System-to-Distribution Parameter Mapping for the Gini Index Detector Test Statistic via Artificial Neural Networks

Alan L. Lemes[a], Dayan A. Guimarães[b], Yvo M. C. Masselli[c]

[a] *National Institute of Telecommunications - Inatel, Santa Rita do Sapucaí, Brazil (e-mail: alan.lima@mtel.inatel.br).*
[b] *National Institute of Telecommunications - Inatel, Santa Rita do Sapucaí, Brazil (e-mail: dayan@inatel.br).*
[c] *National Institute of Telecommunications - Inatel, Santa Rita do Sapucaí, Brazil (e-mail: yvo@inatel.br).*

Abstract

The Gini index detector (GID) was recently proposed for cooperative spectrum sensing (CSS) in cognitive radio networks. It has low computational complexity, robustness against unequal and time-varying noise and received signal powers, and can outperform state-of-the-art detectors. In this article, artificial neural networks (ANNs) are applied to map the CSS system variables into those that parameterize the probability distributions of the GID test statistic under the hypotheses of absence ($\mathcal{H}_0$) and presence ($\mathcal{H}_1$) of the primary sensed signal. The results concerning the goodness-of-fit of the GID test statistic to candidate probability distributions demonstrate that the Stable distribution adequately characterizes the statistic under $\mathcal{H}_0$, whereas the Generalized Extreme Value distribution best applies to $\mathcal{H}_1$. Two ANNs are developed to establish the system-to-distribution parameter mapping, allowing theoretical calculations of the CSS performance metrics and the decision threshold via closed-form expressions. The theoretical results are verified by computer simulations.

Keywords: Artificial neural networks, binary hypothesis test, cognitive radio,

*This work was partially supported by RNP, with resources from MCTIC, Grant No. 01250.075413/2018-04, under the Radio-communications Reference Center (Centro de Referência em Radiocomunicações - CRR) project of the National Institute of Telecommunications (Instituto Nacional de Telecomunicações - Inatel), Brazil

cooperative spectrum sensing, Gini index detector, goodness-of-fit.

1. Introduction

The demand for free bands in the radio-frequency (RF) spectrum has increased with the deployment of existing wireless communication services and technologies, as well as the development of new ones, owed mainly to the fixed allocation policy in which the various services use the spectrum on a primary or licensed basis. This situation may be aggravated by the massive deployment of the Internet of Things (IoT) and the fifth generation (5G) of wireless communication networks.

Despite the apparent RF spectrum scarcity, a study conducted by the Federal Communications Commission (FCC) concluded that a large amount of bands has different degrees of idleness, depending on the time and the geographical location, being considerably underutilized [1].

The concept of cognitive radio (CR) has emerged as a promising solution to the RF spectrum shortage [2], allowing the shared spectrum access between primary users (PUs), those who hold the right to use a certain band, and cognitive secondary users (SUs), improving the efficiency of the spectrum usage [2]. The technique used by the CRs to help the secondary network identify idle bands is called spectrum sensing [3].

Spectrum sensing can be performed independently by each SU or cooperatively, the latter being the most used to overcome the degrading effects of fading, shadowing and hidden terminals, taking advantage of the spatial diversity produced by the different locations of the SUs in cooperation. As a consequence of adopting a cooperative spectrum sensing (CSS) approach, the decisions made on the occupation state of the sensed band become more reliable [4] in comparison with the non-cooperative approach.

In centralized cooperative spectrum sensing with data fusion [4], which is the one adopted herein, the samples collected by the SUs are sent to a fusion center (FC) thorough a control channel. At the FC, the final decision (also called

global decision) is made on the occupancy of the sensed band, and informed to the SUs using broadcast transmissions also via a control channel.

1.1. Problem description

The performance of the spectrum sensing is commonly measured by the probability of false alarm, P_{fa}, and the probability of detection, P_{d}. The former is the probability of making a decision in favor of the presence of the primary signal in the sensed band, given that the band is vacant (hypothesis $\mathcal{H}_0$). The latter is the probability of deciding in favor of the presence of the primary signal given that, in fact, the sensed band is occupied (hypothesis $\mathcal{H}_1$) [4]. Mathematically,

$$P_{\mathrm{fa}} = \Pr[T > \lambda | \mathcal{H}_0] = \int_{\lambda}^{\infty} f(t|\mathcal{H}_0) dt = 1 - F(\lambda|\mathcal{H}_0), \tag{1}$$

$$P_{\mathrm{d}} = \Pr[T > \lambda | \mathcal{H}_1] = \int_{\lambda}^{\infty} f(t|\mathcal{H}_1) dt = 1 - F(\lambda|\mathcal{H}_1), \tag{2}$$

where $\Pr[\cdot]$ denotes the probability of occurrence of the underlying event, T is the test statistic computed according to the spectrum sensing technique adopted, λ is the decision threshold, $f(t|\mathcal{H}_0)$ and $F(t|\mathcal{H}_0)$ are, respectively, the probability density function (PDF) and the cumulative distribution function (CDF) of T under the hypothesis $\mathcal{H}_0$, and $f(t|\mathcal{H}_1)$ and $F(t|\mathcal{H}_1)$ are respectively the PDF and the CDF of T under $\mathcal{H}_1$. A high P_{d} is aimed at, in order to protect the primary network from interferences that may be caused by secondary transmissions in bands mistakenly considered to be unoccupied. It is also desired that P_{fa} is low in order to increase the probability of opportunistic use of the spectrum, consequently raising the data throughput of the secondary network.

From (1) and (2) it can be noticed that P_{fa} and P_{d} can be computed analytically only if $f(t|\mathcal{H}_0)$ and $f(t|\mathcal{H}_1)$ or, equivalently, $F(t|\mathcal{H}_0)$ and $F(t|\mathcal{H}_1)$ are known. The knowledge of these functions is also necessary to calculate the decision threshold λ, which is typically set to achieve a given constant false alarm rate (CFAR) of the spectrum sensing process. In other words, λ is calculated

to attain the target P_{fa}, while P_{d} is governed by the signal-to-noise ratio (SNR) across the SUs receivers. The threshold λ is given by

$$\lambda = F^{-1}(1 - P_{\mathrm{fa}}|\mathcal{H}_0), \tag{3}$$

where F^{-1} is the inverse function of $F(\lambda|\mathcal{H}_0)$, that is, λ is the value of t that makes $F(t|\mathcal{H}_0) = 1 - P_{\mathrm{fa}}$.

The knowledge of PDFs and CDFs of a test statistic carries practical significance only if these functions can be parameterized according to the CSS system variables. This is the condition for allowing the assessment of the CSS performance during the project phase, checking in advance the situations or combinations of system parameters under which the test statistic will be effective in accomplishing the spectrum sensing task as desired. However, it is not always a trivial task to establish the relationship between the CSS system variables and the parameters of the distributions of a test statistic, which often requires very complex analytical methods. There are also cases in which the test statistic is formed in such a way that the mathematical analysis of its distributions under $\mathcal{H}_1$, primarily, and sometimes even under $\mathcal{H}_0$, become intractable.

1.2. Contributions

The mathematical intractability while seeking for the PDFs or CDFs of a test statistic applies to the Gini index detector (GID) [5]. This fact motivated the authors of the present article to apply artificial neural networks (ANNs) specifically designed and trained to establish the system-to-distribution parameter mapping for the GID test statistic. With such mapping it is possible to apply the closed-form expressions also derived herein to compute P_{fa}, P_{d} and λ for a wide range of system parameters of practical relevance. Numerical results are also reported to validate the theoretical findings.

The choice of the GID among many other test statistics carrying such mathematical intractability is owed to the fact that it proved to be quite robust in the scenarios of unequal and possibly time-varying noise and received signal powers across the SUs, a situation that is common in practice. Besides, the

GID exhibits the important property of CFAR and has low complexity, since it only requires the computation of the received signal sample covariance matrix (SCM) and a few more operations to form its test statistic. These features make the GID an option for devices that have low processing power and require high energy efficiency, such as IoT devices and certain mobile communication terminals. Moreover, the GID is blind, meaning that no prior knowledge about the primary signal and the noise variance is required. In spite of the choice by the GID, the methodology adopted herein can be adapted to any test statistic that poses difficulties on finding its distribution under $\mathcal{H}_0$, $\mathcal{H}_1$ or both.

1.3. Related work

Among the spectrum sensing techniques, the mostly known are the energy detection, the matched filter, the cyclostationary feature detection [4], and those techniques based on the eigenvalues of the received signal SCM [6].

As far as the distributions of test statistics and corresponding expressions for computing performance metrics are concerned, one could mention, as extreme cases, the energy detector and some eigenvalue-based test statistics. The former carries a relatively simple analytic development [7], whereas the later may typically resort to extremely complex random matrix analysis. For instance, in [8], the moment-matching method is used to parameterize the Beta distribution for the Hadamard ratio test, under the hypothesis $\mathcal{H}_1$. The same approach is applied in [9] to characterize the Beta distribution of the sphericity test. In [10] it is demonstrated that the Gamma distribution can be used to evaluate the performance of volume-based spectrum sensing techniques.

Regarding the use of ANNs in the wide context of telecommunications, in [11] an ANN is trained to find system parameters that optimize the performance of a communication system. In [12], ANNs are applied to solve the transmitter identification problem. A review about several research papers that deal with the use of ANNs in the context of cognitive radio is presented in [13], where they are explored to assist the choice of modulation, in modulation identification and signal detection, and to improve the performance of spectrum sensing. The

implementation of spectrum sensing assisted by ANNs is also addressed in [14] and [15]. The broader approach of artificial intelligence (AI) is being recently applied to allow radios to learn how spectrum can be shared among competitors, aiming at increasing the overall data throughput [16].

In none of the above-mentioned research the use of the goodness-of-fit (GoF) process has been identified as an alternative for determining the distributions of the test statistics, probably because of the subsequent difficulty on mapping the parameters of the distributions found into the spectrum sensing system parameters. Additionally, the application of ANNs as an alternative to the parameterization of the distributions of test statistics as functions of the system variables has not been found in the literature. Thus, the authors believe that the proposals in this article, as summarized in Section 1.2, can serve as reference to readers for solving similar problems, as an alternative to analytical approaches that can often be highly complex or even mathematically intractable.

The remainder of the article is organized as follows: Section 2 describes the system model and the GID. Section 3 is devoted to the process of fitting the GID test statistic to candidate probability distributions. The design of the neural networks for system-to-distribution mapping is described in Section 4. Section 5 presents theoretical expressions for performance analysis of the GID-based CSS system, and reports numerical results to verify the accuracy of these expressions. Section 6 summarizes the main findings of the work.

2. System model

It is assumed a centralized CSS with data fusion, in which the cognitive network applies m SUs for spectrum sensing, each having a single antenna, or a single SU with m antennas. The number of samples collected per antenna during each sensing interval is n. The total of mn samples are transmitted to the FC through an error-free control channel. The samples received by the FC are arranged in the matrix $\mathbf{Y} \in \mathbb{C}^{m \times n}$, which is given by

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{V}. \tag{4}$$

In this equation, the samples of the signals emitted by p PU transmitters are arranged in the matrix $\mathbf{X} \in \mathbb{C}^{p \times n}$, whose elements have Gaussian distribution with zero mean and variance dependent on the SNR, modeling the envelope fluctuations of typical modulated and filtered signals [6].

The channel matrix $\mathbf{H} \in \mathbb{C}^{m \times p}$ has elements h_{ij}, $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, p$, representing the sensing channel gain between the j-th PU and the i-th SU. Aiming at modeling a flat and slow Ricean fading channel, the channel matrix is given by $\mathbf{H} = \mathbf{GA}$, where $\mathbf{A} \in \mathbb{C}^{m \times p}$ is formed by elements $a_{ij} \sim \mathbb{CN}[\sqrt{K/(2K+2)}, 1/(K+1)]$, yielding unitary second moment for the fading magnitude, where K is the Rice factor [5], [17, pp. 211-219]. The matrix $\mathbf{G} \in \mathbb{R}^{m \times m}$ is defined as $\mathbf{G} = \text{diag}(\sqrt{\mathbf{s}/s_{\text{avg}}})$, where diag$(\cdot)$ returns a diagonal matrix whose main diagonal is formed by the vector in the argument, $\mathbf{s} = [s_1, s_2 ..., s_m]^{\text{T}}$ is the vector with the received signal powers across the SUs, and $s_{\text{avg}} = (1/m) \sum_{i=1}^{m} s_i$ is the received signal power averaged over all SUs.

The matrix $\mathbf{V} \in \mathbb{C}^{m \times n}$ in (4) contains additive white Gaussian noise (AWGN) samples having zero mean and SNR-dependent variance σ^2_{avg}, as follows: given that the average transmitted signal power over all PUs is s_{avg}, the average SNR, in decibels, is $10 \log_{10}(s_{\text{avg}}/\sigma^2_{\text{avg}})$, where $\sigma^2_{\text{avg}} = (1/m) \sum_{i=1}^{m} \sigma_i^2$ is the noise power averaged over all SUs.

Aiming at modeling the scenario in which the noise and the received signal powers at the SUs may be unequal and time-varying, here it is assumed that these powers are uniformly distributed across the SUs, independently varying between consecutive sensing rounds. Specifically, $\sigma_i^2 \sim \mathcal{U}[0.05\sigma^2_{\text{avg}}, 1.95\sigma^2_{\text{avg}}]$ and $s_i \sim \mathcal{U}[0.05 s_{\text{avg}}, 1.95 s_{\text{avg}}]$ in each realization of the spectrum sensing.

From the matrix $\mathbf{Y}$, the received signal SCM is computed at the FC as

$$\mathbf{R} = \frac{1}{n} \mathbf{Y} \mathbf{Y}^{\dagger}, \tag{5}$$

where $\dagger$ denotes the Hermitian operation (complex conjugate and transpose). Subsequently, the FC computes the GID test statistic [5], according to

$$T = \frac{2(m^2 - m) \sum_{i=1}^{m^2} |r_i|}{\sum_{i=1}^{m^2} \sum_{j=1}^{m^2} |r_i - r_j|}, \tag{6}$$

where r_i is the i-th element of the vector $\mathbf{r}$ that is formed by stacking all columns of the matrix $\mathbf{R}$. Originally, the Gini index has been proposed as a measure of social inequality, which means that the values in $\mathbf{r}$ are interpreted as having the social parameter measures, for example incomes. In [5], the Gini index is adapted to the spectrum sensing scenario, where it measures the inequality between the elements of the SCM when the tested hypotheses change.

The global decision upon the occupation state of the sensed band is finally made in favor of $\mathcal{H}_1$ if $T > \lambda$, recalling that λ is the decision threshold. If $T \leq \lambda$, the decision is made in favor of $\mathcal{H}_0$.

3. Goodness-of-fit to the GID test statistic under $\mathcal{H}_0$ and $\mathcal{H}_1$

This section describes the process of fitting the GID test statistic distribution to several candidate probability distributions. Initially, samples containing values of the test statistic T were generated under $\mathcal{H}_0$ and $\mathcal{H}_1$. Such samples were subjected to the maximum likelihood (ML) parameter estimation of the continuous distributions available to the built-in distribution fitting function `fitdist` of the MATLAB software, version 2018a, excluding the Beta and the Kernel distributions. The Beta distribution has been excluded because it is restricted to the range $[0, 1]$, and the GID test statistic may have a different support. The Kernel distribution has been excluded because it is non-parametric, thus not being applicable to the problem at hand. The remaining candidate distributions are listed in Table 1, along with the corresponding number of parameters.

Two metrics have been adopted to assess the degree of adherence between each candidate distribution and the empirical distribution generated from the values of the test statistic T defined in (6): the mean squared error (MSE) between the distributions, and the average of the Kolmogorov-Smirnov statistic (KSS) resulting from the Kolmogorov-Smirnov GoF test [18], here named mean KSS (MKSS). After the GoF test, the candidate distributions were ranked by ordering these metrics.

The use of the MSE in conjunction with the MKSS has been adopted due

Table 1: Selected continuous distributions operated by the MATLAB function `fitdist`

Distribution	Number of parameters
Birnbaum-Saunders	2
Burr	3
Exponential	1
Extreme value	2
Gamma	2
Generalized extreme value	3
Generalized Pareto	3
Half-normal	2
Inverse Gaussian	2
Logistic	2
Log-logistic	2
Log-normal	2
Nakagami	2
Normal	2
Rayleigh	2
Rician	2
Stable	4
t location-scale	3
Weibull	2

to the varying and sometimes small number of occurrences in which each of the candidate distributions is accepted in the Kolmogorov-Smirnov test, under the typical significance level of 0.05. In other words, the simple fact that a particular distribution is accepted as representative of a given sample is not sufficient to carry out a classification in terms of GoF. As demonstrated by the analyzes presented in Sections 3 and 5, this procedure unveiled to be suitable to the problem at hand and precise enough to determine the distributions of T under the hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$.

Specifically, a sample with 70000 values of T was generated under each hypothesis, for 70 random combinations of the CSS parameters: $p = \{1, 2, ..., 4\}$ PUs; $m = \{2, 3, ..., 7\}$ SUs; $n = \{10, 20, ..., 100\}$ samples; Rice factor $K = \{2, 3, ..., 12\}$, and SNR $= \{-15, -14, ..., -5\}$ dB under $\mathcal{H}_1$ (the SNR does not matter under $\mathcal{H}_0$). For each sample, the Kolmogorov-Smirnov GoF test was carried out for each of the distributions given in Table 1, using the MATLAB

function `kstest`. The MSE and the KSS between each empirical and candidate distribution were subsequently computed and stored for each sample, at the end being averaged to yield the global MSE and MKSS. The **Algorithm 1** shows the steps of this process, which has been applied to $\mathcal{H}_0$ and $\mathcal{H}_1$.

Algorithm 1 GoF of the GID test statistic using the MSE and the MKSS

$Runs \leftarrow 70000$, $Loops \leftarrow 70$, $\Sigma_{\text{MSE}} \leftarrow 0$, $\Sigma_{\text{KSS}} \leftarrow 0$
for $i = 1:1:Loops$ **do**
  Generate the system parameters
  for $j = 1:1:Runs$ **do**
    Generate T_i according to (6), under $\mathcal{H}_0$ or $\mathcal{H}_1$
  end
  Generate the empirical CDF (CDF_e) of the sample $\{T_j\}$
  Using $\{T_j\}$, estimate the parameters of the CDFs in Table 1
  For each CDF in Table 1, compute the theoretical CDF (CDF_t)
  Compute the KSS related to CDF_e and CDF_t
  Compute the MSE between CDF_e and CDF_t
  $\Sigma_{\text{MSE}} \leftarrow \Sigma_{\text{MSE}} + \text{MSE}$ for each CDF_t
  $\Sigma_{\text{KSS}} \leftarrow \Sigma_{\text{KSS}} + \text{KSS}$ for each CDF_t
end
$\text{MSE} \leftarrow \Sigma_{\text{MSE}}/Loops$ for each CDF_t
$\text{MKSS} \leftarrow \Sigma_{\text{KSS}}/Loops$ for each CDF_t

Tables 2 and 3 show the ranks of the distributions that attained the five smallest MSEs and MKSSs for the hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. Notice that the ranks are not the same with respect to both metrics under $\mathcal{H}_0$, but are the same up to the third position. Under $\mathcal{H}_1$ the ranks are the same.

Table 2: Top five distributions under $\mathcal{H}_0$

Distribution	MSE (rank)	MKSS (rank)
Stable	6.09×10^{-5} (1)	1.24×10^{-2} (1)
Burr	1.33×10^{-4} (2)	1.86×10^{-2} (2)
Log-logistic	1.74×10^{-4} (3)	2.26×10^{-2} (3)
t location-scale	1.98×10^{-4} (4)	2.68×10^{-2} (5)
Gen. extreme value	2.23×10^{-4} (5)	2.43×10^{-2} (4)

Under $\mathcal{H}_0$, it can be seen in the Table 2 that the Stable distribution is the best-ranked in terms of MSE and MKSS. From Table 3, it can be seen that the

Table 3: Top five distributions under $\mathcal{H}_1$

Distribution	MSE (rank)	MKSS (rank)
Gen. extreme value	4.38×10^{-5} (1)	1.02×10^{-2} (1)
Stable	1.37×10^{-4} (2)	1.93×10^{-2} (2)
Burr	1.73×10^{-4} (3)	2.21×10^{-2} (3)
Log-logistic	4.16×10^{-4} (4)	3.41×10^{-2} (4)
t location-scale	6.96×10^{-4} (5)	4.80×10^{-2} (5)

Generalized extreme value (GEV) distribution occupies the top rank position in terms of both metrics. Hence, from this point onward the Stable distribution is adopted to characterize the GID test statistic T defined in (6) under $\mathcal{H}_0$, whereas the GEV is the selected distribution to characterize T under $\mathcal{H}_1$.

The parameters of the Stable distribution and their roles are: α ($0 < \alpha \leq 2$) and β ($-1 \leq \beta \leq 1$) determine the shape of the distribution, γ ($0 < \gamma < \infty$) is the scale, and δ ($-\infty < \delta < \infty$) determines the location. The Stable PDF and CDF cannot be expressed analytically, except for certain values of its parameters [19]. The parameters of the GEV distribution and their roles are: μ ($-\infty < \mu < \infty$) is the location parameter, σ ($\sigma > 0$) is the scale, and k ($-\infty\ k < \infty$) governs the shape of the distribution. The GEV distribution has a considerably simple mathematical representation [20].

The Stable and the GEV distributions are included in the list of functions whose computations have become numerically stable, sufficiently mature and well accepted, to the point that expressions containing these functions can be considered closed-form. For example, the PDF, the CDF and the inverse CDF of the Stable and the GEV distributions are embedded into the MATLAB environment, allowing for calculations as easily as happens with the complementary error function and other similar functions that can only be operated numerically.

Figures 1 and 2 respectively illustrate the adherence of the Stable and the GEV distributions to empirical CDFs of the GID test statistic generated from three different sets of the CSS system parameters. It can be seen that the theoretical and the empirical CDFs are practically overlapped in all cases. Additional

validation results are presented in Section 5.

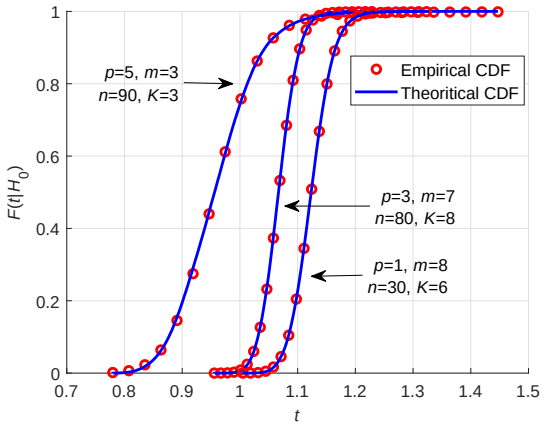

Figure 1: Empirical and theoretical CDFs under $\mathcal{H}_0$

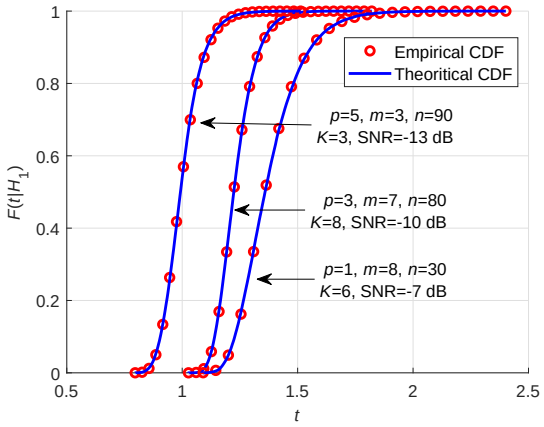

Figure 2: Empirical and theoretical CDFs under $\mathcal{H}_1$

4. Parameter mapping via artificial neural networks

4.1. Data set generation

In order to generate the whole data set, which has been subsequently divided into subsets for training, validation and testing the ANNs, a routine has been developed in the MATLAB software, according to **Algorithm 2**. The CSS system parameters were formed by all combinations of: $p = \{1, 2, 3, 4\}$ PUs, $m = \{2, 3, 4, 5, 6, 7\}$ SUs, $n = \{10, 20, ..., 100\}$ samples per SU, Rice factor $K = \{2, 3, ..., 12\}$, and SNR $= \{-15, -14, ..., -5\}$ dB, yielding 29040 combinations. For each combination, the parameters of the CDFs under $\mathcal{H}_0$ and $\mathcal{H}_1$ were estimated under the ML criterion, applying the MATLAB function `fitdist` to 50000 samples of the GID test statistic T defined in (6). Recall that the actual CDF parameters under $\mathcal{H}_0$ do not change with the SNR, but have been estimated anyway, given the time-domain variations of the noise powers across the SUs, and the noise waveform itself.

Algorithm 2 Generation of the data set

```
Loop ← 1 ; Runs ← 50000
for p = 1 : 1 : 4 do
    for m = 2 : 1 : 7 do
        for SNR = −15 : 1 : −5 do
            for n = 10 : 10 : 100 do
                for K = 2 : 1 : 12 do
                    for i = 1 : 1 : Runs do
                        Generate T_i according to (6), under H_0 and H_1
                    end
                    Estimate α, β, γ and δ of Stable PDF under H_0
                    Estimate μ, σ and k of GEV PDF under H_1
                    Save: Inputs(loop, :) ← [p m n K SNR]
                    Save: Outputs(loop, :) ← [α β γ δ μ σ k]
                    Loop ← Loop + 1
                end
            end
        end
    end
end
```

Table 4 shows some statistics of the estimates of the distribution parameters

associated to the complete data set. Later on in this article, such statistics will allow the assessment of the ANN accuracy in the task of mapping the system variables into the parameters of the Stable and GEV distributions.

Table 4: Statistics of the parameters of the Stable and GEV distributions

	Maximum	Minimum	Mean	Std. deviation
α	1.9741	1.2832	1.8374	0.1251
β	1	1	1	0
γ	0.0975	0.0188	0.0415	0.0217
δ	1.1461	0.8881	1.0118	0.0689
k	0.3448	-0.1253	-0.0101	0.0737
σ	0.2499	0.0311	0.0990	0.0431
μ	1.6603	0.8718	1.1175	0.1390

It is interesting to notice in Table 4 that the parameter β of the Stable distribution is always 1. Besides, it can be observed that the values of α are high on average, which means that β does not significantly influence the shape of the PDF [19]. As a consequence, $\beta = 1$ can be safely adopted hereafter, regardless the system parameters.

In summary, there will be one ANN with five inputs and three outputs to map the system variables $\{p, m, n, K, \text{SNR}\}$ into the parameters $\{k,\ \sigma,\ \mu\}$ of the GEV distribution, and another ANN with four inputs and three outputs to map the system variables $\{p, m, n, K\}$ into the parameters $\{\alpha,\ \gamma,\ \delta\}$ of the Stable distribution. Since $\beta = 1$, it is obviously not necessary to set this parameter as an output of the second ANN.

4.2. Artificial neural network training

Multilayer perceptron (MLP) [21] ANNs have been adopted in this work. For these ANNs, the input-output equation is the general formula of linear interpolation [22]. Hence, after trained such ANNs are capable of approximating any continuous non-linear function under certain precision. Owed to this characteristic, the ANNs of type MLP are suitable for the GID system-to-distribution parameter mapping, additionally allowing, due to their interpolation ability, the

estimation of distribution parameters associated to system variables intermediate to those used in the training phase. Moreover, it will be demonstrated that the extrapolation ability is also possible in specific cases.

The ANNs were simulated using the MATLAB deep learning toolbox (formerly neural network toolbox), using 70% of the data set for for training, 20% for validation, and 10% for testing. The activation function used in the ANN neurons was the hyperbolic tangent, and the Levenberg Marquardt algorithm was used during the training phase by the gradient-descent optimization.

The number of training epochs was defined according to a large number of preliminary tests with large numbers of epochs. It has been observed that, above 2000 epochs, the mean squared error between actual and estimated parameters did not reduce significantly. Thus, in order to avoid over-fitting, the maximum number of epochs was limited to 2000. This relatively small number enables ANN retraining, if necessary, within a very short time.

The learning procedure was terminated whenever the validation error monotonically increased for 15 epochs, after reaching a local (possibly global) minimum. The set of synaptic weights producing the minimum error just before these 15 epochs was defined as the optimal. A zero target MSE and a learning rate of 0.05 were also adopted.

Tables 5 and 6 report the average of the MSEs between the ANNs' output parameters and the actual ones (i.e., those estimated from the data under the ML criterion), respectively for the Stable and the GEV distributions, as a function of the number of neurons in the hidden layer. These MSEs were averaged over all combinations of system parameters used for training. The actual number of training epochs is also given in these tables for reference.

It can be seen that the increase in the number of neurons causes a reduction in the MSE, as expected. However, under the hypothesis $\mathcal{H}_0$ it can be seen that there is no significant reduction of the MSE above 15 neurons. Under $\mathcal{H}_1$, the same happens when the number of neurons approaches 30. Therefore, these were the numbers of neurons adopted in the hidden layers of the ANNs that map the parameters of the Stable and GEV distributions, respectively.

The ANNs were translated into the MATLAB functions[1] `NnH0([p m n K])` and `NnH1([p m SNR n K])`. The former returns a vector containing the parameters α, γ and δ, in this order, of the Stable distribution that characterizes the GID test statistic T under $\mathcal{H}_0$ (recall that $\beta = 1$, always), as a function of the CSS system parameters p, m, n and K. The latter returns a vector with the parameters k, σ and μ, in this order, of the GEV distribution that characterizes T under $\mathcal{H}_1$, as a function of the system parameters p, m, SNR, n and K.

Aiming at assessing the performances of the designed ANNs, the set of system variables used to generate the training data was applied as input to the above functions. After obtaining ANN responses for each of the 29040 combinations of input parameters, the errors between such responses and the parameters obtained by ML estimation were calculated. Figures 3 and 4 depict the histograms of the mapping errors for the hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. Comparing the dispersion of these histograms, measured by the standard deviation (SD), with the corresponding standard deviations given in Table 4, it is concluded that the estimation errors produced by the ANNs are quite small. Besides, it is observed that the errors are symmetric about zero, which means that the mappings performed by the ANNs are unbiased. These conclusions are strengthened in the next section, where the performances of the CSS system obtained from Monte Carlo simulations are compared with those obtained

[1]The source codes for these functions (`NnH0.m` and `NnH1.m`) can be accessed in [23]. These codes must be in the same directory of the main routine that uses them, and do not depend on the MATLAB machine learning toolbox to run.

Table 5: MSE provided by the ANN under $\mathcal{H}_0$

Nº of neurons	MSE	Nº of Epochs
5	9.95×10^{-6}	331
10	6.40×10^{-6}	672
15	6.38×10^{-6}	206
20	6.37×10^{-6}	119
25	6.32×10^{-6}	167

Table 6: MSE provided by the ANN under $\mathcal{H}_1$

Nº of neurons	MSE	Nº of Epochs
5	1.10×10^{-4}	615
10	2.63×10^{-5}	1338
15	1.14×10^{-5}	687
20	8.08×10^{-6}	898
25	6.12×10^{-6}	878
30	4.65×10^{-6}	711
35	4.18×10^{-6}	1087

theoretically, using the parameters mapped by the designed ANNs.

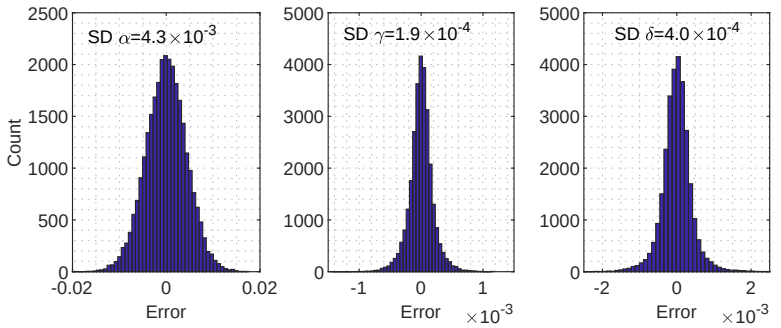

Figure 3: Histograms of mapping errors under $\mathcal{H}_0$

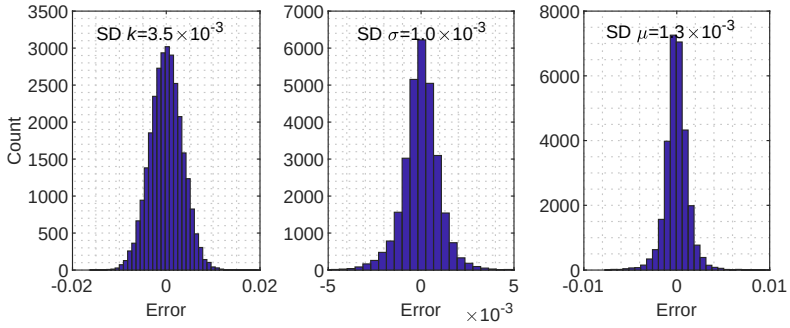

Figure 4: Histograms of mapping errors under $\mathcal{H}_1$

5. Spectrum sensing performance

The probability of detection, P_{d}, and the probability of false alarm, P_{fa}, of a spectrum sensing technique are commonly evaluated using the receiver operating characteristic (ROC) curve, where P_{d} is traded off against P_{fa} as the decision threshold λ is varied. For the simulation results presented subsequently, each point on a ROC curve was determined from 50000 Monte Carlo events in which the GID test statistic was generated for a given system setting, as described in Section 2.

In terms of theoretical performance analysis, taking as references the Equations (1), (2) and (3), it follows that

$$P_{\mathrm{fa}} = 1 - F_{\mathrm{S}}(\lambda, \alpha, 1, \gamma, \delta), \tag{7}$$

where $F_{\mathrm{S}}(\lambda, \alpha, 1, \gamma, \delta)$ is the value of the Stable CDF parameterized by α, $\beta = 1$, γ and δ, at the point $t = \lambda$, and

$$P_{\mathrm{d}} = 1 - F_{\mathrm{G}}(\lambda, k, \sigma, \mu), \tag{8}$$

where $F_{\mathrm{G}}(\lambda, k, \sigma, \mu)$ is the value of the GEV CDF parameterized by λ, k, σ and μ, at the point $t = \lambda$, with the decision threshold determined according to the target probability of false alarm, P_{fa}, using

$$\lambda = {F_{\mathrm{S}}}^{-1}(1 - P_{\mathrm{fa}}, \alpha, 1, \gamma, \delta), \tag{9}$$

where ${F_{\mathrm{S}}}^{-1}(1 - P_{\mathrm{fa}}, \alpha, 1, \gamma, \delta)$ is the value of the inverse Stable CDF at $1 - P_{\mathrm{fa}}$.

It is important to stress that (7), (8) and (9) can be considered closed-form expressions, since they have numerically stable computations algorithms available in several modern mathematical software tools. In the MATLAB environment, these expressions are respectively written as

$$P_{\mathrm{fa}} = \texttt{1-cdf('Stable'}, \lambda, \alpha, 1, \gamma, \delta), \tag{10}$$

$$P_{\mathrm{d}} = \texttt{1-cdf('GeneralizedExtremeValue'}, \lambda, k, \sigma, \mu), \tag{11}$$

$$\lambda = \texttt{icdf('Stable'}, 1 - P_{\text{fa}}, \alpha, 1, \gamma, \delta). \tag{12}$$

In order to verify the accuracy of the system-to-distribution parameter mapping carried out by the ANNs, Figure 5 shows theoretical and simulated performances of the GID for three sets of arbitrarily chosen CSS system variables: $\{p = 4,\ m = 3,\ n = 50,\ K = 12,\ \text{SNR} = -5\ \text{dB}\}$, $\{p = 3,\ m = 5,\ n = 80,\ K = 3,\ \text{SNR} = -11\ \text{dB}\}$, and $\{p = 1,\ m = 4,\ n = 20,\ K = 7,\ \text{SNR} = -10\ \text{dB}\}$. The corresponding parameters of the distributions mapped by the ANNs were: $\{\alpha \approx 1.8011,\ \beta = 1,\ \gamma \approx 0.0487,\ \delta \approx 0.9637,\ k \approx -0.0466,\ \sigma \approx 0.1888,\ \mu \approx 1.2856\}$, $\{\alpha \approx 1.9166,\ \beta = 1,\ \gamma \approx 0.0262,\ \delta \approx 1.0242,\ k \approx -0.0420,\ \sigma \approx 0.0573,\ \mu \approx 1.0825\}$ and $\{\alpha \approx 1.7426,\ \beta = 1,\ \gamma \approx 0.0425,\ \delta \approx 1.0344,\ k \approx 0.0710,\ \sigma \approx 0.0844,\ \mu \approx 1.0864\}$. It can be seen in Figure 5 that empirical and theoretical ROCs are practically overlapped, again demonstrating that the Stable and the GEV distributions satisfactorily represent the GID test statistic under the hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. Moreover, it can be concluded that the ANNs developed herein are capable of estimating the parameters of such distributions with accuracy, allowing equally accurate theoretical performance calculations.

With the purpose of exemplifying the calculations, consider the upper curve in Figure 5, for $P_{\text{fa}} = 0.1$. It follows that $\lambda =$ `icdf('Stable', 0.9, 1.8011, 1, 0.0487, 0.9637)` ≈ 1.0719, and $P_{\text{d}} =$ `1-cdf('GeneralizedExtremeValue', 1.0719, -0.0466, 0.1888, 1.2856)` ≈ 0.9509. For the bottom curve, also for $P_{\text{fa}} = 0.1$, it can be obtained $\lambda =$ `icdf('Stable', 0.9, 1.7426, 1, 0.0425, 1.0344)` ≈ 1.1351, and $P_{\text{d}} =$ `1-cdf('GeneralizedExtremeValue', 1.1351, 0.0710, 0.0844, 1.0864)` ≈ 0.4414. It can be noticed that these values are in agreement with those read from the corresponding ROC curves in Figure 5.

As already mentioned, it is possible to map CSS system parameters different from those used during the training phase, owed to the fact that an ANN of type MLP is an interpolator by construction. To illustrate this possibility, Figure 6 presents theoretical and simulated ROC curves considering the following combi-

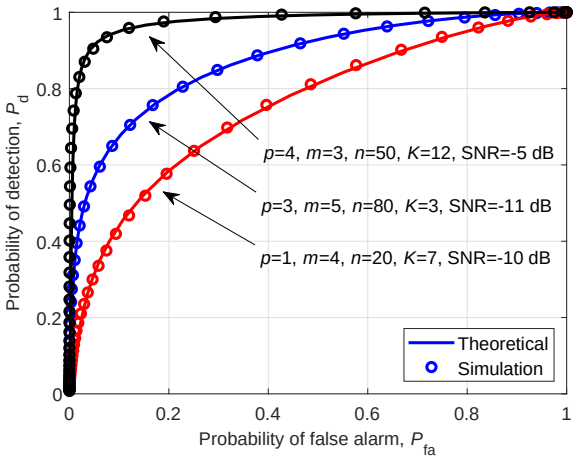

Figure 5: Theoretical and simulated GID performances for parameters present in the ANNs training data

nations of system variables: $\{p = 2, m = 4, n = 25, K = 7.7, \text{SNR} = -10.3 \text{ dB}\}$ and $\{p = 3, m = 6, n = 64, K = 4.8, \text{SNR} = -12,5 \text{ dB}\}$. These variables were mapped by the ANNs onto the following parameters of the Stable and GEV distributions: $\{\alpha \approx 1.9804, \beta = 1, \gamma \approx 0.0175, \delta \approx 1.0486, k \approx -0,1046, \sigma \approx 0.0577, \mu \approx 1.1591\}$, and $\{\alpha \approx 1.7679, \beta = 1, \gamma \approx 0.0407, \delta \approx 1.0276, k \approx 0.0479, \sigma \approx 0.0838, \mu \approx 1.0811\}$, respectively. From this figure it can be seen that theoretical and simulation results are in close agreement.

It was also verified that it is possible to use the ANNs developed to extrapolate results in situations of good performance of the CSS scheme, for instance $P_{\text{d}} > 0.8$ and $P_{\text{fa}} < 0.2$. This possibility is confirmed in Figure 6, where theoretical and simulated performances are very close to each other. The system variables considered in this case were $\{p = 4, m = 7, n = 150, K = 8, \text{SNR} = -11.5 \text{ dB}\}$, which were mapped into the following parameters of the Stable and GEV distributions: $\{\alpha \approx 1.9264, \beta = 1, \gamma \approx 0.0234, \delta \approx 1.0553, k \approx -0.0493, \sigma \approx 0.0467, \mu \approx 1.0965\}$.

The probabilities P_{d} and P_{fa} can be merged into a single quantity, the area under the ROC curve (AUC), which can be used as an alternative spectrum

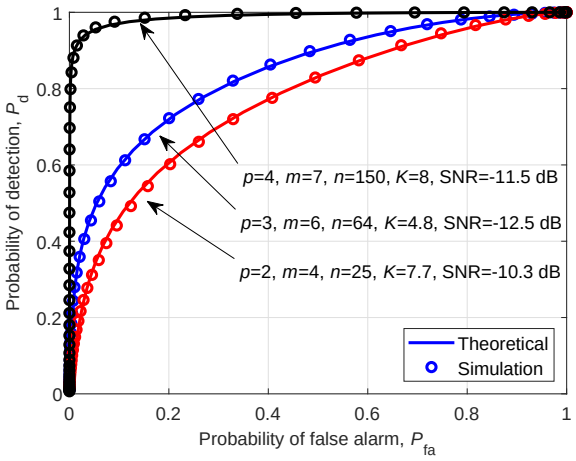

Figure 6: Theoretical and simulated GID performances for parameters not present in the ANNs training data

sensing performance metric. When AUC = 0.5, the worst performance is unveiled, corresponding to a useless detector with $P_{\rm d} = P_{\rm fa}$, which is equivalent to deciding arbitrarily for $\mathcal{H}_0$ or $\mathcal{H}_1$. When AUC = 1, the best performance is attained, corresponding to an ideal detector for which the ROC curve passes through the points ($P_{\rm d} = 1$, $P_{\rm fa} = 1$), ($P_{\rm d} = 1$, $P_{\rm fa} = 0$), and ($P_{\rm d} = 0$, $P_{\rm fa} = 0$). Thus, a nontrivial detector has $0.5 < \text{AUC} \leq 1$.

With the objective of identifying unsatisfactory system-to-distribution parameter mapping results, if any, the 29040 combinations of system variables used to train the ANNs were used to generate parameters of the Stable and the GEV distributions estimated by the ML criterion, as well as the ones mapped by the developed ANNs. Each pair of estimated and mapped parameters was used in a Monte Carlo simulation, generating the corresponding pair of ROC curves. The histogram of the difference (or error) between AUCs of all ROC pairs is shown in Figure 7. This histogram has a mean $\approx 3.39 \times 10^{-5}$ and a standard deviation ≈ 0.0028. Taking into account the limits of an AUC, it is concluded that the mapping errors produced by the ANNs are quite small, yielding practically imperceptible differences between two ROC curves.

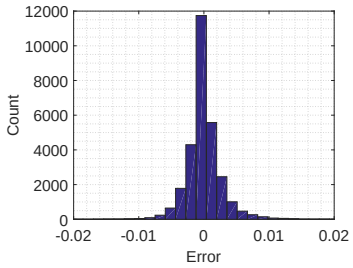

Figure 7: Histogram of the error between theoretical and empirical AUCs

6. Conclusions

This paper addressed the problem of mapping the cooperative spectrum sensing system variables into those that parameterize the probability distributions of the Gini index detector (GID) test statistic, under the null ($\mathcal{H}_0$) and the alternative ($\mathcal{H}_1$) hypotheses of absence and presence of the primary signal in the sensed band, respectively. A goodness-of-fit analysis unveiled the Stable and the Generalized Extreme Value distributions as representative of the GID test statistic under $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. Two artificial neural networks were developed to establish the system-to-distribution parameter mapping, allowing theoretical calculations of the spectrum sensing performance and the decision threshold via numerically-computable closed-form expressions.

Our findings can aid GID-related research, and also as support for the parameterization of this detector in practical applications, without the need for time-consuming computational simulations or field trials with real hardware. Additionally, it is possible to translate the functions that describe the developed neural networks into more comprehensive language environments like C, Java or Python, or even implement them in hardware.

The results presented may encourage the application of the method described herein to other test statistics whose analytical development to obtain their distributions under $\mathcal{H}_0$ and $\mathcal{H}_1$ is very complex or even mathematically intractable.

Acknowledgment

The authors are indebted to the anonymous referees for their valuable comments and feedback, which greatly contributed to the improvement of the article.

References

[1] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, S. Mohanty, Next generation/dynamic spectrum access/cognitive radio wireless networks: A survey, Computer networks 50 (13) (2006) 2127–2159.

[2] F. Hu, B. Chen, K. Zhu, Full spectrum sharing in cognitive radio networks toward 5G: A survey, IEEE Access 6 (2018) 15754–15776.

[3] Y. Arjoune, N. Kaabouch, A comprehensive survey on spectrum sensing in cognitive radio networks: Recent advances, new challenges, and future research directions, Sensors 19 (1), (2019).

[4] I. F. Akyildiz, B. F. Lo, R. Balakrishnan, Cooperative spectrum sensing in cognitive radio networks: A survey, Physical communication 4 (1) (2011) 40–62.

[5] D. A. Guimarães, Gini index inspired robust detector for spectrum sensing over Ricean channels, Electronics Letters 55 (12) (2019) 713–714.

[6] B. Nadler, F. Penna, R. Garello, Performance of eigenvalue-based signal detectors with known and unknown noise level, in: 2011 IEEE Int. Conf. on Commun. (ICC), IEEE, 2011, pp. 1–5.

[7] F. F. Digham, M.-S. Alouini, M. K. Simon, On the energy detection of unknown signals over fading channels, IEEE Trans. Commun. 55 (1) (2007) 21–24.

[8] L. Huang, Y. Xiao, H. C. So, J. Fang, Accurate performance analysis of hadamard ratio test for robust spectrum sensing, IEEE Trans. Wireless Commun. 14 (2) (2015) 750–758.

[9] L. Wei, O. Tirkkonen, Spectrum sensing in the presence of multiple primary users, IEEE Trans. Commun. 60 (5) (2012) 1268–1277.

[10] L. Huang, C. Qian, Y. Xiao, K. Q. Zhang, Performance analysis of volume-based spectrum sensing for cognitive radio, IEEE Trans. Wireless Commun. 14 (1) (2015) 317–330.

[11] N. Baldo, M. Zorzi, Learning and adaptation in cognitive radios using neural networks, in: 2008 5th IEEE Consumer Communications and Networking Conf., IEEE, 2008, pp. 998–1003.

[12] K. Merchant, S. Revay, G. Stantchev, B. Nousain, Deep learning for RF device fingerprinting in cognitive communication networks, IEEE J. Sel. Topics Signal Process. 12 (1) (2018) 160–167.

[13] S. Pattanayak, P. Venkateswaran, R. Nandi, Artificial neural networks for cognitive radio: a preliminary survey, in: 2012 8th Int. Conf. on Wireless Communications, Networking and Mobile Computing, IEEE, 2012, pp. 1–4.

[14] M. Saber, K. A Hatim, E. R. Abdessamad, S. Rachid, Energy detection and artifcial neural networks for spectrum sensing based on real world signals, Workshop on Advancing Technology for Humanitarian Sciences.

[15] A. Ghasemi, C. Parekh, P. Guinand, Spectrum sensing for modulated radio signals using deep temporal convolutional networks, in: 2019 IEEE Wireless Commun. and Netw. Conf. Workshop (WCNCW), 2019, pp. 1–5.

[16] P. Tilghman, Will rule the airwaves: A DARPA grand challenge seeks autonomous radios to manage the wireless spectrum, IEEE Spectrum 56 (6) (2019) 28–33.

[17] D. A. Guimarães, Digital Transmission: A Simulation-Aided Introduction with VisSim/Comm, Springer-Verlag, 2009.

[18] L. Sachs, Applied Statistics: A Handbook of Techniques, Springer Science & Business Media, 2012.

[19] J. Nolan, Stable distributions: models for heavy-tailed data, Birkhauser New York, 2003.

[20] S. Kotz, S. Nadarajah, Extreme value distributions: theory and applications, Imperial College Press, 2000.

[21] S. Haykin, Neural Networks, Vol. 2, Prentice Hall, 1994.

[22] H.-X. Li, E. S. Lee, Interpolation functions of feedforward neural networks, Computers & Mathematics with Applications 46 (12) (2003) 1861–1874.

[23] A. L. Lemes, MATLAB functions for the GID system-to-distribution parameter mapping via ANNs (Feb 2020).
URL `https://www.dropbox.com/s/169x2wuodmnwgt1/ANNs_GID.zip?dl=0`

Alan Lima Lemes holds an MSc in Telecommunications Engineering from the National Institute of Telecommunications, Inatel, Brazil. He is currently a professor at the Technical School of Electronics, ETE-FMC, Brazil. He is also a Development and Application Engineer at Custom Brazil. His research interests are signal processing, data science and machine learning applied to wireless communications systems.

Dayan Adionel Guimarães holds a PhD in Electrical Engineering from the State University of Campinas, Unicamp, Brazil. He is currently a Senior Lecturer and Researcher with the National Institute of Telecommunications, Inatel, Brazil. His research interests are wireless communications, simulation of communication systems, spectrum sensing for cognitive radio, and convex optimization and signal processing applied to communications.

Yvo Marcelo Chiaradia Masselli holds a PhD in Electrical Engineering from the Federal University of Itajubá, Unifei, Brazil. He is currently a professor of Artificial Intelligence and Postgraduate Coordinator of the National Institute of Telecommunications, Inatel, Brazil. His research interests are in specialized systems and computer vision. He published papers in national and international newspapers related to Automated Systems and Artificial Intelligence.