

## Research Article

# Influence of Training Set Selection in Artificial Neural Network-Based Propagation Path Loss Predictions

Ignacio Fernández Anitzine,<sup>1</sup> Juan Antonio Romo Argota,<sup>1</sup> and Fernando Pérez Fontán<sup>2</sup>

<sup>1</sup>Department of Electronics and Telecommunications, University of the Basque Country, Alameda Urquijo s/n, 48013 Bilbao, Spain

<sup>2</sup>Department of Signal Theory and Communications, University of Vigo, Campus Universitario s/n, 36200 Vigo, Spain

Correspondence should be addressed to Juan Antonio Romo Argota, juanantonio.romo@ehu.es

Received 26 July 2012; Accepted 19 October 2012

Academic Editor: César Briso Rodríguez

Copyright © 2012 Ignacio Fernández Anitzine et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper analyzes the use of artificial neural networks (ANNs) for predicting the received power/path loss in both outdoor and indoor links. The approach followed has been a combined use of ANNs and ray-tracing, the latter allowing the identification and parameterization of the so-called dominant path. A complete description of the process for creating and training an ANN-based model is presented with special emphasis on the training process. More specifically, we will be discussing various techniques to arrive at valid predictions focusing on an optimum selection of the training set. A quantitative analysis based on results from two narrowband measurement campaigns, one outdoors and the other indoors, is also presented.

## 1. Introduction

The need for connectivity anywhere, added to the increment in the number of users, has triggered the development of various generations of mobile communication standards in the last decades. The demand for greater traffic capacity involving both voice and data transmission requires the planning of mobile communication networks comprised of smaller and smaller cells, thus making the number of base stations grow exponentially, and complicating the process of determining and optimizing the location of these stations. Because of this, accurate and fast prediction models are needed for making received signal level/path loss predictions prior to actual network deployment. In this paper, we analyze the performance achievable with an intermediate technique between purely empirical and purely deterministic, based on the use of artificial neural networks (ANNs).

## 2. Prediction Models

A great variety of methods [1] has been proposed for predicting the expected received electric field level or, alternatively, the path loss. These calculations can be made using empirical

or deterministic models. An intermediate alternative is using artificial neural network-based (ANN) models.

Empirical models are based on measurement campaigns carried out in specific, representative environments. Regression techniques are then used for obtaining mathematical expressions describing the propagation loss as a function of the path length. The computational efficiency of these models is satisfactory, while having a limited accuracy. A typical example is the well-known Okumura-Hata model [2, 3].

On the other hand, deterministic models apply accurate electromagnetic techniques or simplified versions of them. These require accurate input information of the propagation environment: buildings, and so forth. Their main advantage is their precision, despite their lack of computational efficiency. It is quite common to see high frequency approximations of the full wave solutions which make use of ray-tracing techniques for identifying all possible paths between the transmitter and the receiver including multiple reflections, diffractions and transmissions through walls. The contribution of each ray is then calculated by using Fresnel's transmission and reflection coefficients, and GTD/UTD [4, 5] for diffracted contributions.

On the other hand, ANN-base models try to combine the advantages of empirical and deterministic models. ANNs are composed of several nodes or neurons divided into different levels with connections between them. The neurons may receive several input signals which are combined using appropriate weights and passed through specific transfer functions. To specify the various weights, the network must be trained. Training is carried out using measured data. Depending on the quality of the training process so will be the ability of the ANN to make predictions in unknown situations: generalization property.

In the literature, the most common choice is using feedforward networks, commonly referred to as multilayer perceptrons (MLPs) [6]. An alternative is to use the so-called radial basis function networks (RBFs) for their fast convergence, robustness, and small size [7].

Most implementations for our application use ANNs with two hidden layers. In the first, a number of neurons greater than the number of inputs is usually found [8]. However, other studies show that more complex networks do not necessarily increase the prediction accuracy. Moreover, it has been found that the generalization properties of ANNs may be reduced, that is, they may be more sensitive to the training set data [9].

In the hidden layers, nonlinear activation functions are normally used, for example, sigmoid-type functions. For the output level, linear functions are normally used. In the hidden layers, also wavelet functions can be found in received field prediction applications [10]. However, even though they show faster computation times, in contrast, they require much larger training data sets.

Different algorithms can be used for training an ANN. In [11], their efficiencies were analyzed showing that the best results are obtained with Bayesian regularization and Levenberg-Marquardt techniques, the latter being the most used option. Another algorithm also used [12, 13], which offers good performances is the resilient propagation algorithm.

ANNs can also be combined with other techniques for characterizing the effects of RF propagation. When simulation time is critical, the so-called “dominant path,” selected by means of a ray-tracing tool, can be used to provide the necessary inputs to the ANN. This leads to acceptable results both in terms of time and accuracy. The dominant path is the propagation path between the transmitter and receiver showing the smallest loss. Thus, instead of searching for all possible ray combinations, the problem is simplified while an acceptable generalization performance may be achieved. The dominant path can be calculated using two main techniques: the recursive neighboring model [14] and the convex corners approach [15].

In the last few years, many researchers have applied ANNs for predicting the path loss in indoor [8, 16], outdoor urban [17, 18], and rural [9] environments. In the above references, extensive descriptions and optimizations of ANN architectures, trainings, and generalizations have been presented. However, special attention must still be paid to the repercussions of using different criteria for selecting the

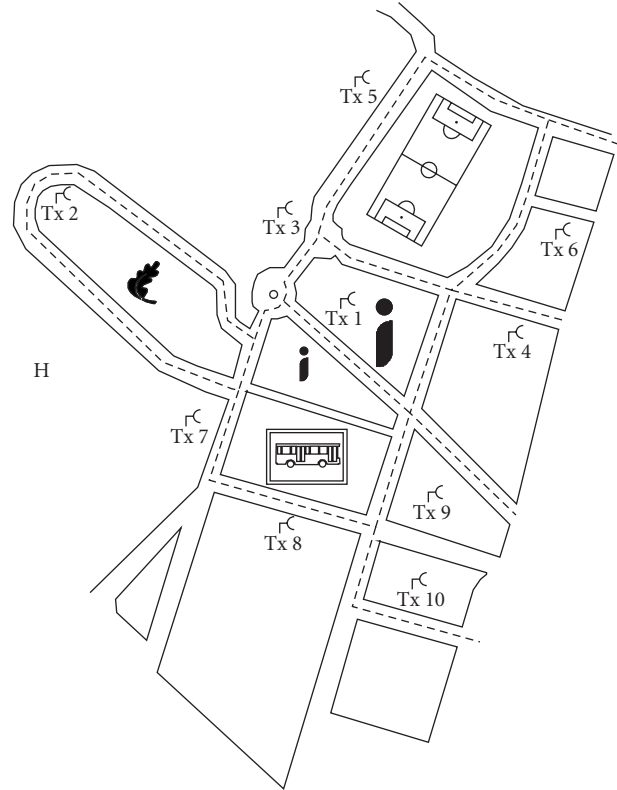


FIGURE 1: Outdoor measurement routes and transmitters.

training data set. This is the main issue discussed in this paper.

### 3. Measurements and Tools

In this section, the main features of the measured data are presented, then we go on to present the developed ANN tool which operates in combination with a ray-tracing tool able to identify the dominant path between transmitter and receiver. Typically, a single transmitter is assumed while various receive locations can be defined as part of a route or a meshed grid. The route option is very well suited for the training process.

A continuous wave (CW) transmitter was set up at a number of sites, while the received power was measured at several points along a number of routes. Measurements were repeated several times so as to average out the signal cancellations and enhancements due to multipath. For each measurement point, information on its coordinates and the received power level in dBm were recorded. All the outdoor and indoor measurement routes and the transmit locations are shown in Figures 1 and 2, respectively. The CW measurements were made at the 900 and 1800 MHz bands, using in both cases a vertically polarized 4 dBi gain antenna and 35 dBm transmit power. The receiver was a spectrum analyzer connected to a PC. Measurements were triggered every 350 cm along the route. The receive antennas were also vertically polarized, with omnidirectional patterns and 0 dBi gains.

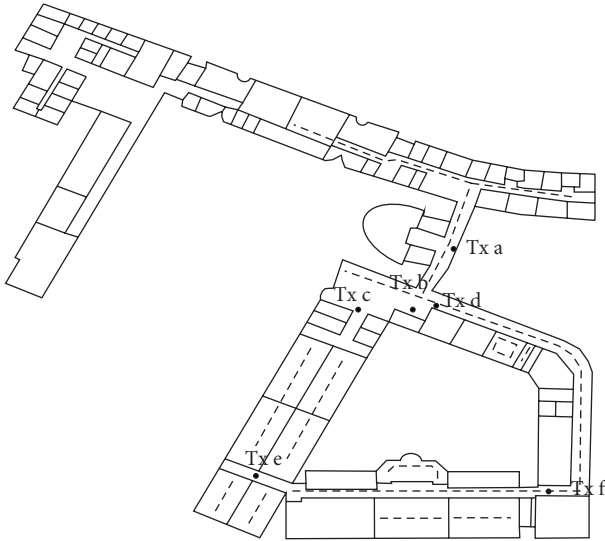


FIGURE 2: Indoor measurement routes and transmitters.

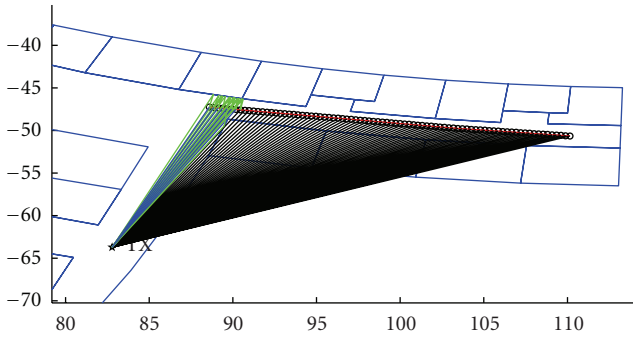


FIGURE 3: Example of dominant path calculation in the indoor case. Blue lines represent direct ray paths, Green lines represent reflection paths and black lines represent diffraction paths.

Our ANN model works in combination with a simplified ray tracing tool. This performs CAD tasks as well as basic ray tracing for finding the dominant propagation path for each Tx-Rx pair, then it calculates this path's parameters.

For both outdoor and indoor links, the dominant path can belong to any of four different types: (a) direct ray paths, when the line-of-sight, LOS, path is not blocked, (b) wall-reflection paths, (c) corner-diffraction paths, and (d) propagation through-obstacle paths, when it is not possible to link the transmitter and receiver with one of other three path types. In this last case, a straight line is drawn from one end to the other. Each time this line crosses an obstacle, for example, a wall, the corresponding loss is added. Figure 3 illustrates this classification for the indoor case.

#### 4. ANN-Based Model

Starting from an earlier version of the tool [16], we have implemented a new one using the dominant path approach. Then, this implementation has been trained with measurements. Finally, comparisons between predictions

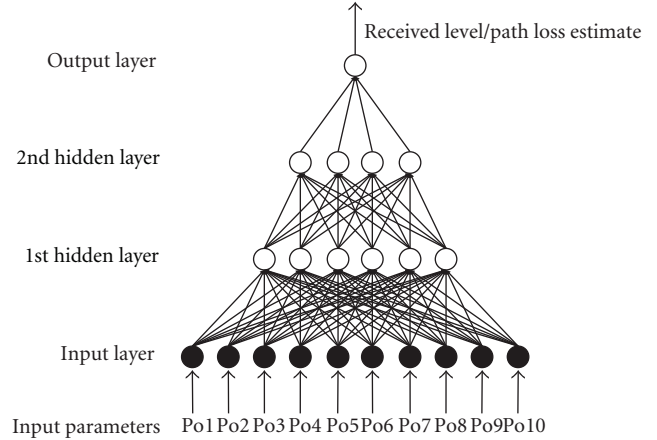


FIGURE 4: Architecture of the indoor neural network.

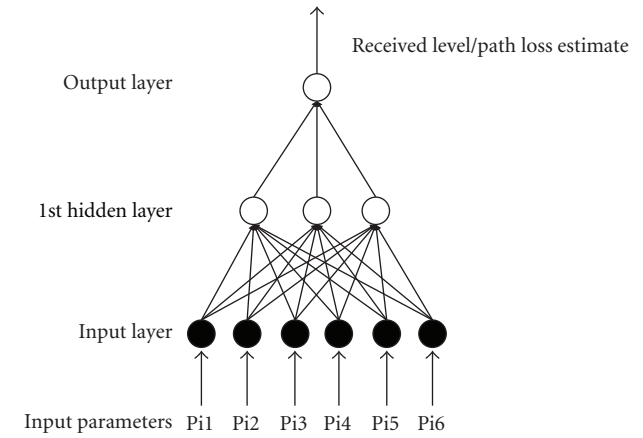


FIGURE 5: Architecture of the outdoor neural network.

and measurements for data sets different from those used for training were carried out. Two different ANNs have been implemented for indoor and outdoor scenarios, respectively. Two different networks were necessary due to the significant differences in propagation conditions in the two scenarios.

The indoor ANN is a MLP network with pyramidal structure consisting of three main parts: an input layer with 8 neurons, each associated with one of the 8 selected input parameters, two hidden layers with 6 and 4 neurons, respectively, with sigmoid-type activation functions and, finally, an output layer with a single neuron with a linear function (Figure 4). The outdoor ANN uses fewer inputs resulting in a simpler structure (Figure 5).

The input parameters must characterize the propagation path between transmitter and receiver in the most faithful way. Numerous parameters could have been selected. After several trials, we selected the parameters listed below.

##### (a) Indoor Scenarios

- (i) *Screen effect, Po1, Po2.* It occurs when there are walls near the transmitter or receiver blocking the direct ray.

TABLE 1: Classification of receive locations.

	Outdoor receivers	% Of total	Indoor receivers	% Of total
LOS direct ray	1352	20.2	572	16.7
NLOS reflection	526	7.8	—	—
NLOS diffraction	1972	29.4	635	18.6
NLOS obstacles	2861	42.6	2213	64.7
Total	6711	100	3420	100

TABLE 2: Distribution of measurement points in sets A and B according to their dominant paths.

Training routes	Set A	Set B
LOS direct ray	194	268
NLOS reflection	5	56
NLOS diffraction	250	494
NLOS obstacles	237	662
Total	686 (8.9% of 6711)	1480 (22% of 6711)

- (ii) *Local reflections, Po3, Po4.* They exist when either the receiver or the transmitter are located close to a corner giving rise to multiple reflections.
- (iii) *Waveguide effect, Po5.* It appears in corridors.
- (iv) *Change of direction, Po6.* It occurs when diffraction takes place.
- (v) *Transmission loss, Po7.* It is introduced when the signal must pass through an obstacle.
- (vi) *Free space loss, Po8.* It depends on the distance between the transmitter and receiver, and the working frequency.

#### (b) Outdoor Scenarios

- (i) *Distances L1 and L2, Pi1, Pi2.* They are defined as the separations between the transmitter/receiver and the interaction point (reflection or diffraction point). The longer these distances are, the larger the loss will be.
- (ii) *Incidence and scattering angles, Pi3.* They are defined with respect to a wall's normal.
- (iii) *Reflection and diffraction coefficients, Pi4, Pi5.* Fresnel's reflection coefficients and UTD edge diffraction coefficients.
- (iv) *Free space loss, Pi6.* It depends on the distance between the transmitter and receiver, and the working frequency.

The most critical step when designing an ANN-based model is the training process which will condition the achievable prediction accuracy. The *back-propagation* technique was selected as learning method, where the predicted power is compared with the actual measurement, and the difference (error) is fed back to the network for correcting the various network connection weights.

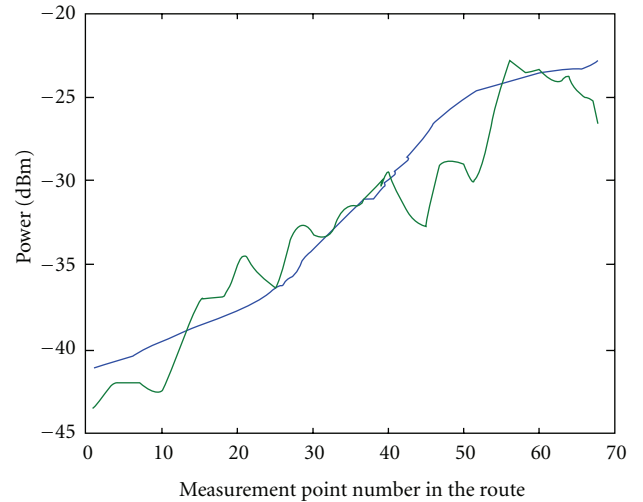


FIGURE 6: Example of prediction result and comparison with measurements.

The *Levenberg-Marquardt* algorithm was used for training the model. This method uses the evolution of the gradient changing the coefficient for each neuron connection in the direction that causes a larger error reduction. The chosen number of training cycles was one thousand. This is a tradeoff between error, and time. As said, the selection of the training set is the most critical issue and will be discussed in depth below.

After the ANNs were trained, we analyzed the prediction errors by comparing the results of the ANN-based model and the received power levels measured at points different from those used in the training phase. Figure 6 illustrates a measurement route and the obtained prediction. For each route, the mean error, mean squared error and standard deviation were calculated. In the figure we can observe how the prediction curve is much smoother than that of the measurement. This is because the ANN input parameters, obtained from the ray-tracer, are very similar for neighboring points along the route. The user of such a prediction tool must be aware of this limitation. Still, as observed, the average error and its spread are very small.

## 5. Selecting the Training Set

As discussed in previous sections, a wise selection of real propagation paths from which the neural network will learn how to calculate the received power is the most critical factor

TABLE 3: Error statistics for the outdoor case with the ANN trained with set A and with set B.

Test routes	Route 1	Route 2	Route 3	Route 4	Route 5
Network trained with training Set A					
Mean error	10.03	8.09	11.41	5.76	3.86
RMS error	14.17	10.11	13.26	8.29	5.15
std	10.01	6.07	6.76	5.97	3.41
Network trained with training Set B					
Mean error	6.66	6.70	5.40	4.26	1.77
RMS error	10.54	9.49	7.62	6.58	2.52
std	8.17	6.72	5.24	5.01	1.79

TABLE 4: Path types used in indoor trainings.

Training routes	Set C	Set D
LOS direct ray	399	406
NLOS diffraction	30	101
NLOS obstacles	219	390
Total	648 (17.5% of 3420)	897 (26% of 3420)

in the training phase. Those real situations form the so-called “training set.”

To optimize the training set several routes with different characteristics must be selected so as to provide the ANN with all the propagation conditions (reflection paths, direct ray paths, etc.) likely to be encountered. In addition, the selected routes have to include received positions showing different ranges of input parameters. In this way, the network will learn to behave in many different situations and will be able to make correct generalizations when applied to new cases. After learning from a number of routes, the network must be tested with other data sets from different routes. Predictions for those test routes must show similar errors to those for the training routes. If this is the case, network will be correctly trained.

The first and essential step in the training process involves a suitable characterization of the measurements points in the training routes according to their dominant path type. The choice of training routes must be a planned process based on supplying a sufficient and balanced number of measured points belonging to the various propagation conditions to be expected. Based on the dominant path concept, we have to be careful when training the ANN to provide an appropriate mix of the four path types identified.

A total of 29 measurement routes were recorded, each with a different number of receive positions depending on its length. For outdoor links, a total of 50 routes were measured. Hence, the available measurements correspond to a total of 79 routes with 3420 sampling or receive points for the indoor case and 6711 for outdoor locations. As indicated earlier, each route was measured several times and, then, point-wise averages were calculated. The number of transmitter sites in the indoor case was 6, while for the outdoor case 5 sites were used. Table 1 presents a summary of all measurement locations according to their corresponding path types.

Two strategies have been analyzed in the selection of the training set. In the first, we selected entire routes while the second focused on selecting specific receive points according to the dominant path category to which they belonged.

We now analyze the first, that is, route-wise strategy. From the available measurements, a subset of the routes was used for training while the rest was used for testing. To illustrate the effect of the number of routes considered in the training process in relation to the achieved prediction accuracy, several training sets were used as discussed below, both for the indoor and outdoor cases.

To train the outdoor network, two different sets were used. Set A consisted of data gathered from a single transmit site and three different routes. In all, 686 data points: 194 corresponding to direct ray paths, 250 to diffraction paths, 5 to reflection paths, and 237 to through-obstacle paths. Set-B consisted of data from seven routes and 2 transmit sites, in all 1480 data points classified as follows: 268 were direct ray paths, 494 diffraction paths, 56 reflection paths and 662 through-obstacle paths, Table 2.

After training, measurements from 5 routes corresponding to a different transmit site were used to test the ANNs trained with sets A and B. Table 3 shows the results of this analysis. For set A, acceptable error levels were obtained when the test routes showed similar propagation characteristics to those used in the training process. However, for the other routes, all three error parameters (mean, RMS and standard deviation) were rather high, even over 10 dB. At some locations such as those corresponding to reflection paths, predictions were worse than those observed when training the network with set A. This is due to the fact that only 5 data points corresponding to this path type were used in the training. Thus, the network could not learn how to behave in reflection-dominated paths. It is clear that the training needed improvement for this type of paths. On the other hand, set B contained a more balanced mix of data points corresponding to all four classes. In this case, the error statistics are drastically reduced.

For training of the indoor network, two sets were also used. Set C consisted of data from two transmitters and four different routes. In all, 648 measurements were used: 399 points corresponded to direct-ray paths, 219 to through-obstacle paths and 30 to diffraction paths. Set D consisted of eight routes corresponding to four transmitters. Now, 897 training points were used (26.3% of a total of 3420). The distribution of path types is as follows: 406 were direct ray

TABLE 5: Numerical results of simulations, with two and four transmitters, for the indoor routes.

Test routes	Route 6	Route 7	Route 8	Route 9
Network trained with set C				
Mean error	1.23	2.58	2.19	2.37
RMS error	1.72	3.46	3.14	3.27
std	1.20	2.31	2.25	2.25
Network trained with set D				
Mean error	0.96	1.99	1.36	1.47
RMS error	1.48	2.84	1.99	2.60
std	1.13	2.03	1.45	2.15

TABLE 6: Errors for the path-type oriented analysis for the outdoor case.

Test routes	LOS direct ray	NLOS reflection	NLOS diffraction	NLOS obstacles
Mean error	5.77	7.49	8.78	7.32
RMS error	9.34	9.04	11.88	8.84
std	7.35	5.07	8.00	4.95

TABLE 7: Errors for the path-type oriented analysis for the indoor case.

Test routes	LOS direct ray	NLOS diffraction	NLOS obstacles
Mean error	1.77	3.26	2.15
RMS error	2.52	4.74	3.15
std	1.79	3.44	2.30

paths, 101 diffraction paths, and 390 through-obstacle paths, Table 4. For the test set four complete routes were used, Table 5.

Again, in the case of Set-D, the errors were much smaller than for the Set-C. With the new training, the same routes were simulated. Due to the path type mix in Set-C, routes with diffraction paths were badly predicted: the network so trained cannot properly simulate those measurement points where the dominating conditions are not sufficiently well represented in the training set. Training Set-D introduces more measurements and also covers a more balanced mix of propagation path types. Thus, the selected routes in Set-D encompass an appropriate assortment of paths from all types.

Now we analyze the second strategy to selecting the training set, that is, a path-type oriented selection. In this case, the training process was separately carried out for each type of propagation path. Training the ANN with separate receiver locations according to their propagation path types could, in principle, allow achieving a much better prediction accuracy. According to this approach, several routes were split into subsets, as a function of their dominant path, so that all receive points with a direct-ray predominant path were placed into the same subset. Then, some of those points were used to train the ANN and others for testing it. The same was done for reflection, diffraction, and through-obstacle paths.

As shown in Table 6, results for reflection, diffraction and through-obstacle paths show a similar error parameter range, in the order of 7-8 dB. Meanwhile, the variability of direct ray paths proved to be lower than in the other cases. A similar analysis was carried out for the indoor case, Table 7. Now, the error parameter range in through-obstacle and diffraction paths is in the order of 2-3 dB, whereas for direct ray paths it again shows a lower value. In any case, even though both in outdoor and indoor situations, the general performance is quite good, it does not seem to be much better than the one achieved in the previous analyses.

## 6. Conclusions

To create an effective ANN and properly make path loss predictions, a correct training strategy must be devised. The selection of the training sets is the most critical factor to ANN prediction performance in this application. An appropriate assortment of different propagation conditions represented by different types of propagation paths is required so the net can learn how to behave and make suitable generalizations in as many different situations as possible.

In this paper, we have focused on an implementation combining a simplified ray-tracing tool which takes care of identifying the so-called “dominant path” and calculating a number of propagation path-related parameters used as inputs to the ANN which, in turn, makes the final prediction.

When we indicate that there is a need for an appropriate assortment of paths with different propagation conditions, the selection has to be based on a classes defined according to the dominant path. Both for indoor and outdoor conditions, four different dominant path classes have been identified. When the above premises are fulfilled, ANNs may very well represent a good alternative to predict radio propagation with errors in a similar range to other, more complex methods with more computational load. From our experimental analyses the error parameters, mean, rms, and standard deviation were always below 7 dB.

To achieve these results in a training strategy oriented toward the dominant path, the training points need to be adequately selected so that they are representative of the ensemble of the possible types in the coverage area. This selection requires an in-depth knowledge of the propagation scenario, and hence an elevated cost for collating the data in the set which in practice is unfeasible.

In a complete route-oriented strategy, the accuracy of the achieved results will depend on the total number of routes in the training set. It was observed that as the number of samples is increased so does the accuracy, especially for a small number of routes. If the sample size is properly balanced, further increments will not produce significant performance improvements while the cost increases.

In this paper, the balanced size corresponds to a route selection approximately encompassing 25% of the foreseen coverage area. The selected routes should provide diversity of cases while they are validated through a simple process. Such a set produces similar results as with a set based on the dominant path types found in the coverage area. In summary, adopting this strategy will lead to the generation of a less complex training set at much smaller cost than using a path type-oriented strategy and achieving similar accuracies.

A word of caution must be said, however. As illustrated in Figure 6, ANN predictions for consecutive points belonging to the same route cannot follow some of the sharp variations encountered in the measurements, where the measurements are already the results of averaging over several repeated passes, that is, they contain the slow channel variations due to shadowing, but the multipath has been removed. This is because the inputs to the net provided by the ray-tracing plus dominant path tool do not change so drastically from point to point. This shortcoming needs to be born in mind when considering the application of this approach.

## References

- [1] S. Saunders and A. A. Aragon-Zavala, *Antennas and Propagation for Wireless Communications*, John Wiley & Sons, London, UK, 2nd edition, 2007.
- [2] Y. Okumura, E. Ohmori, T. Kawano, and K. Fukuda, "Field strength and its variability in VHF and UHF land-mobile radio service," *Review of the Electrical Communications Laboratory*, vol. 16, no. 9-10, pp. 825-973, 1968.
- [3] M. Hata, "Empirical formula for propagation loss in land mobile radio services," *IEEE Transactions on Vehicular Technology*, vol. 29, no. 3, pp. 317-325, 1980.
- [4] J. H. Tarnng, "Three-dimensional modeling of 900-MHz and 2.44-GHz radio propagation in corridors," *IEEE Transactions on Vehicular Technology*, vol. 46, no. 2, pp. 519-527, 1997.
- [5] M. F. C atedra and J. P erez Arriaga, *Cell Planning For Wireless Communications*, Artech House, Norwood, Mass, USA, 1999.
- [6] S. Haykin, *Neural Networks: A Comprehensive Foundation*, IEEE Press, McMillan College Publishing Co, 1994.
- [7] Y. Sun, Y. Xu, and L. Ma, "The implementation of fuzzy RBF neural network on indoor location," in *Proceedings of the Pacific-Asia Conference on Knowledge Engineering and Software Engineering (KESE '09)*, pp. 90-93, December 2009.
- [8] A. Nešković, N. Nešković, and D. Paunović, "Indoor electric field level prediction model based on the artificial neural networks," *IEEE Communications Letters*, vol. 4, no. 6, pp. 190-192, 2000.
- [9] E. Ostlin, H. J. Zepernick, and H. Suzuki, "Macrocell path-loss prediction using artificial neural networks," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 6, pp. 2735-2747, 2010.
- [10] F. Cheng and H. Shen, "Field strength prediction based on wavelet neural network," in *Proceedings of the 2nd International Conference on Education Technology and Computer (ICETC '10)*, pp. V2255-V2258, June 2010.
- [11] I. Vilović, N. Burum, and Z. Šipuš, "A comparison of neural network models for indoor field strength prediction," in *Proceedings of the 49th International Symposium ELMAR-2007 focused on Mobile Multimedia*, pp. 251-254, Zadar, Croatia, September 2007.
- [12] I. Popescu, I. Naforni a, Gh. Gavriloaia, P. Constantinou, and C. Gordan, "Field strenght prediction in indoor environment with a neural model," in *Proceedings of the 5th International Conference on Telecommunications in Modern Satellite, Cable and Broadcasting Service (TELSIKS '01)*, September 2001.
- [13] B. Monteiro, G. P. S. Cavalcante, H. S. Gomes, D. M. Ros ario, F. F. Lima, and H. A. Junior, "Evaluation of radio propagation parameters for field strength prediction using neural network," in *Proceedings of the SBMO/IEEE MTT-S International Microwave and Optoelectronics Conference (IMOC '07)*, pp. 888-892, November 2007.
- [14] G. W ofle and F. M. Landstorfer, "A recursive model for the field strength prediction with neural networks," in *Proceedings of the 20th International Conference on Antennas and Propagation*, vol. 2, pp. 174-1177, Edinburgh, UK, April 1997.
- [15] G. Woelfle and F. M. Landstorfer, "Dominant paths for the field strength prediction," in *Proceedings of the 48th IEEE Vehicular Technology Conference (VTC '98)*, vol. 1, pp. 552-556, May 1998.
- [16] J. A. Romo, I. F. Anitzine, and F. P. Font an, "Application of neural networks to field strength prediction for indoor environments," in *Proceedings of the European Conference on Antennas and Propagation (EuCAP '06)*, November 2006.
- [17] I. Popescu, D. Nikitopoulos, P. Constantinou, and I. Nafornita, "ANN prediction models for outdoor environment," in *Proceedings of the IEEE 17th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '06)*, September 2006.
- [18] Z. Stankovic B, Milovanovic, M. Veljkovic, and A. Dordevic, "The hybrid-neural empirical model for the electromagnetic field level prediction in urban environments," in *Proceedings of the 7th Seminar Neural Network Applications Electrical Engineering*, Belgrade, Serbia, 2004.