidly with G, reaching, respectively, 0.921, 0.897 and 0.876 for G = 1 mm/h. Although the probabilities of detection of the three detectors show significant differences for G < 2 mm/h, for rainfall intensities higher than this value the three detectors provide similar P_d. It is therefore possible to design acoustic detectors capable of detecting rainfall of intensity greater than 2 mm/h with a probability greater than 0.9, while keeping a false alarm probability of 0.003.



(a)



(b)

Fig. 4. ROC curves for the supervised learning models listed in Table II. (a) LDA, LR and RF, with a zoom for the RF model. (b) SVM with three kernel functions.

The sharp P_d increase with G observed in Fig. 5 shows that the missed detections are mainly related to drizzle phenomena characterized by low precipitation intensity. This relation is confirmed by the average of the rainfall intensities recorded by the surface rain gauge when the precipitation is detected or missed by the underwater acoustic device. Among the 876 acoustic samples collected in rainy conditions (with intensity greater than 0.1 mm/h), the RF-based classifier correctly detects

620 of them (70.8%) and misses the remaining 256 samples (29.2%). The average rainfall intensity measured for the detected samples is 2.98 mm/h, whereas the average intensity for the missed samples is 0.71 mm/h.

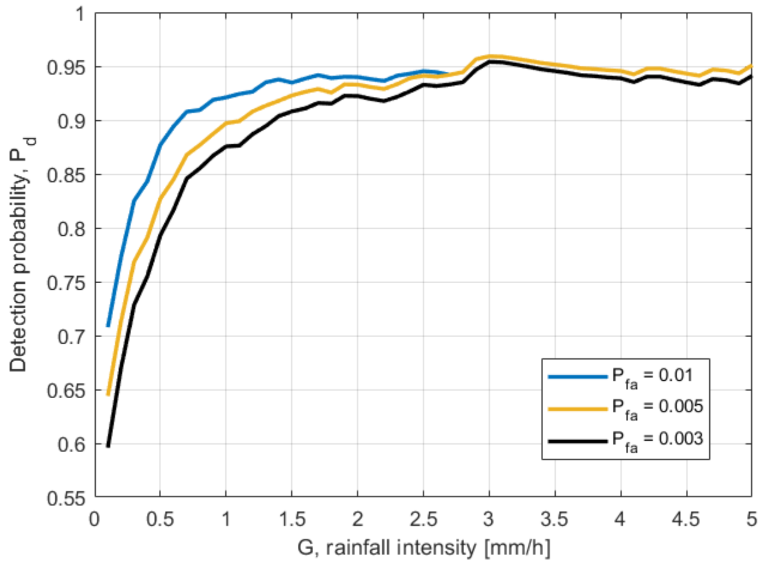


Fig. 5. Detection probability for the rainy samples with a rainfall intensity greater than or equal to G. Three RF-based classifiers, with different false alarm probabilities, are considered.

Moving from missed detections to false alarms, an analysis of wind distribution provides some interesting insights. Fig. 6 shows the wind speed histograms for rainy samples correctly detected (620 samples), non-rainy samples correctly classified (17,145 samples), and non-rainy samples raising false alarms (172 samples, corresponding to P_{fa} = 0.01). The average wind speeds for these three categories are, respectively, 8.5, 4.6 and 9.3 m/s. It is evident that the false alarm samples present a wind distribution more similar to that of the rainy samples than to that of non-rainy samples. However, the histogram of non-rainy samples shows that there are over a thousand samples with wind speed greater than 10 m/s that are correctly classified. To analyze this issue in detail, Fig. 7 shows the estimated P_{fa} when the samples for which the wind speed is greater than W, $W \in [0.1, 10]$ m/s, are considered. The three detectors already examined in Fig. 5 are included. Notwithstanding the considerable rise of the P_{fa} with increasing wind speed, the probability of correct classification for non-rainy samples remains satisfactory (e.g., for W = 10 m/s, P_{fa} increases from 0.01 to 0.08, but the probability of correctly classifying a non-rainfall sample is still high:

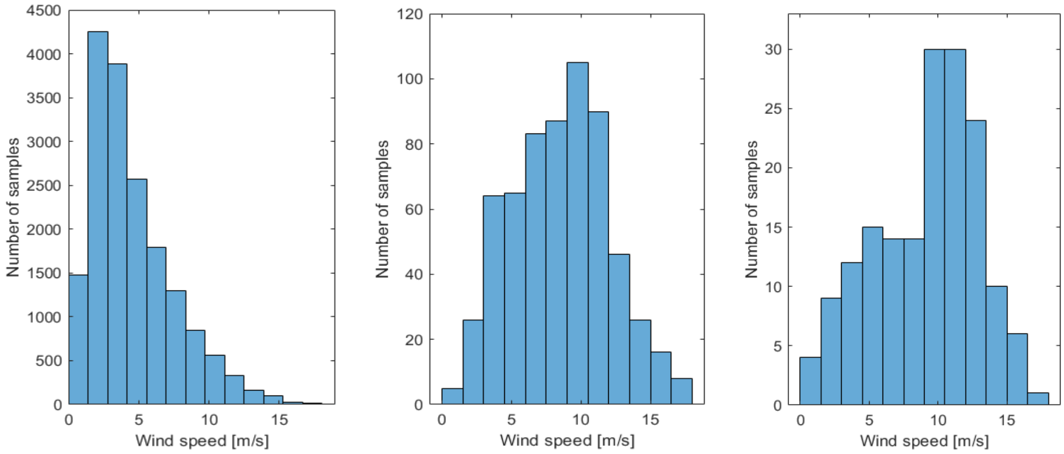
0.92). Therefore, the wind-related similarity only partially explains why the detector is misled and false alarms occur.

Finally, the performance of the RF-based detector during the period of data collection is examined in Fig. 8, where the height of the bars indicates the rainfall intensity measured by the rain gauge and the colors distinguish samples correctly detected (light blue bars) from missed alarm samples (orange bars). Samples raising false alarms are inserted as white bars with black edges, and an arbitrary height of 2 is set for them. The two zoom panels show the typical behavior that characterizes the 25-month span of data collection with good uniformity. As described in Section III.A, 1,999 out of the 18,193 dataset samples are characterized by the passage of a ship within 5 km of the platform during the observation hour. These samples are not discarded and are used, like all others, to train and test the statistical model. It is verified *a posteriori* that the P_d and P_{fa} values estimated on these samples do not differ significantly from those already reported, thus supporting the robustness of the proposed detector.

The performance achieved by the RF-based detector acting on hourly-averaged spectra can also be compared with those obtained by other underwater acoustic systems [3,7,15] acting on short-term spectra, summarized in Section II. By using data in Fig. 5, it is immediately possible to observe that the proposed system, at the same P_{fa} values and rainfall intensities, always provides a significantly higher detection capability. Moving from short-term spectra to hourly-averaged spectra, according to the data presented in Section IV.A, the performance obtained from the detection algorithms used in [3,7,15] worsen. As a result, the supervised learning models adopted in this study achieve a detection performance significantly better than those obtained from rule-based detection algorithms and better even than that obtained from the binary classifier proposed in [20].

Another useful comparison is with the rainfall detection capability of the weather radar described in Section III.A. According to [15], rainfall detection by radar at the W1M3A observatory is characterized by P_{fa} = 0.009 accompanied by P_d = 0.728 for G = 0.1 mm/h and P_d = 0.846 for G =

1 mm/h. The data in Fig. 5 show that the performance of the proposed acoustic system is very close to that of radar: slightly worse for $G = 0.1$ mm/h and slightly better for $G = 1$ mm/h.



(a) (b) (c)

Fig. 6. Histograms of wind speed for: (a) non-rainy samples correctly classified; (b) rainy samples correctly detected; (c) non-rainy samples raising false alarms.

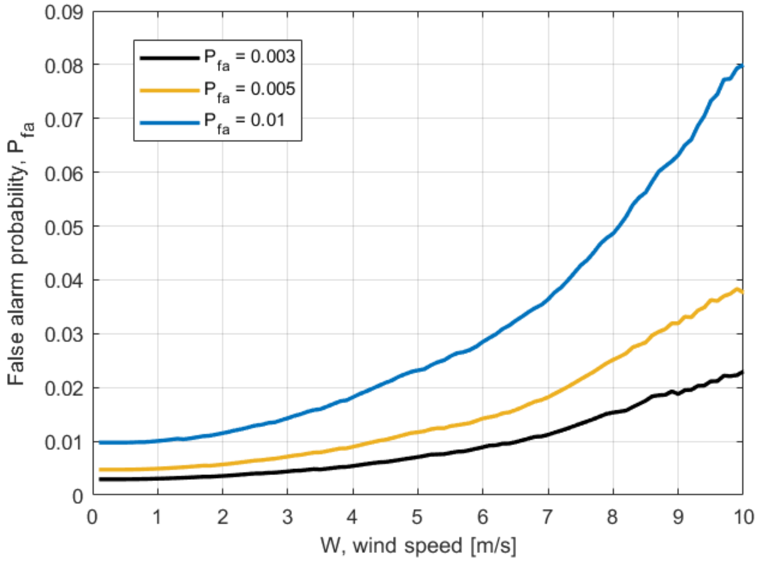


Fig. 7. False alarm probability for non-rainy samples with a wind speed greater than W. Three RF-based classifiers with different probabilities of false alarm (on the entire dataset) are considered.

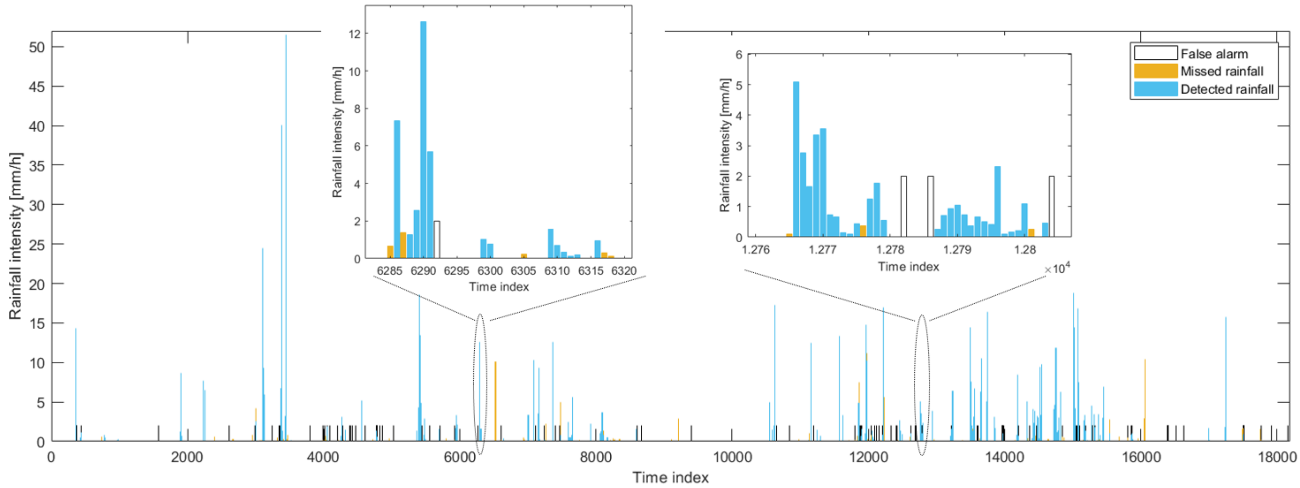


Fig. 8. Rainfall intensity during the 18,193 hours of observation (one sample per hour; about 25 months of data collection), with indication of detected rainy samples (620 hours), missed rainy samples (256 hours), false alarm samples (172 hours). The zoom panels show two examples of the occurrence of the three cases on a fine scale.

V. CONCLUSIONS

This study concerned the possibility of detecting precipitation, from drizzle phenomena to events of high intensity, using the underwater acoustic noise spectrum obtained from the average of the instantaneous spectra acquired, at various times, over the course of an hour. Since each sample is representative of an entire hour, to maintain sufficient temporal coverage it was necessary to analyze all the spectra acquired, even those altered by the passage of ships, high wind and other concurrent noises.

A dataset composed of more than 18,000 hours of measurements at sea allowed an in-depth experimentation of different rainfall detection methods. Although the rainfall detection by rule-based algorithms taken from the literature have not provided satisfying performance on this type of spectrum, machine learning methods have shown that the detection can be carried out successfully. In this analysis, kernel-based and ensemble-learning models have demonstrated the best performance among the experimented supervised classifiers. In particular, the RF-based binary classifier has shown a satisfactory balance between computational burden and performance, reaching a detection probability greater than 90% when precipitation exceeds 0.7 mm/h and P_{fa} is

1% or, alternatively, when precipitation exceeds 1.4 mm/h and P_{fa} is 0.3 %. This level of performance is slightly better than that obtained by a weather radar operating in the experiment area, and therefore the proposed method represents a promising alternative to obtain an estimate of rainfall intensity in areas where environmental constraints do not allow the installation of rain gauges or radar systems. This is even more noteworthy in polar areas, where global warming is changing the hydrological cycle of those regions, thus increasing rainfall with respect to snow precipitation [31].

While the presence of high wind, especially above 10 m/s, induced a noticeable increase in the probability of false alarm, the performances did not undergo significant alterations in the hours in which a ship transited in the area where the underwater measurement device was placed. Similarly, no fluctuations in performance were observed on a seasonal basis, attributable to varying underwater propagation conditions.

Although very promising, supervised learning models require a training phase that necessitates extensive collection of underwater acoustic spectra, accompanied by concomitant precipitation measurements to be used as ground truth. On the other hand, this is also partially necessary for rule-based algorithms that need specific calibrations to account for geographic location and hydrophone sensitivity. The possibility of using the trained detector in geographic areas other than the one in which the training data were collected is a topic for future investigation. However, it is reasonable to assume that in similar environmental settings, a trained detector can continue to operate successfully.

The performance obtained working on averaged spectra suggests that machine learning models may also be advantageous for rain detection using short-term acoustic spectra. This future research development is accompanied by a farther-reaching one: to design statistical learning models that act as regressors for accurate estimation of precipitation intensity and wind speed, making best use of information contained in multi-year time series of underwater acoustic noise.

ACKNOWLEDGMENT

The authors would like to acknowledge the contribution of the Regional Agency for environmental protection of Piedmont (ARPA) who jointly operate the radar system with the Regional Agency for environmental protection of Liguria (ARPAL) for providing weather radar data.

This work is dedicated to the memory of Jeffrey A. Nystuen who started the pioneering research on the underwater sound produced by rain with Walter Munk at Scripps Institute of Oceanography/UCSD in the early 1980s and provided us with one of the devices he developed at the Applied Physical Laboratory, University of Washington, Seattle, WA to initiate underwater noise collection in the Ligurian Sea.

REFERENCES

[1] Black, P. G., Proni, J. R., Wilkerson, J. C., Samsury, C. E. (1997). Oceanic rainfall detection and classification in tropical and subtropical mesoscale convective systems using underwater acoustic methods. *Monthly Weather Review*, *125*(9), 2014-2042.

[2] Riser, S. C., Nystuen, J., Rogers, A. (2008). Monsoon effects in the Bay of Bengal inferred from profiling float-based measurements of wind speed and rainfall. *Limnology and Oceanography*, *53*(5part2), 2080-2093.

[3] Yang, J., Riser, S. C., Nystuen, J. A., Asher, W. E., Jessup, A. T. (2015). Regional rainfall measurements: using the Passive Aquatic Listener during the SPURS field campaign. *Oceanography*, *28*(1), 124-133.

[4] Wentz, F. J., Ricciardulli, L., Rodriguez, E., Stiles, B. W., Bourassa, *et al.* (2017). Evaluating and extending the ocean wind climate data record. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *10*(5), 2165-2185.

[5] Hou, A. Y., Kakar, R. K., Neeck, S., Azarbarzin, A. A., Kummerow, *et al.* (2014). The global precipitation measurement mission. *Bulletin of the American Meteorological Society*, *95*(5), 701-722.

[6] Amitai, E., Nystuen, J. A., Liao, L., Meneghini, R., Morin, E. (2004). Uniting space, ground, and underwater measurements for improved estimates of rain rate. *IEEE Geoscience and Remote Sensing Letters*, *1*(2), 35-38.

[7] Ma, B. B., Nystuen, J. A. (2005). Passive acoustic detection and measurement of rainfall at sea. *Journal of Atmospheric and Oceanic Technology*, *22*(8), 1225-1248.

[8] Amitai, E., Nystuen, J. A., Anagnostou, E. N., Anagnostou, M. N. (2007). Comparison of deep underwater measurements and radar observations of rainfall. *IEEE Geoscience and Remote Sensing Letters*, *4*(3), 406-410.

[9] Vagle, S., Large, W. G., Farmer, D. M. (1990). An evaluation of the WOTAN technique of inferring oceanic winds from underwater ambient sound. *Journal of Atmospheric and Oceanic Technology*, *7*(4), 576-595.

[10] Nystuen, J. A., Selsor, H. D. (1997). Weather classification using passive acoustic drifters. *Journal of Atmospheric and Oceanic Technology*, *14*(3), 656-666.

[11] Nystuen, J. A., Amitai, E., Anagnostou, E. N., Anagnostou, M. N. (2008). Spatial averaging of oceanic rainfall variability using underwater sound: Ionian Sea rainfall experiment 2004. *The Journal of the Acoustical Society of America*, *123*(4), 1952-1962.

[12] Nystuen, J. A., Moore, S. E., Stabeno, P. J. (2010). A sound budget for the southeastern Bering Sea: Measuring wind, rainfall, shipping, and other sources of underwater sound. *The Journal of the Acoustical Society of America*, *128*(1), 58-65.

[13] Nystuen, J. A., Anagnostou, M. N., Anagnostou, E. N., Papadopoulos, A. (2015). Monitoring Greek seas using passive underwater acoustics. *Journal of Atmospheric and Oceanic Technology, 32*(2), 334-349.

[14] Kuhner, J. (2018, October). Automating the Detection of Precipitation and Wind Characteristics in Navy Ocean Acoustic Data. In *OCEANS 2018 MTS/IEEE Charleston* (pp. 1-7). IEEE.

[15] Pensieri, S., Bozzano, R., Nystuen, J. A., Anagnostou, E. N., Anagnostou, M. N., *et al.* (2015). Underwater acoustic measurements to estimate wind and rainfall in the Mediterranean Sea. *Advances in Meteorology*, *2015*(612512), 1-19.

[16] Riser, S. C., Yang, J., Drucker, R. (2019). Observations of large-scale rainfall, wind, and sea surface salinity variability in the eastern tropical Pacific. *Oceanography*, *32*(2), 42-49.

[17] Cazau, D., Bonnel, J., Baumgartner, M. (2019). Wind speed estimation using acoustic underwater glider in a near-shore marine environment. *IEEE Transactions on Geoscience and Remote Sensing*, *57*(4), 2097-2106.

[18] Quartly, G. D., Guymer, T. H., Birch, K. G., Smithers, J., Goy, K., *et al.* (2000, July). Listening for rain: theory and practice. In *5th European Conference on Underwater Acoustics, Vol. 1* (pp. 1-6).

[19] Anagnostou, M. N., Nystuen, J. A., Anagnostou, E. N., Nikolopoulos, E. I., Amitai, E. (2008). Evaluation of underwater rainfall measurements during the Ionian Sea rainfall experiment. *IEEE Transactions on Geoscience and Remote Sensing*, *46*(10), 2936-2946.

[20] Taylor, W. O., Anagnostou, M. N., Cerrai, D., Anagnostou, E. N. (2020). Machine Learning Methods to Approximate Rainfall and Wind From Acoustic Underwater Measurements. *IEEE Transactions on Geoscience and Remote Sensing*, 1-12 (in press).

[21] Nystuen, J. A. (2011, June). Quantifying physical processes in the marine environment using underwater sound. In *Proceedings of 4th Underwater Acoustics & Measurements conference* (pp. 20-24).

[22] Yang, J., Asher, W. E., Riser, S. C. (2016, January). Rainfall measurements in the North Atlantic Ocean using underwater ambient sound. In *2016 IEEE/OES China Ocean Acoustics* (*COA*) (pp. 1-4). IEEE.

[23] Trucco, A. (2001). Detection of objects buried in the seafloor by a pattern-recognition approach. *IEEE Journal of Oceanic Engineering*, *26*(4), 769-782.

[24] Barngrover, C., Kastner, R., Belongie, S. (2014). Semisynthetic versus real-world sonar training data for the classification of mine-like objects. *IEEE Journal of Oceanic Engineering*, *40*(1), 48-56.

[25] Klausner, N. H., & Azimi-Sadjadi, M. R. (2020). Manifold-Based Classification of Underwater Unexploded Ordnance in Low-Frequency Sonar. *IEEE Journal of Oceanic Engineering*, *45*(3), 1034-1044.

[26] Canepa, E., Pensieri, S., Bozzano, R., Faimali, M., Traverso, P., *et al.* (2015). The ODAS Italia 1 buoy: More than forty years of activity in the Ligurian Sea. *Progress in Oceanography*, *135*, 48-63.

[27] Bozzano, R., Pensieri, S., Pensieri, L., Cardin, V., Brunetti, F., *et al*. (2013, June). The M3A Network of Open Ocean Observatories in the Mediterranean Sea. In *2013 MTS/IEEE OCEANS-Bergen* (pp. 10-14). IEEE.

[28] Anagnostou, M. N., Nystuen, J. A., Anagnostou, E. N., Papadopoulos, A., Lykousis, V. (2011). Passive aquatic listener (PAL): An adoptive underwater acoustic recording system for the marine environment. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, *626*, S94-S98.

[29] Pensieri, S., Bozzano, R., Anagnostou, M. N., Anagnostou, E. N., Bechini, R., *et al.* (2013, June). Monitoring the oceanic environment through passive underwater acoustics. In *2013 MTS/IEEE OCEANS-Bergen* (pp. 1-10). IEEE.

[30] Hastie, T., Tibshirani, R., Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.

[31] Bintanja, R. (2018). The impact of Arctic warming on increased rainfall. *Scientific reports*, *8*(1), 1-6.