

Parametrização das distribuições da estatística de teste do detector PRIDe via redes neurais artificiais

Dayan Adionel Guimarães e Alan Lima Lemes

Resumo—No contexto de sensoriamento espectral para redes de rádios cognitivos, este artigo reporta resultados de testes de aderência de distribuições de probabilidade à estatística de teste do detector PRIDe (*Pietra-Ricci index detector*) sob as hipóteses de ausência e presença do sinal primário. Adicionalmente apresenta-se o desenvolvimento de redes neurais artificiais para mapear os parâmetros do sistema de sensoriamento espectral nos parâmetros das distribuições selecionadas. Resultados de simulação são então confrontados com resultados teóricos, demonstrando a adequação dessas distribuições e a acurácia de suas parametrizações.

Palavras-Chave—Rádio cognitivo, redes neurais artificiais, sensoriamento espectral, teste de aderência.

I. INTRODUÇÃO

A grande quantidade de sistemas de comunicações sem fio hoje em operação resulta na escassez de faixas de radiofrequência (RF) livres, configurando um dos principais problemas para o desenvolvimento de novos serviços e sistemas, bem como para a expansão dos existentes. Tal escassez tem como principal causa a política de alocação fixa de espectro atualmente adotada, na qual se destina ao usuário primário (*primary user*, PU), de forma fixa, a fatia do espectro correspondente ao serviço prestado. Se essa política for mantida, o problema tenderá a se agravar com a massificação das redes de quinta geração (*fifth generation*, 5G) e da Internet das coisas (*Internet of things*, IoT), devido ao número sem precedentes de terminais em rede.

Por outro lado, estudos demonstram que a escassez de bandas livres é na verdade relativa, ou seja, grande parte das faixas de frequência já alocadas são subutilizadas, revelando-se livres durante certos intervalos de tempo e em certas localizações geográficas [1].

O conceito de rádio cognitivo (*cognitive radio*, CR) tem sido considerado uma promissora solução do problema supramencionado, à medida que pode viabilizar o uso compartilhado do espectro de RF entre PUs e usuários secundários (*secondary users*, SUs), elevando assim a eficiência de utilização espectral [2]. O processo chave para a realização dessa nova política de alocação dinâmica do espectro é o sensoriamento espectral [3], o qual, como o nome sugere, monitora o espectro em busca de bandas ociosas que possam ser utilizadas de forma oportunista, ou seja, enquanto e onde não estiverem em uso pela rede primária.

Dayan A. Guimarães, Instituto Nacional de Telecomunicações - Inatel, Santa Rita do Sapucaí, MG, e-mail: dayan@inatel.br; Alan L. Lemes, Custom Brasil, Santa Rita do Sapucaí, MG, e-mail: alan.lima@custom.biz. Este trabalho foi parcialmente financiado pela RNP, com recursos do MCTIC, processo No 01250.075413/2018-04, sob o projeto Centro de Referência em Radiocomunicações (CRR) do Inatel.

Dentre as diferentes arquiteturas de implementação do sensoriamento espectral destaca-se o sensoriamento cooperativo (*cooperative spectrum sensing*, CSS), no qual vários SUs colaboram para que a decisão sobre o estado de ocupação da banda sensoriada seja mais acurada em comparação com o sensoriamento individual. No sensoriamento cooperativo centralizado com fusão de amostras, objeto do presente artigo, as amostras do sinal recebido por cada SU são transmitidas a um elemento central da rede secundária denominado centro de fusão (*fusion center*, FC), onde são processadas para formar a estatística de teste das hipóteses \mathcal{H}_0 (sinal primário ausente na banda sensoriada) e \mathcal{H}_1 (sinal primário presente).

A estatística de teste no FC pode ser formada por diferentes técnicas, estando entre as mais conhecidas a detecção de energia, a detecção por filtro casado e a detecção por propriedades cicloestacionárias do sinal sensoriado [3].

Recentemente foi proposto o detector baseado no índice de Pietra-Ricci (*Pietra-Ricci index detector*, PRIDe) [4], considerado um dos mais simples já desenvolvidos e tendo adicionalmente os atributos de taxa de falso alarme constante (*constant false alarm rate*, CFAR), robustez frente a variações de potência de sinal recebido e de ruído, e potencial para suplantarem o desempenho de renomados detectores.

Neste artigo, o detector PRIDe é explorado objetivando sua análise teórica de desempenho, sendo proposta a aplicação de redes neurais artificiais para traduzir os parâmetros do sistema de sensoriamento espectral nos parâmetros das distribuições de probabilidade da estatística de teste, conforme detalhado no restante desta seção.

A. Definição detalhada do problema

O desempenho do sensoriamento é geralmente medido pela probabilidade de falso alarme, P_{fa} , e pela probabilidade de detecção, P_d . A primeira é a probabilidade de ser tomada uma decisão em favor da presença do sinal primário na banda sensoriada, sendo que a banda está desocupada. A segunda é a probabilidade de decidir em favor da presença do sinal primário, dado que, de fato, a banda monitorada está ocupada. Matematicamente tem-se

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

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

em que $\Pr[\cdot]$ denota a probabilidade de ocorrência do evento em questão, T é a estatística de teste, λ é o limiar de decisão, $f(t | \mathcal{H}_0)$ e $F(t | \mathcal{H}_0)$ são, respectivamente, a função

densidade de probabilidade (*probability density function*, PDF) e a função distribuição cumulativa (*cumulative distribution function*, CDF) de T sob \mathcal{H}_0 , e $f(t|\mathcal{H}_1)$ e $F(t|\mathcal{H}_1)$ são respectivamente a PDF e a CDF de T sob \mathcal{H}_1 .

Por meio de (1) e (2) se pode notar que é possível calcular P_{fa} e P_d analiticamente somente se $f(t|\mathcal{H}_0)$ e $f(t|\mathcal{H}_1)$ ou, equivalentemente, $F(t|\mathcal{H}_0)$ e $F(t|\mathcal{H}_1)$ são conhecidas. O cálculo do limiar de decisão λ , o qual é tipicamente configurado para se atingir uma determinada taxa de falso alarme, também necessita do conhecimento de tais funções, pois

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

sendo F^{-1} a função inversa de $F(\lambda|\mathcal{H}_0)$, ou seja, λ é o valor de t tal que $F(t|\mathcal{H}_0) = 1 - P_{fa}$.

Há estatísticas de teste, como é o caso daquela referente ao PRIDE, para as quais é matematicamente intratável determinar de forma analítica suas PDFs ou CDFs. Por outro lado, mesmo que tais funções sejam determinadas, é premente que os parâmetros do sistema de sensoriamento espectral nelas estejam presentes, direta ou indiretamente, para que o desempenho e a configuração do sistema possam ser realizados de acordo com a parametrização de tais distribuições.

B. Contribuições

Devido à impossibilidade de determinar de forma analítica as distribuições da estatística de teste do detector PRIDE sob as hipóteses \mathcal{H}_0 e \mathcal{H}_1 , neste artigo são apresentados resultados de testes de aderência que objetivaram determinar numericamente as distribuições aproximadas dessa estatística, utilizando procedimento similar àquele proposto em [5]. Em seguida apresenta-se o desenvolvimento de redes neurais artificiais (*artificial neural networks*, ANNs) com o objetivo de mapear os parâmetros do sistema de sensoriamento espectral nos parâmetros das distribuições selecionadas. Expressões em forma fechada são então propostas para cálculo de P_d , P_{fa} e λ via solução numérica. Como forma de validação, resultados de simulação computacional são confrontados com resultados teóricos obtidos a partir da aplicação das distribuições parametrizadas de acordo com o mapeamento dado pelas ANNs.

C. Pesquisas relacionadas

A análise matemática de distribuições de estatísticas de teste para sensoriamento espectral pode ser relativamente simples, como é o caso do detector de energia [6], mas pode ser extremamente complexa, situação verificada no caso dos detectores baseados nos autovalores da matriz de covariância do sinal recebido, nos quais aplica-se a teoria de matrizes aleatórias [7], [8], [9]. Em [10], um amplo conjunto de gráficos foi utilizado como forma de mapear os parâmetros sistêmicos do detector GRCR (*Gershgorin radii and centers ratio*) nos parâmetros das distribuições da correspondente estatística de teste.

Em se tratando da aplicação de redes neurais artificiais no amplo contexto das telecomunicações, podem ser citados [11], [12], [13]. Já no contexto de sensoriamento espectral destacam-se [14] e [15], em que ANNs são utilizadas como forma de melhoria do processo. O amplo tema da inteligência artificial também já é explorado no contexto de sistemas que

fazem uso de cognição para otimização do compartilhamento de espectro [16].

Com exceção de [5], onde procedimento similar àquele aqui utilizado é aplicado para se realizar o mapeamento dos parâmetros sistêmicos do detector GID (*Gini index detector*) [17] nos parâmetros das distribuições de sua estatística de teste, não se tem conhecimento do uso de ANNs como solução para tal problema no caso do detector PRIDE.

As demais seções deste artigo estão assim organizadas: o modelo de sistema é descrito na Seção II. A Seção III descreve os procedimentos e apresenta os resultados dos testes de aderência da estatística de teste do detector PRIDE. A Seção IV é dedicada ao procedimento e aos resultados do mapeamento de parâmetros por meio de ANNs. A Seção V apresenta resultados teóricos e de simulação utilizando mapeamentos realizados pelas ANNs desenvolvidas. As principais conclusões da pesquisa são apresentadas na Seção VI.

II. MODELO DE SISTEMA

O sistema CSS consiste de m SUs, cada um coletando n amostras do sinal recebido de p transmissores primários durante cada intervalo de sensoriamento. No FC, tais amostras formam a matriz $\mathbf{Y} \in \mathbb{C}^{m \times n}$ dada por

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{V}. \quad (4)$$

Nessa equação, as amostras associadas aos sinais transmitidos pelos p PUs são arranjadas na matriz $\mathbf{X} \in \mathbb{C}^{p \times n}$. Com o objetivo de modelar as flutuações de envoltória de sinais modulados e filtrados [18], tais amostras são variáveis aleatórias Gaussianas complexas de média zero e variância dependente da razão sinal-ruído (*signal-to-noise ratio*, SNR) média na entrada dos receptores dos SUs.

A matriz de canal $\mathbf{H} \in \mathbb{C}^{m \times p}$ em (4) é formada por elementos h_{ij} , $i = 1, 2, \dots, m$, $j = 1, 2, \dots, p$ que representam os ganhos dos canais entre o j -ésimo PU e o i -ésimo SU. Esses ganhos são constantes durante o intervalo de sensoriamento, sendo independentes e identicamente distribuídos entre consecutivas execuções do sensoriamento. Adicionalmente, a matriz de canal é dada por $\mathbf{H} = \mathbf{G}\mathbf{A}$, em que $\mathbf{A} \in \mathbb{C}^{m \times p}$ tem elementos $a_{ij} \sim \mathcal{CN}[\sqrt{\kappa_{ij}}/(2\kappa_{ij} + 2), 1/(\kappa_{ij} + 1)]$, tais que $\mathbb{E}\{a_{ij}^2\} = 1$, com κ_{ij} sendo o fator de Rice do canal entre o j -ésimo PU e o i -ésimo SU. Assim como em [4], aqui modelam-se canais de sensoriamento realistas em termos do fator de Rice $\kappa_{ij} = 10^{K_{ij}/10}$, sendo $K_{ij} \sim \mathcal{N}[\mu_K, \sigma_K]$, com $\mu_K = 1,88$ dB e $\sigma_K = 4,13$ dB para áreas urbanas, $\mu_K = 2,41$ dB e $\sigma_K = 3,84$ dB para áreas suburbanas, e $\mu_K = 2,63$ dB e $\sigma_K = 3,82$ dB para áreas rurais [19].

Admite-se a possibilidade de níveis desiguais de potência dos sinais recebidos pelos PUs. Nesse caso, $\mathbf{G} \in \mathbb{R}^{m \times m} = \text{diag}(\sqrt{\mathbf{p}/p_{\text{avg}}})$, sendo $\mathbf{p} = [p_1, p_2, \dots, p_m]$ o vetor com as potências dos sinais recebidos pelos SUs, com $p_{\text{avg}} = \frac{1}{m} \sum_{i=1}^m p_i$ denotando a potência média desses sinais. Como o ganho médio de potência do canal é unitário, sem perda de generalidade considera-se que cada PU transmite com uma potência constante p_{avg}/p . Quando as potências dos sinais recebidos são variantes no tempo, tem-se $p_i \sim \mathcal{U}[(1 - \rho)p_{\text{avg}}, (1 + \rho)p_{\text{avg}}]$ em cada execução do sensoriamento, sendo ρ a fração de variação em relação à média.

Analogamente admite-se a possibilidade de variação temporal dos níveis de potência de ruído nos PUs, caso em que a i -ésima linha da matriz $\mathbf{V} \in \mathbb{C}^{m \times n}$ em (4) é composta por variáveis aleatórias Gaussianas de média zero e variância $\sigma_i^2 \sim \mathcal{U}[(1 - \rho/2)\sigma_{\text{avg}}^2, (1 + \rho/2)\sigma_{\text{avg}}^2]$ em cada execução do sensoriamento, com $\sigma_{\text{avg}}^2 = \frac{1}{m} \sum_{i=1}^m \sigma_i^2$ sendo a potência média de ruído nos SUs, e onde arbitrariamente adota-se uma fração de variação das potências de ruído igual à metade daquela referente às potências de sinal. A SNR média nos SUs, em dB, é $\text{SNR} = 10 \log_{10}(p_{\text{avg}}/\sigma_{\text{avg}}^2)$.

De posse de \mathbf{Y} o FC computa a matriz de covariância amostral (*sample covariance matrix*, SCM) de ordem m ,

$$\mathbf{R} = \frac{1}{n} \mathbf{Y} \mathbf{Y}^\dagger, \quad (5)$$

em que \mathbf{Y}^\dagger denota o complexo conjugado e transposto da matriz \mathbf{Y} . Sob a hipótese \mathcal{H}_0 o sinal primário está ausente na banda de interesse, isto é, $\mathbf{Y} = \mathbf{V}$. Sob a hipótese \mathcal{H}_1 o sinal está presente, ou seja, $\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{V}$.

Seja r_i o i -ésimo elemento do vetor formado pelo empilhamento de todas as colunas da matriz \mathbf{R} , e $\bar{r} = (1/m^2) \sum_i r_i$. A estatística de teste do detector PRIDE [4] é dada por

$$T = \frac{\sum_{i=1}^{m^2} |r_i|}{\sum_{i=1}^{m^2} |r_i - \bar{r}|}. \quad (6)$$

Se $T > \lambda$ decide-se por \mathcal{H}_1 como o estado de ocupação do canal sensoriado. Caso contrário decide-se por \mathcal{H}_0 .

III. TESTES DE ADERÊNCIA DA ESTATÍSTICA DE TESTE DO DETECTOR PRIDE

Esta seção descreve, de forma resumida, o procedimento¹ de teste de aderência (*goodness-of-fit*, GoF) da estatística de teste T definida em (6) a várias distribuições de probabilidade. Os seguintes passos resumizam tal procedimento, aplicável tanto a \mathcal{H}_0 quanto a \mathcal{H}_1 :

- 1) Geração de amostras de T ;
- 2) Estimção de parâmetros das distribuições² contínuas operadas pelo MATLAB a partir das amostras de T ;
- 3) Cômputo das métricas de aderência entre as CDFs teóricas candidatas e a CDF empírica de T ;
- 4) Ranqueamento e seleção das melhores distribuições.

Dois métricas de aderência foram utilizadas: o erro quadrático médio (*mean squared error*, MSE) e a média da estatística de Kolmogorov-Smirnov (*mean Kolmogorov-Smirnov statistic*, MKSS) resultante do correspondente teste de aderência [20]. Especificamente, uma amostra com 50000 valores de T foi gerada para cada hipótese, para 1000 combinações dos parâmetros sistêmicos: $p = 1, 2, \dots, 5$ PUs; $m = 3, 4, \dots, 10$ SUs; $n = 20, 40, \dots, 200$ amostras; $\text{SNR} = -15, -14, \dots, -5$ dB sob \mathcal{H}_1 (já que a SNR não influencia T sob \mathcal{H}_0), e um par

¹O procedimento é, em essência, aquele aplicado em [5] no caso do detector GID. Detalhes são aqui omitidos por limitação de espaço, recomendando-se ao leitor consultar [5] para complementação das informações.

²A estimção de máxima verossimilhança de parâmetros foi realizada pela função `fitdist` do MATLAB, versão 2018a, utilizando as seguintes distribuições candidatas: Birnbaum-Saunders, Burr, exponencial, extreme value, gamma, generalized extreme value, generalized Pareto, half-normal, inverse Gaussian, logistic, log-logistic, log-normal, Nakagami, normal, Rayleigh, Rician, stable, t location-scale e Weibull.

(μ_K, σ_K) para cada um dos ambientes analisados (urbano, suburbano e rural), conforme descrito na seção anterior.

Para cada amostra foi realizado o teste de Kolmogorov-Smirnov via função `kstest` do MATLAB, para as distribuições candidatas anteriormente citadas. As métricas MSE e MKSS foram subsequentemente computadas, levando ao ranqueamento apresentado nas Tabelas I e II para as cinco distribuições melhor classificadas, respectivamente para \mathcal{H}_0 e \mathcal{H}_1 . Nota-se nessas tabelas que o ranqueamento das distribuições foi o mesmo no que diz respeito a ambas as métricas.

Na hipótese \mathcal{H}_0 , a distribuição melhor classificada foi a *stable*, sendo a distribuição *generalized extreme value* (GEV) a melhor classificada sob \mathcal{H}_1 . Essas serão, portanto, as distribuições utilizadas para caracterizar a estatística de teste PRIDE deste ponto em diante.

TABELA I
CINCO DISTRIBUIÇÕES MELHOR CLASSIFICADAS SOB \mathcal{H}_0

Distribuição	MSE (ranque)	MKSS (ranque)
Stable	$6,11 \times 10^{-5}$ (1)	$1,29 \times 10^{-2}$ (1)
Generalized extreme value	$8,71 \times 10^{-5}$ (2)	$1,38 \times 10^{-2}$ (2)
Burr	$1,10 \times 10^{-4}$ (3)	$1,71 \times 10^{-2}$ (3)
Loglogistic	$2,07 \times 10^{-4}$ (4)	$2,24 \times 10^{-2}$ (4)
t Location scale	$2,57 \times 10^{-4}$ (5)	$2,50 \times 10^{-2}$ (5)

Os parâmetros da distribuição *stable* são: α ($0 < \alpha \leq 2$) e β ($-1 \leq \beta \leq 1$) determinam seu formato, γ ($0 < \gamma < \infty$) é o fator de escala e δ ($-\infty < \delta < \infty$) determina sua localização. A PDF e a CDF *stable* não podem ser expressas analiticamente, exceto para certos valores de seus parâmetros [21]. No caso da distribuição GEV tem-se: μ ($-\infty < \mu < \infty$) determina sua localização, σ ($\sigma > 0$) é o fator de escala e k ($-\infty < k < \infty$) governa o seu formato. Tanto a PDF quanto a CDF GEV têm representação matemática consideravelmente simples [22].

As distribuições *stable* e GEV estão incluídas na lista de funções cujos cálculos numéricos se tornaram suficientemente maduros a ponto de permitir que uma expressão que contenha essas funções possa ser considerada forma fechada. A CDF e a CDF inversa dessas distribuições estão incorporadas ao ambiente MATLAB, seguindo as sintaxes

```

cdf('Stable', alpha, beta, gamma, delta),
cdf('GeneralizedExtremeValue', lambda, k, sigma, mu),
icdf('Stable', P, alpha, beta, gamma, delta),
icdf('GeneralizedExtremeValue', P, lambda, k, sigma, mu),
    
```

sendo P a probabilidade de referência para o cálculo da CDF inversa. No presente contexto, $P = 1 - P_{\text{fa}}$, conforme (3).

TABELA II
CINCO DISTRIBUIÇÕES MELHOR CLASSIFICADAS SOB \mathcal{H}_1

Distribuição	MSE (ranque)	MKSS (ranque)
Generalized extreme value	$3,89 \times 10^{-5}$ (1)	$9,39 \times 10^{-3}$ (1)
Stable	$1,47 \times 10^{-4}$ (2)	$2,01 \times 10^{-2}$ (2)
Burr	$1,84 \times 10^{-4}$ (3)	$2,31 \times 10^{-2}$ (3)
Loglogistic	$4,59 \times 10^{-4}$ (4)	$3,49 \times 10^{-2}$ (4)
t Location scale	$6,81 \times 10^{-4}$ (5)	$4,45 \times 10^{-2}$ (5)

As Figs. 1 e 2 apresentam CDFs empíricas (obtidas por simulação) e teóricas (obtidas pelas correspondentes funções

embutidas no MATLAB e parametrizadas de acordo com o critério de máxima verossimilhança adotado na função `fitdist`), sob as hipóteses \mathcal{H}_0 e \mathcal{H}_1 , respectivamente, para três cenários escolhidos como representativos. Nota-se que as CDFs teóricas e empíricas praticamente se sobrepõem em todos os casos analisados, uma primeira demonstração da adequação das distribuições selecionadas. Resultados adicionais de validação são apresentados na Seção V.

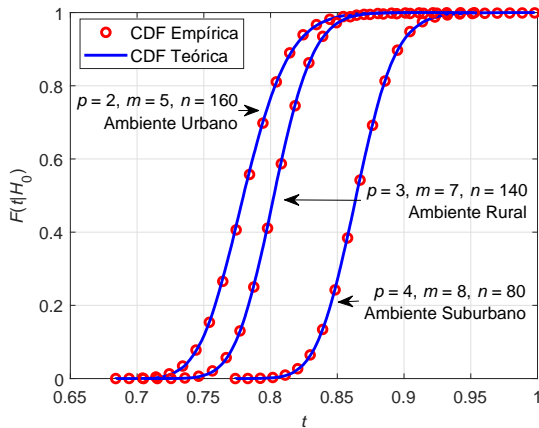


Fig. 1. CDFs empíricas e teóricas sob \mathcal{H}_0 .

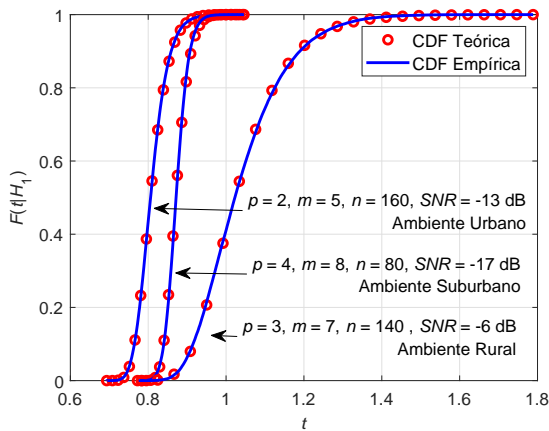


Fig. 2. CDFs empírica e teóricas sob \mathcal{H}_1 .

IV. MAPEAMENTO DE PARÂMETROS POR MEIO DE REDES NEURAIS ARTIFICIAIS

Esta seção descreve resumidamente o procedimento utilizado para geração do conjunto de treinamento e o dimensionamento das ANNs. Sendo tal procedimento similar àquele adotado em [5], omitem-se aqui os detalhes que podem ser consultados nessa referência.

Foram gerados três conjuntos de treinamento, cada um utilizando um par (μ_K, σ_K) adequado para representar ambiente urbano, suburbano e rural, conforme Seção II. Nos três ambientes os parâmetros sistêmicos utilizados foram: $p = 1, 2, \dots, 4$ PUs; $m = 3, 4, \dots, 10$ SUs; $n = 20, 40, \dots, 200$ amostras e $\text{SNR} = -17, -15, \dots, -5$ dB, totalizando 2240 combinações. Para cada combinação, 50000 amostras da estatística de teste

T foram geradas sob \mathcal{H}_0 e \mathcal{H}_1 . O procedimento a seguir foi repetido para todas as combinações dos parâmetros sistêmicos:

- 1) Geração dos parâmetros sistêmicos;
- 2) Geração de amostras de T de acordo com (6);
- 3) Estimação de parâmetros das distribuições utilizando a função `fitdist` do MATLAB para as distribuições *stable* sob \mathcal{H}_0 e GEV sob \mathcal{H}_1 ;
- 4) Armazenamento dos parâmetros sistêmicos e das CDFs para composição do conjunto de treinamento das ANNs.

Quatro ANNs do tipo MLP (*multilayer perceptron*) foram treinadas para mapear os parâmetros das CDFs nos parâmetros sistêmicos do CSS. Adotou-se uma camada oculta com vinte e cinco neurônios, função de ativação do tipo tangente hiperbólica e otimizador Levenberg–Marquardt. A ANN para mapeamento sob \mathcal{H}_0 possui como entrada os parâmetros p , m e n , por ser independente do valor da SNR e da matriz de canal, sendo a mesma para todos ambientes. Sob \mathcal{H}_1 foram treinadas três ANNs, uma para cada ambiente (urbano, suburbano, rural), tendo como entradas p , m , n e SNR.

As ANNs desenvolvidas foram traduzidas nos códigos `NnH0.m`, `NnH1u.m`, `NnH1s.m` e `NnH1r.m` do MATLAB, os quais estão disponíveis em [23]. Esses códigos devem estar no mesmo diretório da rotina que for utilizá-los e não necessitam do módulo *machine learning toolbox*. A função `NnH0([p m n])` retorna um vetor com os parâmetros α , γ e δ da distribuição *stable*, nesta ordem (o parâmetro β é sempre unitário), em função dos parâmetros sistêmicos p , m e n . As funções `NnH1u([p m n SNR])`, `NnH1s([p m n SNR])` e `NnH1r([p m n SNR])` retornam vetores com os parâmetros k , σ e μ da distribuição GEV, nesta ordem, em função dos parâmetros sistêmicos p , m , n e SNR.

V. ANÁLISE DE DESEMPENHO

Para a análise de desempenho de uma técnica de sensoriamento espectral, a probabilidade de detecção, P_d , e a probabilidade de falso alarm, P_{fa} , são comumente avaliadas por meio da curva característica de operação do receptor (*receiver operating characteristic, ROC*), na qual tem-se P_d versus P_{fa} de acordo com a variação do limiar de detecção λ . Nos resultados apresentados a seguir, cada ponto sobre uma curva ROC simulada foi gerado a partir de 50000 eventos de Monte Carlo, nos quais a estatística de teste PRIDE foi gerada para determinados conjuntos de parâmetros sistêmicos. Estes conjuntos foram então aplicados às ANNs descritas na seção anterior, permitindo determinar os parâmetros das distribuições da estatística de teste e calcular o desempenho teórico do sensoriamento espectral aplicando as expressões

$$P_{fa} = 1 - F_S(\lambda, \alpha, 1, \gamma, \delta), \quad (7)$$

em que $F_S(\lambda, \alpha, 1, \gamma, \delta)$ é o valor da CDF *stable* parametrizada por $\alpha, \beta = 1, \gamma, \delta$ no ponto $t = \lambda$, e

$$P_d = 1 - F_G(\lambda, k, \sigma, \mu), \quad (8)$$

em que $F_G(\lambda, k, \sigma, \mu)$ é o valor da CDF GEV parametrizada por λ, k, σ, μ no ponto $t = \lambda$, com o limiar de decisão calculado de acordo com a probabilidade de falso alarme por meio de

$$\lambda = F_S^{-1}(1 - P_{fa}, \alpha, 1, \gamma, \delta), \quad (9)$$

sendo $F_S^{-1}(1-P_{fa}, \alpha, 1, \gamma, \delta)$ a CDF *stable* inversa em $1-P_{fa}$.

É importante ressaltar que (7), (8) e (9) podem ser consideradas expressões em forma fechada, pois utilizam funções operadas por algoritmos numericamente estáveis e disponíveis em várias ferramentas computacionais de cálculo. No ambiente MATLAB, tais expressões são escritas aplicando as sintaxes

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

$$P_d = 1 - \text{cdf}('GeneralizedExtremeValue', \lambda, k, \sigma, \mu), \quad (11)$$

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

Objetivando verificar a acurácia do mapeamento de parâmetros sistêmicos em parâmetros das distribuições por meio das ANNs desenvolvidas, a Fig. 3 apresenta resultados de desempenho teóricos e simulados do detector PRIDE, para parâmetros sistêmicos arbitrariamente escolhidos, mas representativos de situações práticas comuns. Pode-se notar que as ROCs simuladas praticamente se sobrepõem às correspondentes ROCs teóricas, mais uma vez demonstrando que as distribuições *stable* e GEV representam satisfatoriamente a estatística de teste PRIDE sob as hipóteses \mathcal{H}_0 e \mathcal{H}_1 , respectivamente. Adicionalmente, conclui-se que as ANNs desenvolvidas são capazes de mapear os parâmetros de tais distribuições com precisão, permitindo igualmente acuradas análises teóricas de desempenho do detector PRIDE.

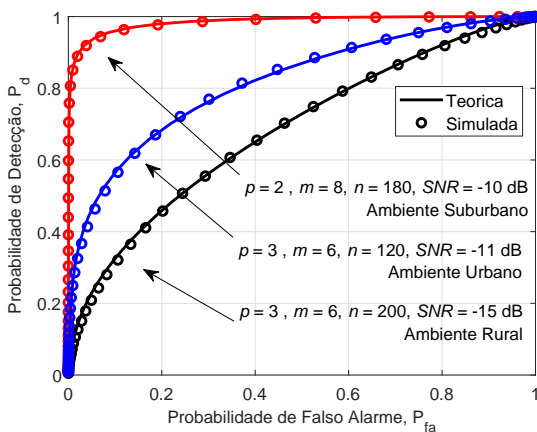


Fig. 3. Desempenhos teóricos e simulados do detector PRIDE.

Vale ressaltar que é possível mapear parâmetros sistêmicos diferentes daqueles utilizados no treinamento das ANNs, graças à inerente capacidade interpoladora de ANNs do tipo MLP. Além disso, é também possível extrapolar valores de parâmetros sistêmicos além das faixas de variação escolhidas para treinamento [5].

VI. CONCLUSÕES

Este artigo apresentou resultados de testes de aderência de distribuições de probabilidade à estatística de teste do detector PRIDE, sob as hipóteses \mathcal{H}_0 e \mathcal{H}_1 , bem como o desenvolvimento de quatro ANNs para mapear os parâmetros do sistema de sensoriamento espectral nos parâmetros das distribuições selecionadas. Resultados de simulação foram confrontados com resultados teóricos, demonstrando a adequação dessas distribuições e a acurácia de suas parametrizações.

REFERÊNCIAS

- [1] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "Next generation/dynamic spectrum access/cognitive radio wireless networks: A survey," *Computer networks*, vol. 50, no. 13, pp. 2127–2159, 2006.
- [2] F. Hu, B. Chen, and K. Zhu, "Full spectrum sharing in cognitive radio networks toward 5G: A survey," *IEEE Access*, vol. 6, pp. 15 754–15 776, 2018.
- [3] Y. Arjoun and N. Kaabouch, "A comprehensive survey on spectrum sensing in cognitive radio networks: Recent advances, new challenges, and future research directions," *Sensors*, vol. 19, no. 1, 2019.
- [4] D. A. Guimarães, "Pietra-Ricci index detector for centralized data fusion cooperative spectrum sensing," *IEEE Trans. Veh. Technol.*, 2020, doi: 10.1109/TVT.2020.3009440. [Online]. Available: <https://ieeexplore.ieee.org/document/9141423>
- [5] A. L. Lemes, D. A. Guimarães, and Y. M. C. Masselli, "System-to-distribution parameter mapping for the Gini index detector test statistic via artificial neural networks," *Computers & Electrical Engineering*, vol. 85, p. 106692, 2020.
- [6] F. F. Digham, M.-S. Alouini, and M. K. Simon, "On the energy detection of unknown signals over fading channels," *IEEE Trans. Commun.*, vol. 55, no. 1, pp. 21–24, 2007.
- [7] L. Huang, Y. Xiao, H. C. So, and J. Fang, "Accurate performance analysis of hadamard ratio test for robust spectrum sensing," *IEEE Trans. Wireless Commun.*, vol. 14, no. 2, pp. 750–758, 2015.
- [8] L. Wei and O. Tirkkonen, "Spectrum sensing in the presence of multiple primary users," *IEEE Trans. Commun.*, vol. 60, no. 5, pp. 1268–1277, 2012.
- [9] L. Huang, C. Qian, Y. Xiao, and K. Q. Zhang, "Performance analysis of volume-based spectrum sensing for cognitive radio," *IEEE Trans. Wireless Commun.*, vol. 14, no. 1, pp. 317–330, 2015.
- [10] A. L. Lemes and D. A. Guimarães, "Estimação das distribuições da estatística de teste GRCR sob as hipóteses \mathcal{H}_0 e \mathcal{H}_1 ," in *XXXVI Simpósio Brasileiro de Telecomunicações e Processamento de Sinais, SBRT'18*, Sep. 2018, Campina Grande, PB, Brasil.
- [11] N. Baldo and 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, and B. Noursain, "Deep learning for RF device fingerprinting in cognitive communication networks," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 1, pp. 160–167, 2018.
- [13] S. Pattanayak, P. Venkateswaran, and 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, and S. Rachid, "Energy detection and artificial neural networks for spectrum sensing based on real world signals," *Workshop on Advancing Technology for Humanitarian Sciences*, 07 2018.
- [15] A. Ghasemi, C. Parekh, and P. Guinand, "Spectrum sensing for modulated radio signals using deep temporal convolutional networks," in *2019 IEEE Wireless Commun. and Netw. Conf. Workshop (WCNCW)*, Apr 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*, vol. 56, no. 6, pp. 28–33, June 2019.
- [17] D. A. Guimarães, "Gini index inspired robust detector for spectrum sensing over Ricean channels," *Electronics Letters*, vol. 55, no. 12, pp. 713–714, 2019.
- [18] B. Nadler, F. Penna, and R. Garello, "Performance of eigenvalue-based signal detectors with known and unknown noise level," in *IEEE Int. Conf. Commun.*, Jun 2011, pp. 1–5.
- [19] S. Zhu, T. S. Ghazaany, S. M. R. Jones, R. A. Abd-Alhameed, J. M. Noras, T. Van Buren, J. Wilson, T. Suggett, and S. Marker, "Probability distribution of Rician K -factor in urban, suburban and rural areas using real-world captured data," *IEEE Trans. Antennas Propag.*, vol. 62, no. 7, pp. 3835–3839, Jul 2014.
- [20] L. Sachs, *Applied Statistics: A Handbook of Techniques*. Springer Science & Business Media, 2012.
- [21] J. Nolan, *Stable distributions: models for heavy-tailed data*. Birkhauser New York, 2003.
- [22] S. Kotz and S. Nadarajah, *Extreme value distributions: theory and applications*. Imperial College Press, 2000.
- [23] A. L. Lemes, "Funções do MATLAB para mapeamento de parâmetros do detector PRIDE via ANNs," Mar 2020. [Online]. Available: <https://www.dropbox.com/s/k8a264mnc903qv/ANNs.zip?dl=0>