Digital TV Channel Prediction Using Clustering Algorithms and Statistical Learning

Daniel C. Vidal, Pedro V. G. Castellanos and Tadeu N. Ferreira

Abstract—**Due to the rise of new communication services, portions of the electromagnetic spectrum must be relocated and their distribution optimized. With the digitization of the open TV service, the distribution of channels in the frequency band destined for this service generates an inefficient use of the radio spectrum. These unused frequency bands are denominated void spaces. To establish efficient spectrum use, it is important to identify these spectrum gaps and use it according to some certain criteria. In this article, machine learning algorithms are proposed to identify new spectrum opportunities, through the signal levels received in the UHF frequency range of the Digital TV system. These spectrum opportunities are generated from natural or artificial obstacles present in the propagation environment. Two measurement campaigns were carried out in a suburban environment to obtain the level of the received signal in an area of approximately 240,000 square meters. From the received power values, machine learning algorithms were used to make prediction of the received signal levels. By using a reception threshold, it is possible to identify the shadow regions and the availability of spectrum opportunities.**

Index Terms—**Machine Learning, Spectrum Occupancy, TV White Space, SBTVD.**

I. Introduction

DUE to the evolution of communication technology and the search for new communication services, the demand for wider frequency bands is progressively increasing. The frequency spectrum is a finite and scarce resource, which disables the rapid inclusion of new wireless communication services or the expansion of an existing service due to current spectrum allocation policies. The spectrum policy is statically defined for each type of service, i.e., a portion of the spectrum band is intended solely and exclusively for a service. Several authors [1], [2] have carried out research on the use of the spectrum in the current model which is not efficient. Thus, the current scenario of spectrum usage needs a new efficient spectrum allocation paradigm.

The coverage area of a group of broadcasting stations is the area where the desired field strength is greater than or equal to the minimum field strength necessary to permit a pre-defined reception quality. This reception quality is pre-defined for a specified reception conditions and for an envisaged percentage of the covered receiving locations [3]. According to the Brazilian Association of Technical Standards (ABNT), the threshold to insure the minimal quality of reception in a Digital Terrestrial Television (DTT) receiver is -77 dBm [4]. Some of the measured locations have a direct line-of-sight to the site of the antennas from the TV channels. Then, it was possible to determine the portion of the area which was receiving the DTT signal according to the ABNT specifications, and if the levels received in that specific area were actually within the threshold.

Previous prediction models, such as the ITU-R P.1546 recommendation [5], use statistical models based on data which are specific to the regions where they were collected. In order to obtain a good result with this recommendation, the analyzed area must have certain similarities with those used to generate conventional models. With the current generation of digital TV systems, changes in the transmitted signal are more noticeable when the aforementioned statistical propagation models are used. This observation raises the question of what can be done to try to minimize the effects that degrade the signal, and to some extent, to improve the results of the propagation models.

There are several models for propagation path loss estimation in the literature for outdoor environments. To illustrate that, Suh *et al.* [6] proposed a methodology for measurements on field strength

Daniel Vidal is with the Universidade de São Paulo, São Paulo, SP,Brasil e-mail:danielvidal@usp.br

Pedro Castellanos and Tadeu Ferreira are with Universidade Federal Fluminense, Niterói, RJ, Brasil e-mails:{pcastellanos, tadeu_ferreira}@id.uff.br

The authors thank CAPES Code 0001 for the financial support of the project.

Digital Object Identifier: 10.14209/jcis.2022.6

predictions and compared them to well-known models, such as Okumura-Hata, on suburban areas for 1 kW effective radiated power in the 30 MHz - 3000 MHz frequency range.

Silva *et al.* [7] made a comparison with predicted values for field strength measurements of DTV in five different Brazilian cities using models such as: ITU-R Recommendation P.525, Deygout-Assis Knife Edge, Deygout-Assis Rounded, Deygout-Assis Main Rounded, CRC-Predict, and ITU-R Recommendation P.1546. The paper pointed out that the ITU-R P.1546 is the model that best fit the field measurements by far, in the cities of São Paulo, Recife and Brasília. They concluded that more investigations are needed to have more accurate results and that the urban environment needs to be taken in consideration due to the chances in the environments morphology.

The studies mentioned above use traditional statistical methods, which require a detailed knowledge of the environment where they will be carried out. On the other hand, our work is based on geographic data and measurements of the field strength, in an attempt to decrease the amount of information needed to make predictions. Our approach makes use of a well-known, low-complexity and comprehensive machine learning algorithms such as K-Nearest Neighbors (KNN), Random Forest Regressor (RFR) and Support Vector Regressor (SVR), in order to predict the presence and the location of the usable channels in the UHF band. We also used the K-Means algorithm to group the measurement point by similarities in the latitude, longitude and altitude, which is also necessary due to the grid-like nature of the measurement campaign.

This article proposes an approach using machine learning algorithms for resource allocation of the electromagnetic spectrum, more specifically in the UHF band, exploring the white spaces and the shadow areas, where the received signal is very low or practically nonexistent. This could facilitate other services to use this frequencies to enhance its services, as mobile carriers could use this zones of shadow to improve the coverage of mobile signals. Data were obtained by measurement campaigns that were carried out in a suburban area, where the transmitted signals pass through different types of terrain until it reaches the receiver. The transmitters are located 700 m above sea level and approximately 14 km away from the measurement area.

Section II presents some propagation aspects of DTV. In Section IV, the machine learning algorithms used in this article are presented. Section V shows the results, while Section VI concludes this article.

II. Digital Terrestrial Television

The terrestrial broadcasting services use a large part of the frequency spectrum, comprising the bands from 54 MHz to 216 MHz in the VHF band for channels 2 to 13 and from 470 MHz to 806 MHz in the UHF band for channels 14 to 69. One of the main characteristics of digital TV systems is the possibility of having different transmission configurations, which allows to adjust the signal according to the propagation conditions and data rate. Through BST-OFDM (Band-Segmented Transmission Orthogonal Frequency-Division Multiplexing) modulation, it is possible to transmit up to 4 channels on a single channel of approximately 6 MHz, increasing the diversity of multimedia features on a single 6 MHz channel. In communication system planning, several aspects must be taken into account, including the coverage area interface analysis and channel allocation [8]. The coverage area of a terrestrial TV is subdivided into units of area known as pixels, which are usually 100 m by 100 m in size. Within the coverage area, the availability of the TV signal for each pixel varies according to the propagation conditions (received power), noise in the band and interference condition. This availability is defined through the Location Probability (LP) parameter [8].

LP is widely used in DTT networks planning, its values vary from 100% in points close to the transmitter to 0% in points outside the coverage area, in reference to the ideal model. The target values of 90% of the time in 50% of the location are defined as an obligation to be reached in a given coverage area [9]. This value is used to define the system's protection area, where the service must be provided by obligation. The LP calculation considers the statistical variation of the signal within the pixel and then evaluating which fraction of pixel has a Carrier-to-Interference-plus-Noise ratio ($C/(I+N)$), where C is the power of the incoming signal of interest, I is the power of the interference signal and N is the noise power. The value of $C/(I+N)$ should be sufficient to support the demodulation of a Digital

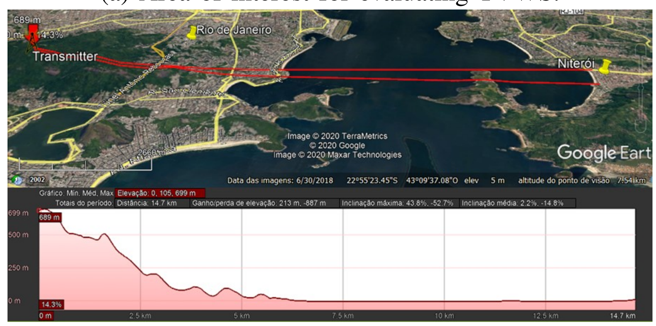
Television (DTV) system. In other words, the LP is determined by $C/(I+N)$. For the actual operating configuration where the LTE (Long Term Evolution) signal is close to the DTV, the interference levels may increase, thus increasing the number of pixel in an unavailability condition. In reality, the pixels in a transmission network are degraded by both noise and interference from other transmission networks where there is a frequency reuse (Single Frequency Network – SFN). They are also degraded by interference from systems operating at adjacent frequencies.

A. *ITU-R P.1546 Recommendation*

The ITU-R P.1546 Recommendation is a standard defined by the International Telecommunication Union (ITU) which consists of a set of recommendations for the calculation of prediction of radio coverage of terrestrial links, in the frequency range of 30 MHz to 3000 MHz, in point-to-area predictions, within the distances of 1 km and 1000 km [6] [10]. The recommendation is valid for some types of routes, such as: land routes, sea routes and mixed (land and sea) routes. It is important to note that, for mixed routes, between land and river regions, river routes must be considered as land routes.

The results presented in this recommendation are based on values of regions where measurements were performed. The measurements were carried out in temperate climate environments, such as in North America and Europe. For maritime routes, the Mediterranean Sea and North Sea regions were considered, that is, respectively regions of hot and cold seas. In regions of maritime routes in hot regions, the propagation characteristics are different. This modeling is considered empirical or semi-empirical. The calculations consist of electric field intensity curves and tabs, obtained from measurements on site, and these values are presented in curves, based on an effective radiated power of 1 kW, for nominal frequency values in 100, 600 and 2000 MHz and values of 1%, 10% and 50%, which represent the percentage of time exceeded for attendance. These parameters are based on the link distance and the antennas height. Typically, the values shown on the curves apply to areas with size 200 m by 200 m.

(a) Area of interest for evaluating TVWS.



(b) Different paths of propagation.

Fig. 1: Representation of different paths and the area of interest.

III. Collected Data

A. *Measurement Campaign*

In this work, a measurement campaign was performed aiming the DTV signal availability using a received signal power in an area of approximately 240,000 square meters, as shown in Fig.1a. The area of interest and the transmitter antennas are on opposite sides of the Guanabara Bay, over 14 km apart, as shown in Fig.1b. This is an interesting characteristic of this work, since the transmitted signal has to travel over different paths, over both land and water. We use as transmitted signals the actual broadcasted signals by the local television operators. They are all located at the same geographical point, at 700 meters above sea level.

The results presented in this campaign were the reception levels in the UHF frequency range of DTV with a frequency range from 470 MHz to 700 MHz. In order to not lose resolution, the band was divided into segments of approximately 57 MHz. The data was collected through Anritsu's MS2034A spectrum analyzer, using a of 10 kHz Resolution Bandwidth (RBW). The spectrum analyzer power provides us with 551 measurement points for each scan, where each point provides the received power value for each 103.45 kHz band. To obtain the power received on a 6 MHz DTV channel, equation (1) was used.

$$P = 10log_{10}\left[B_s\left(\frac{\frac{1}{n}\sum_{i=1}^{n}10^{\frac{P(i)}{10}}}{NBW}\right)\right] \qquad (1)$$

where *P* is the total power in the channel width, B_s is the channel width, *NBW* is the equivalent analyzer noise bandwidth, *n* is the number of sample points and $P(i)$ is the power read on the spectrum analyzer [11].

IV. Machine Learning Algorithms for Clustering and Regression

In order to estimate the received signal levels and thus detect the presence or absence of the signal from a DTV operator, the data collected from the measurements were processed and the received power value for each channel present in the measurement locations was calculated. A preliminary analysis using the prediction model ITU-R P.1546 was performed. Due to the power variability of the received signal at each measurement point, the model prediction does not present an adequate approximation to the collected data. As described in section I, the objective of this work is the use of machine learning algorithms to estimate the received channel power with some type of pre-classification in the data. Those models are compared with traditional prediction models. Based on this principle, there is a manner to improve the accuracy of the response of the prediction algorithm, which is by grouping the data according to common characteristics. The chosen characteristics are the conditions of inspection at the measuring points, that is, the coordinates that are often used to determine a geographical point, such as latitude, longitude and altitude. In addition, some machine learning techniques and algorithms used in forecasting will be presented in this article.

A. Clustering using K-Means

From the collected measurements, we propose to divide the measurement points into groups based on their geographic characteristics, such as latitude, longitude and altitude, in a way that a regression algorithm can achieve a better estimation of the area of interest. Clustering techniques [12] seems to adequately group the points together. The K-Means algorithm was chosen because of its simplicity. This algorithm is widely used for clustering points spread in a Euclidean space and it has a fast convergence time.

K-Means is an unsupervised algorithm that classifies data into *k* different groups by an iterative method, so the results generated are compact and independent groups of samples [13]. In order to achieve a correct behavior of the K-Means algorithm we transformed the latitude and longitude coordinates (spherical coordinates) into a Universal Transverse Mercator (UTM) coordinate system, which provides a projection of the GPS coordinates over a 2-dimensional Cartesian coordinate system for a correct operation.

One main issue in clustering is to define an optimal number of clusters for a data-set. For this purpose, we applied the elbow method [14], which consists of executing the K-Means algorithm over the data-set for a range of *k* clusters. This is a heuristic method, which involves graphing the variation explained according to the number of clusters and choosing the elbow of the curve as the number of clusters to be used. Then for each candidate number of clusters it evaluates the Within Cluster Sum of Squares (WCSS), which is the sum of the square distance of the samples to their closest cluster center. Suppose that a given set of clusters $C = (C_1, \ldots, C_n)$, whose centers are $(\phi_1, \ldots, \phi_n)$. Therefore, the WCSS can be expressed as in equation (2).

$$WCSS = \sum_{i=1}^{n}\sum_{x\in C_i}\|x-\phi_i\|^2 \qquad (2)$$

After running the test to select the value of *k*, in this case $k = 5$, the measurement points were assigned to their respective groups and the model was trained in independent instances for each created group, applying the regression algorithms that were selected for this work, as shown in Fig. 2.

B. Random Forest Regressor

Random Forest algorithms are an ensemble method of classification or regression, by cronstructing a multitude of Decision Trees, that is a simple algorithm which rarely produces results as

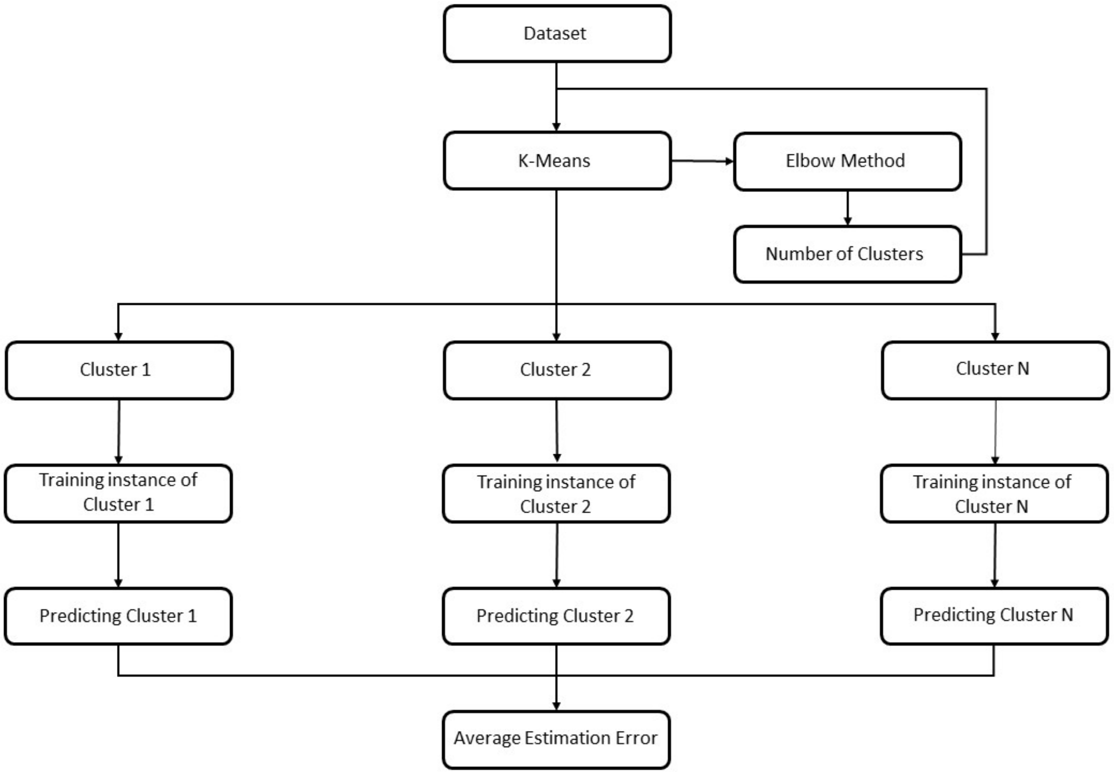


Fig. 2: Method of Evaluation of the K-Means algorithm.

satisfactory as the other supervised methods used in this work. Therefore, the Random Forest Regressor (RFR) algorithm was chosen in this work, which consists of an ensemble of decision trees that are then combined to produce a consensus on the forecast.

In the RFR algorithm, a number of decision trees are built on a bootstrap training samples. But in the process of building these decision trees, each time a split in a tree is considered, a random sample of *m* predictors is chosen as split candidate from the complete set of *p* predictors [15]. This method does not allow the trees to consider the majority of the available predictors. Assume that one predictor is stronger than the others, this means that each decision tree chooses the same predictor at the top split. By not allowing the trees to choose from the full set of predictors, those trees will not be highly correlated on average.

In the training phase, let $D = (x_i, L_i)$ a data set containing the vector $x_i \in \mathbb{R}^3, i \in 1, 2, \ldots, q$, *q* being the number of samples, x_i a vector composed of latitude, longitude and altitude, and $L_i \in \mathbb{R}$ which is the propagation loss.

The Random Forest divides the data set *D* into *T* bootstraps, which is a tool used to quantify the uncertainty with a given estimator and widely used when it is very difficult to calculate the standard deviation, $D_S \subseteq D$ and each one is inserted into the root node of a regression tree. Let D_p be a bootstrap subset of D_S on node *p*. A subset of samples is selected at random on each split node to develop the binary test, t_c, where $x \in D_p$. The binary test is shown by Equation (3).

$$t_{c,\tau}(x) = \begin{cases} 1 & \text{if } x^c > \tau \\ 0 & otherwise \end{cases} \quad (3)$$

The test that has the lowest mean Square Error (MSE) value is selected to move forward on the tree, and it is kept constant as the tree grows. The RFR prediction results correspond to the average of the *T* regression trees [16] following the Equation (4).

$$L_{pred}(x) = \frac{1}{T} \sum_{i=1}^{T} \hat{L}_i(x) \quad (4)$$

C. *Support Vector Regressor*

The Suopport Vector Regressor (SVR) is an extension of the Support Vector Machine (SVM), which constructs a hyperplane, or a set of hyperplanes, in a high dimensional-space. The main idea behind the SVM is to non-linearly map a data set contained in a finite-dimensional space to a larger space, so that the data set is linearly separable. The SVR maps the samples of a hyper-plane of high dimension which can be described by Equation (5):

$$f(x) = w^T \phi(x) + b \tag{5}$$

where w is a normal vector that determines the hyper-plane directions, x is the entry vector, $\phi(.)$ is the nonlinear map function and b is the displacement.

The solution to equation (5) needs a kernel function, which is the key to the performance of the SVR-based predictor. In this work, we choose the Radial Basis Function (RBF) that suits it better for low-dimension tasks and for lacking of previous knowledge [16] [17].

D. K-Nearest Neighbors Regression

The K-Nearest Neighbors (KNN) algorithm finds a group of *k* neighboring points that are the closest in distance to the point *x*, which is being investigated, thus making a comparison with the *y* values assigned to each of the *k* neighbors [15]. The associated response of the training point is compared through a consensus of the *k* associated neighbour responses.

Entry: *D*, the set of *k* training samples and a test set $z = (x', y')$.
Training: Calculate $d(x', x)$, which is the distance between z and all the samples $(x, y) \in D$.
Output: $y' = \max_v \sum_{(x_i,y_i)\in D_z} I(v = y_i)$.

Given a training set *D* and a test point $z = (x', y')$, the algorithm calculates the distance between z and all other training points $(x', y') \in D$, to determine the list of nearest neighbors. Once the list of closest neighbors is completed, the test points have the final response value calculated by averaging their neighbors, as shown by Equation (6).

$$\textbf{Mean} : y' = \frac{1}{K} \sum_{x_i \in D_z} y_i \tag{6}$$

V. Numerical Results

The power received on a 6 MHz channel of the digital television system, collected from 551 points, was used to train the machine learning algorithms mentioned in the previous section. The SVR processing is made based on the frequency 533 MHz, that is, channel 24, in order to compare it with the other channels in the UHF band. After generating the parameters of the algorithm, they were applied to the measured data of channel 29 (563 MHz) to evaluate its accuracy. Firstly, the elbow method [14] was applied to select the number of groups to be executed by the K-Means algorithm. To apply this method, it was identified that the ideal number of groups would be 5, as it has a low WCSS value and does not represent a high value of *k*. Next, the data were divided into their designated group and thus were ready for simulations with the other algorithms.

We selected two error metrics, Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), which are explained below, to evaluate the estimation error in the considered algorithms.

- RMSE indicates the error through the square root of the squared differences of $\hat{y}$, which is the predicted value of the loss on the path, and the observation of the real value of the loss y. This metric is defined by equation (7).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \tag{7}$$

- MAE indicates the average error of the predicted values by calculating the absolute value of the difference between the predicted measurement $\hat{y}$ and the corresponding true value y. This metric is defined by equation (8):

$$MAE = \frac{1}{n}\sum_{i=0}^{n} |y_i - \hat{y}_i| \tag{8}$$

Processing was performed by dividing the groups into two subgroups, one for training with 70% of the data from the original group and another group for testing with the remaining 30% of the data. In addition, in all algorithms the value 42 was used as the value of random state. As the first simulation, the RFR algorithm was used with 30 trees and with the MSE decision criterion, Simulations with the SVR algorithm were performed, after an exhaustive search through a grid search, using the RBF function as the chosen function for the kernel and with gamma and error penalty equal to 1. Finally, the simulation of the KNN algorithm was performed with the value $k = 7$, which is the number of neighbors used to determine the final value.

After regarding the evaluation of the MAE, depicted in Figure 3 and Table 1, the best performance is given by the RFR algorithm with an average value of 2.22, followed by KNN equal to 5.20 and SVR equal to 5.63. We can observe that these new methods have a superior performance to the ITU recommendation model, as seen in Table 1. All other methods obtained lower values in both criteria adopted for the error. By observing the RMSE we can also see that the performance of the RFR algorithm was superior when compared to all others, with a value equal to 3.19, followed by KNN and SVR with 6.66 and 7.19, respectively. In all comparisons, the model presented by ITU is inferior to all the models tested in this work.

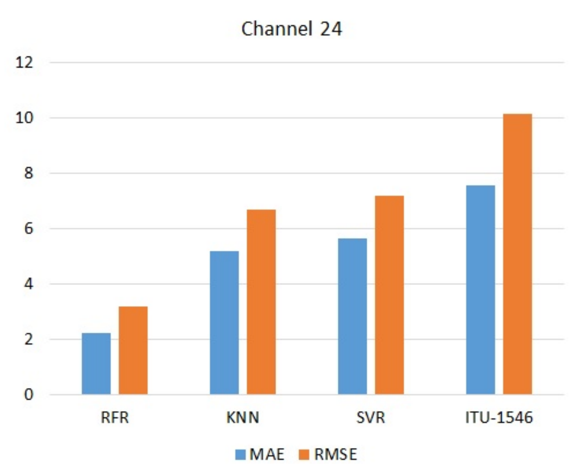


Fig. 3: Performance comparison for Machine Learning algorithms on Channel 24.

TABLE I: Algorithm comparison for MAE.

Channel	RFR	KNN	SVR	ITU-1546
21	3.91	5.63	5.47	10.76
22	4.84	6.61	6.68	8.37
24	2.22	5.20	5.63	7.54
27	4.02	5.59	6.28	12.53
29	5.08	7.10	7.18	8.37

VI. Conclusion

A measurement campaign was carried out in a suburban area, where 551 measurements were taken to obtain power values received on the digital TV channels in the UHF band. This article covers well-known and simple machine learning algorithms to identify spectrum opportunities based on reception conditions. In addition, it was also possible to use easily obtainable input data to be able to use such algorithms. When comparing the error metrics with the results of the ITU model, the machine learning models obtained better results, for the frequency 533 MHz, channel 24, which is used for training the algorithms. By comparing the algorithms already trained, on channel 24 and 29, it was observed that they performed better than the ITU-R P.1546-6 Recommendation. From the analysis it is possible to observe that the predicted values for each point tends to follow the measured data better because the points were grouped according to their geographical position, allowing the training of the algorithm to take into account the reception characteristics of each group. From the results, it is possible to see that the machine learning algorithms proposed in this work present a better prediction of the received power signal when compared with the traditional prediction models that provides an average or median value of its forecast results. From the obtained

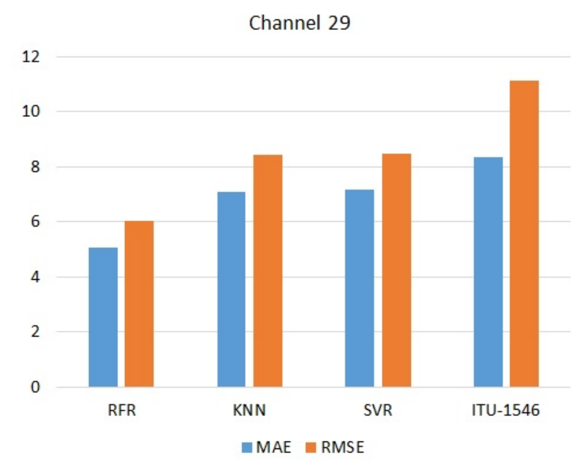


Fig. 4: Performance comparison for Machine Learning algorithms on Channel 29.

TABLE II: Algorithm comparison for RMSE.

Channel	RFR	KNN	SVR	ITU-1546
21	4.79	6.85	6.96	12.65
22	5.66	7.95	7.84	10.97
24	3.19	6.66	7.19	10.13
27	5.02	7.24	8.09	14.81
29	6.02	8.44	8.46	11.14

results, we believe that the algorithm could be used to define shadow areas that produce spectrum opportunities by applying a reception threshold criterion that depends on the system in operation, in a more appropriate way than traditional methods.

For future work we can propose to carry out simulations using the machine learning models already trained, in other locations similar to those covered in this work; new measurement campaigns to compare results and also for new training. Another suggestion is to use other algorithms, such as Artificial Neural Networks [18], to compare the gain in results with the computational cost of the methods involved.

References

[1] Q. Song, Y. Hu, X. Song and T. Zhang, “Views on Further Improving the Spectrum Efficiency,” in *Proc. International Symposium on Microwave, Antenna, Propagation and EMC Technologies for Wireless Communications*, Hangzhou, 2007, pp. 281-284, doi: 10.1109/MAPE.2007.4393601.

[2] K. Xiao, Q. Song and Y. Xie, “Challenges and strategies of the spectrum management,” in *Proc. IEEE International Symposium on Microwave, Antenna, Propagation and EMC Technologies for Wireless Communications,* Beijing, 2005, pp. 131-134 Vol. 1, doi: 10.1109/MAPE.2005.1617865.

[3] International Telecommunications Union, *Handbook on Digital Terrestrial Television Broadcasting Networks and Systems Implementation,* Geneva, Switzerland, 2016.

[4] Associação Brasileira de Normas Técnicas, *Televisão digital terrestre - Receptores - NBR15604:2015,* Standard (in portuguese), 2015.

[5] ”Method for point-to-area predictions for terrestrial services in the frequency range 30 MHz to 4 000 MHz,” Int. Telecommun. Union, Geneva, Switzerland, ITU Recommendation P.1546-6, 2019, [Online]. Available: https://www.itu.int/rec/R-REC-P.1546-6-201908-I/en

[6] K. Suh, Hyuk Jung, J. Lee and Jeong-Seok Jang, ”The calculation of field strength for DTV receiver by Rec. ITU-R P.1546,” *2010 IEEE Asia-Pacific Conference on Applied Electromagnetics (APACE)*, Port Dickson, 2010, pp. 1-4, doi: 10.1109/APACE.2010.5719757.

[7] F. S. Silva, L. J. Matos, F. A. C. Peres and G. L. Siqueira, ”Coverage prediction models fitted to the signal measurements of digital TV in Brazilian cities,” *2013 SBMO/IEEE MTT-S International Microwave & Optoelectronics Conference (IMOC)*, Rio de Janeiro, 2013, pp. 1-5, doi: 10.1109/IMOC.2013.6646535.

[8] European Conference of Postal and Telecommunications, “Technical and operational requirements for the operation of white space devices under geo-location approach,” *in ECC Report 186*, *Technical Report*, January, 2013.

[9] Ministério das Comunicações do Brasil, *Portaria Nº 925,* Standard [in portuguese] 2014. [Online]. Available: https://www2.camara.leg.br/comunicacao/rede-legislativa-radio-tv/arquivos/legislacao-arquivos/portarias-ministerio/Portarian92522AGO2014AspectostcnicosdaTVdigital.pdf. Acessed in July 2020.

[10] P. V. G. Castellanos, *Caracterização do canal de propagação banda larga na faixa UHF para aplicações de TV Digital,* D.Sc. Thesis, PUC-Rio, Rio de Janeiro, 2008 (in portuguese).

[11] Agilent, *Agilent 8560E/8590E Spectrum Analyzer Comparing Power Measurements on Digitally Modulated Signals,* White Paper, 2017.

[12] D. Xu and Y. Tian, “A comprehensive survey of clustering algorithms,” *Ann. Data Science*, vol. 2, no. 2, pp. 165-193, 2015,doi:https://doi.org/10.1007/s40745-015-0040-1.

[13] J. MacQueen, “Some methods for classification and analysis of multivariate observations,” *in Proc. Fifth Berkeley Symposium*, pp. 281–297, Los Angeles, 1965.

[14] D. J. Ketchen and C. L. Shook, “The Application of Cluster Analysis in Strategic Management Research: An Analysis and Critique,” *Strategic Management Journal,* vol. 17, no. 6, pp. 441-458, Jun. 1996.

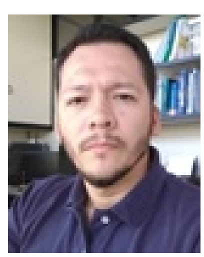
[15] G. James, D. Witten, T. Hastie and R. Tibshirani, *An Intrduction to Statistical Learning,* Springer, 2013.

[16] K. Hung, K. Lin, G. Yang and Y. Tsai, “Hybrid support vector regression and GA/TS for radio-wave path-loss prediction,” in *Proc. International Conference on Computational Collective Intelligence: Technologies and Appications*, pp. 243-251, Berlin, 2010, doi: 10.1007/978-3-642-16693-8_26.

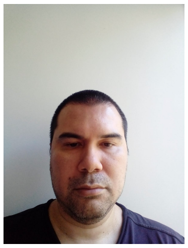
[17] K. Lin, K. Hung, J. Lin and C. Wang, “Applying least squares support vector regression with genetic algorithms for radio-wave path loss prediction in suburban environment,” in *Advances in Neural Network Research and Applications,* Springer, Heidelberg, 2010, doi:10.1007/978-3-642-12990-2_100

[18] C. E. G. Moreta, M. R. C. Acosta and I. Koo, “Prediction of Digital Terrestrial Television Coverage Using Machine Learning Regression,” *IEEE Transactions on Broadcasting*, vol. 65, n. 4, pp. 702-712, Dec. 2019, doi: 10.1109/TBC.2019.2901409.

Daniel C. Vidal received his B.Sc. and M.Sc. degree in Telecommunication Engineering at Fluminense Federal University (UFF) in 2016 and 2019, respectively.He is currently pursuing his PhD degree at University of São Paulo (USP), Brazil. His research interests are digital signal processing, adaptive algorithms and digital communication.



Pedro V. G. Castellanos was born in Cúcuta, NdeS, Colombia. He received the degree of Information Systems Engineer from Francisco de Paula Santander University in 2001.In 2003, he received the MSc and in 2008 the DScdegrees from theCenter for Telecomm. Studies (CETUC) at Pontifical Catholic University Rio de Janeiro (PUC/RJ), in Brazil. Since 2015, he is a professor and researcher at Fluminense Federal University (UFF), Brazil



Tadeu N. Ferreira is an associate professor at Universidade Federal Fluminense (UFF). He obtained a doctorate in Electrical Engineering at UFRJ in 2009. In 2019, he was part of the team that received the best demonstration award of the Brazilian Telecommunications Symposium (SBrT19). In 2015, he was awarded as the 2nd place in the Excellence in Teaching at UFF. He has written 20 articles in journals indexed by CAPES, and he has advised 1 doctorate thesis and 17 masters dissertations.