LOCATION FINGERPRINTING IN GSM NETWORK AND IMPACT OF DATA PRE-PROCESSING

C. M. Takenga, Student Member IEEE[1], K. Kyamakya[2]

[1]IKT, University of Hannover, Hannover, Germany, takenga@ant.uni-hannover.de
[2]Alpen Adria University Klagenfurt, Chair of Transportation Informatics, Klagenfurt Austria, kyamakya@isys.uni-klu.ac.at

Abstract

Different techniques for locating the user terminal in cellular networks have been developed in the recent years. This paper proposes a method which is designed to obtain optimal performance in urban environments. Such performance metrics as accuracy, cost, reliability, scalability, and demanding effort are the key points in this paper. A fingerprint method is used for some of these reasons: firstly, no additional hardware is required for it implementation to existing networks. Secondly, compared to other methods, this one performs better in areas with significant multipath propagation. Due to the big demanding effort during radio signal strength (RSS) collection from the streets to realize a fingerprint database, predicted RSS data of the concerned area are rather used. The high sensitivity of the RSS to the environmental changes is taken into account. A system is developed which calibrates the predicted RSS on basis of a sample real collected RSS. A robust neural network (NN) architecture and training algorithm are used in the positioning unit in order to get a good mapping between locations and RSS.

I. Introduction.

Positioning methods based on Angle of arrival (AOA) and time of arrival (TOA) are expensive to implement in current networks and impractical in micro cell sites. With these methods, the location of the mobile system (MS) is calculated under the assumption of the line of sight propagation between the base station (BS) and MS. This assumption is not valid in city centers where high buildings often obstruct the line of sight. Moreover, severe multipath propagation characteristic to these environments makes it difficult to detect the angle or time of arrival of the direct component. Consequently, methods based on the AOA and TOA are not suitable for dense cellular network of urban environments. In this work, we use a location fingerprinting method. This method does not need any additional equipment to existing networks and it is designed to obtain optimal performance in urban environment.

A NN is trained to do the RSS-Position mapping .

Intelligent systems require training or expert knowledge. A large amount of training data is required to build the database for training the network. This training data is often hard to obtain, time consuming and may not be a good representation of the total dataset. The signatures are the RSS. The high sensitivity of the RSS to the environmental changes makes it difficult to get the best estimated value. A system which remedies to this issue is applied in this work.

The remainder of this paper is organized as follows. In section 2, we survey related works in RSS-fingerprinting position determination technologies. In the third section, we present our fingerprint database and the experimental environment. We describe our system in section 4. In 4.a. we show our performance metrics. Our research methodology for calibration is discussed in 4.b. The noise cancellation algorithm appears in section 4.c. The location technology and the positioning results are in section 4.d. Finally, we conclude in section 5.

II. Related works

Many researchers have paid an attention on this issue of location of a mobile system using the received RSS[1-6]. In [1] different positioning methods have been tested in urban, suburban and rural environments and the location fingerprinting has appeared to perform the best in urban and indoor situations.

Some works have used the real collected RSS to train the NN, which requires RSS collection from many streets for a better accuracy[3, 5, 6]. Others have used the predicted data, which was less accurate than the first due to the difference between the prediction and the real RSS. This paper finds a compromise between the demanding effort and accuracy. The predicted data are corrected by only a sample of real collected data.

The RSS were used to train the NN used for positioning in[3] but the issue due to the random behavior of the RSS was not taken into account.

In [4, 6], positioning is tackled as a multi-class classification problem. The area of interest is divided into small square sections and location estimation of the MS section using received power levels from different base stations was the classification task.

III. Fingerprint database and experimental environment

Several works have been focused on the RSS prediction models [7,8]. A database of the predicted RSS for our interested area is from the E-Plus Mobile Network (Germany). This database is from a prediction model based on Outdoor and Outdoor-to-indoor coverage in urban areas at 1.8 GHz[7].
The predicted RSS used are from 10 Base stations, see Fig.1. whereby four of them are indoor antennas. Each of the 6 outdoor Base stations has three sectors. The experimental area is an urban area with some high buildings. Data set contains 365 148 data records from each of the 22 cell antennas generated from 365 148 locations (5m x 5m resolution) in an area 3 km x 3 km. The red spots on Fig.1 correspond to the location of the BS antennas.

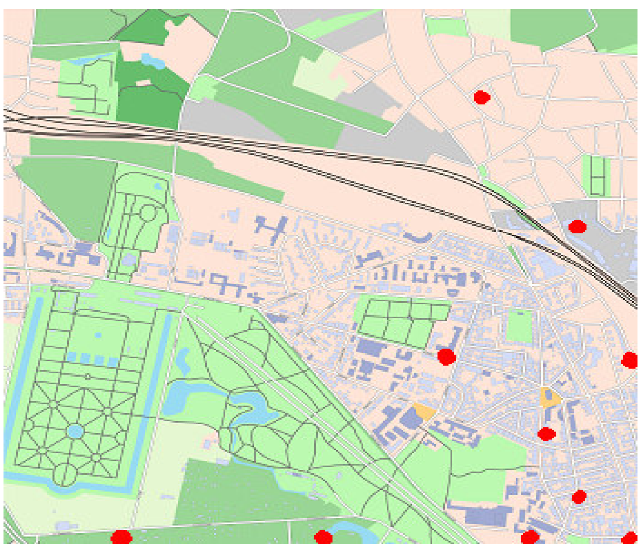

Figure 1: Experimental area(3x3km) and BS positions in Hannover-Nord Stadt, Germany.

IV. System Description and experimental results

A. *Performance metrics.*

The most important performance metric in positioning issue is the accuracy of the location information. This is usually reported as an error distance between the estimated location and the actual mobile location. The report accuracy should include the confidence interval or the error probability function.
An other essential performance metric is the scalability of the positioning system. The scalability is a metric that suggests how well the system performs when it operates with a larger number of location requests and a larger coverage. In location fingerprinting, this issue is of big interest. In order to get a good generalization, NN training data should be selected to cover the entire region where the network is expected to operate. In case of a large area, the training data is so large that the NN is unable to perform the mapping.
In order to remedy to this issue, the area used for the experiment can be divided into sections, see Fig. 2. With experiment, we realized that one trained NN can provide positioning results satisfying the FCC requirements in urban area 3x3 Km. Therefore, sections are of size 3x3 Km. The known parameter (current active cell) determines which trained NN should be used in the positioning algorithm according to Fig.3.

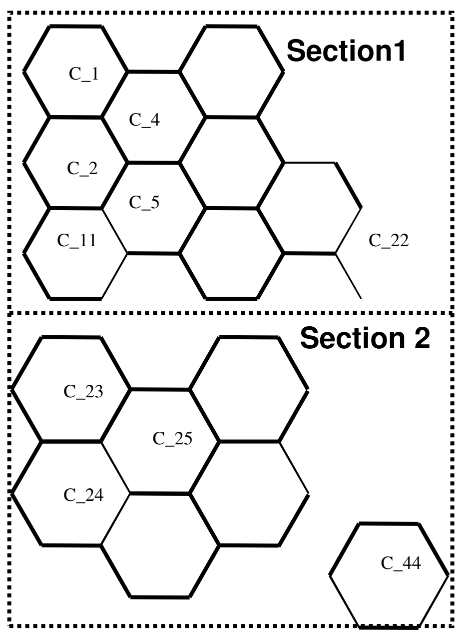

Figure 2: Division of big area in sections

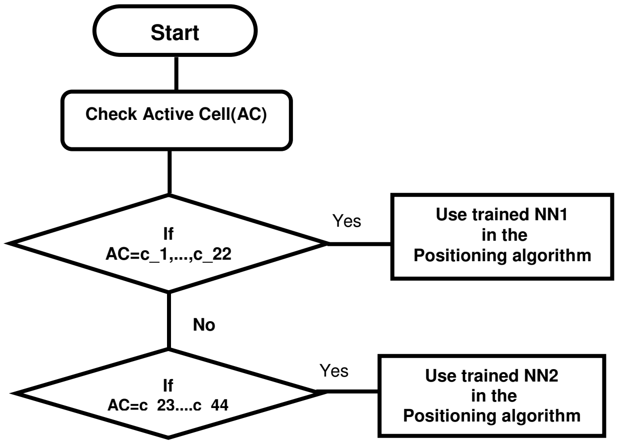

Figure 3: Flowchart showing the NN to be used in the positioning algorithm

The cost is the next performance metrics. This can come from the cost of extra infrastructure, additional band width and nature of the deployed technology. This Fingerprinting method does not need any additional equipment to the existing network.
The demanding effort is also one of the performance metrics considered in this work. Due to the big demanding effort during RSS collection from the streets to realize a fingerprint database, predicted data of the area of concern are rather used.
Reliability of our positioning algorithm is among the major performance metrics. A system including data pre-processing, suitable NN architecture and robust training algorithm will remedy to this issue.

B. *Methodology for calibration*

In this work, calibration is referred to as correction of predicted RSS by a sample of measured data from the streets.
Due to the big effort during RSS collection from the streets to realize a fingerprint database, predicted RSS-data of the area of concern are rather used in our experiment. The wave propagation modeling in urban areas is a very complex task. Various propagation effects have to be considered. An exact modeling of all these effects in such a complex environment is not appropriate in a practical implementation due to computing time constraint. Only the most important factors are taken into account. Due to this above factor, the predicted RSS provided differ to the real data. For this reason a calibration system will be developed to correct these predicted data on the basis of sample RSS measurements.
Neural network is used to make the relationship between the predicted and measured RSS as shown in Fig.4. The flexibility and ability of the NN to deal with uncertain data is the main reason of its use in this calibration issue.

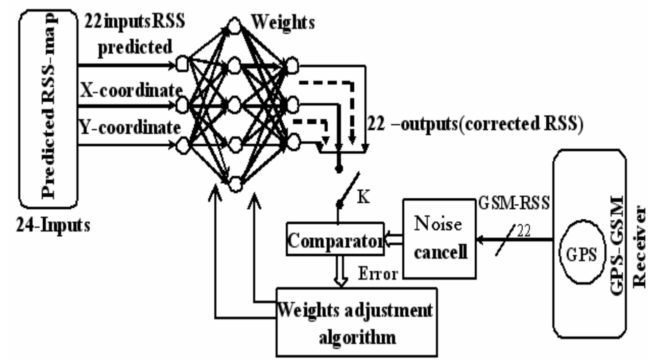

Figure 4: Block diagram used for calibration

The GSM-RSS used to correct the predicted data are collected along the streets of our experimental area by using a GPS-GSM antenna with a FALCOM A2D-3 modem . Fig.5

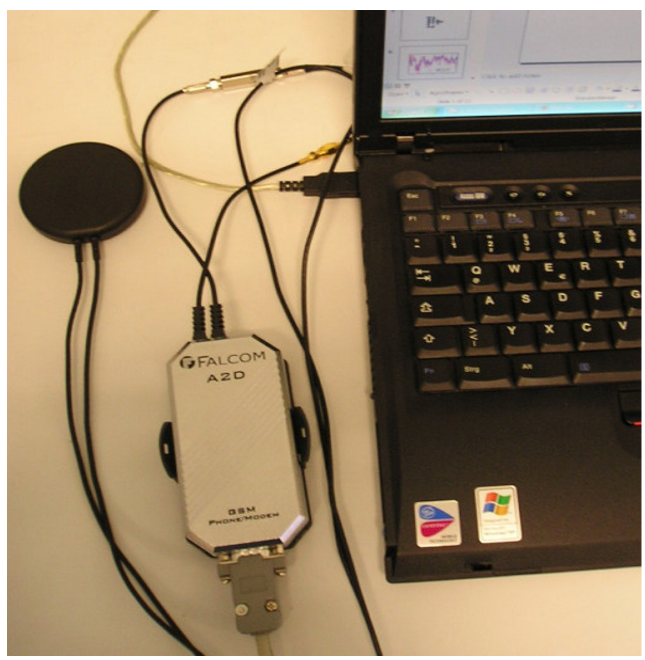

Figure 5: GSM-GPS antenna and modem. Used for RSS collection

The NN has 24 inputs, i.e. 22 predicted RSS from 22 antennas and 2 inputs which are the coordinates of the points used.
The two position coordinates are involved to the input in order to make a better generalization for the entire experimental region. 22 outputs of the NN are compared to a string of 22 RSS collected data from the street. Due to

the limitation of standard GSM receiver, only a maximum of 7 strongest RSS values appear in measured record. Therefore, missing values are assumed to be equal to the limited sensitivity level. The collected RSS from the GSM-GPS receiver are filtered from noise in the Noise cancellation unit.
Most standard RF receivers have a limited receiver sensitivity level. Therefore, all the predicted RSS which power values are smaller than -113dBm or RxLev = 0 are fixed equal to this limited value during training phase.

The error coming from the comparator, which is the difference between the desired output and the actual one, is then utilized in weight-adjusting algorithm to determine the amount of adjustment to be made in the weights in both layers.
In order to start the training process, the connector K is closed and weights are randomly adjusted to small random values. When the input vector with predicted RSS and the corresponding coordinates is applied to the NN, it produces an output vector which is compared to the target vector (filtered GSM-RSS) to produce the error. The weight-adjusting algorithm then modifies the weights in the direction that reduces this error. When the input vector is again applied, it produces a new output and this process is repeated over and over until the error is minimized to some specified value or to an irreducible small quantity. At this time, the NN is said to have been trained to map input vector into desired output vector.
The connector K can then be opened and any predicted RSS at the input will give us a corrected value of that predicted RSS at the output.

The algorithm used to adapt the weights in this work is gradient descent for its simplicity and low memory requirement for generalization. The mean squared error of the system is minimized by moving down to the gradient of the error curve. A risk of stacking in local minima is present, and in this case is to start over by reinitializing the weights to some new set of small random values. Geometrically, this changes the starting position of the network.
NN are very sensitive to absolute magnitudes, if some of the input range from 0dB to -113dB and other are very big values (x-UTM, y-UTM), fluctuations in the big range input will tend to swamp any importance given to the first, even if the first input is much more important in predicting the desired output. To minimize the influence of absolute scale, all inputs to a NN are scaled in our case and normalized so that they correspond to roughly the same range of values 0 to 1.

During the RSS samples collection, it does appear that some data are missing. In spite of the robustness property of the NN that it can work with incomplete data sets, missing data can create serious problems. If a data cannot be found, the common sense and technically correct thing to do is to replace every missing value with the best estimate of what it would have been. In our case, the limited sensitive value appears to be the best estimate.

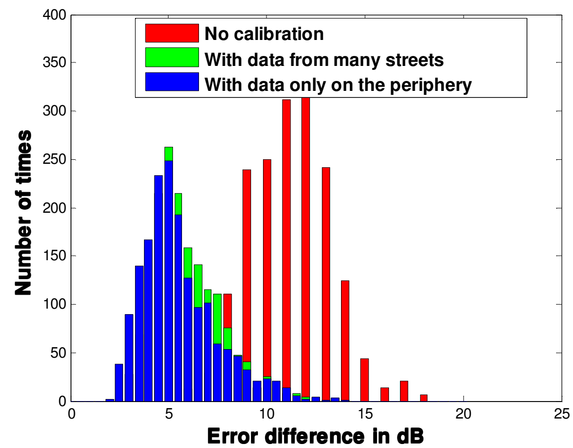

Figure 6: Histogram showing the dependence of calibration error on different scenarios

Fig.6 shows the calibration impact on the data. Without calibration, RMSE is around 13dB. With calibration, the error is reduced to a value around 5dB of difference. It is also shown that, only a sample of collected data (data collected from the periphery only) is able to calibrate the whole experimental area. This reduces the demanding effort.

$$RMSE = \sqrt{\left[RSSmean_k - RSS'mean_k\right]^2},$$

where $RSSmean_k = \frac{1}{22} * \sum_{i=1}^{22} \left[RSS_{ki}\right]$, is the mean value of the predicted RSS from the 22 BS antennas at k position, i is the index of the cell antenna, $RMSE$ is the root mean square error, i.e. the difference between predicted and measured values at a k point. $RSS'mean_k$ is the mean value of the measured RSS from the all the cell antennas at k position.

C. Noise cancellation algorithm

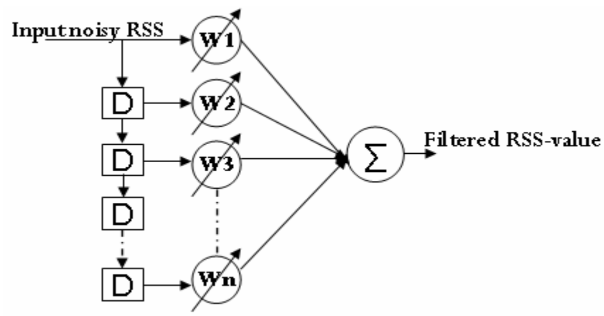

Figure 7: Noise cancellation with TDNN (Time delay NN)

The noise cancellation block on Fig.4 has a big role. This block solves the problem related to the stochastic behaviour of the received power of the radio signal. The high sensitivity of the RSS to the environmental changes makes it difficult to get the best estimated value. A time delayed NN (a single input adaptive transverse filter) is able to remove a large portion of the noise in the received RSS, see Fig.7 It actually performs this operation by calculating a linear weighted average over a window. In this paper, the window is six time steps long, that means we used 5 delay units (D). This procedure can be performed either online or off line. A string of RSS from a BS is connected to the input of the TDNN, and at the output we have the filtered values. Fig.8 shows us the RSS before and after the noise cancellation unit. The result shows that a large portion of the noise in the received RSS has been removed. This TDNN uses not only the actual RSS to reduce the noise but also the previous ones. The RSS history helps a TDNN to make a best estimation of the values.

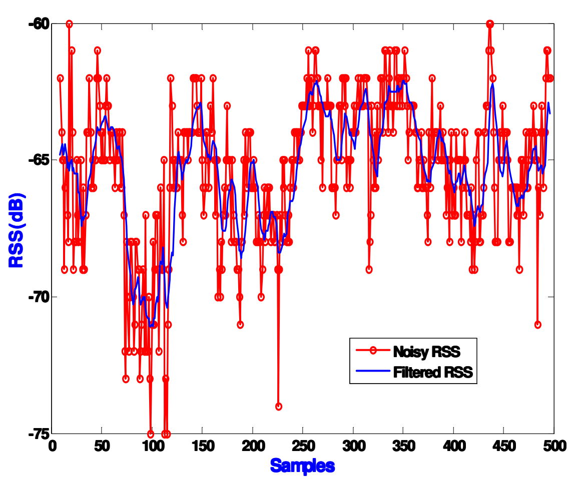

Figure 8: Noisy and filtered RSS

D. Positioning algorithm

NN training data should be selected to cover the entire region where the network is expected to operate. We use a supervised learning. Our input patterns are the corrected predicted RSS and the correct outputs are their corresponding positions. Identification of the appropriate NN architecture is of great importance. A large number of hidden layers increases the processing power of the NN but requires significantly more time for training and a larger number of training examples to train the NN properly. The number of neurons in each layer is determined by the nature of the problem. A systematic analysis of a series of candidate NN architectures was conducted for this RSS-based fingerprint localization.

A multilayer perceptron architecture (MLP) with 22 inputs, 2 hidden layers and 2 outputs was the best candidate. 36 and 32 neurons were applied to the hidden layers respectively, see Fig.9

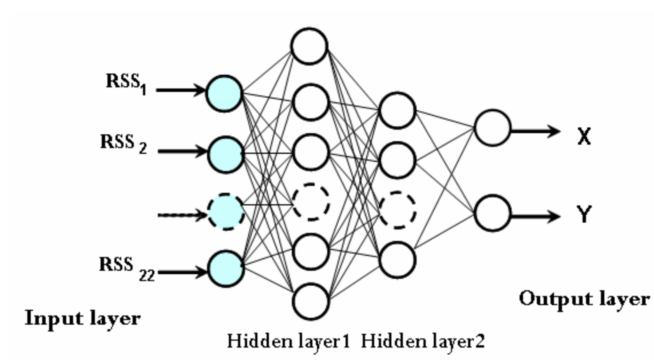

Figure 9: NN architecture used for this positioning issue

Once again, a gradient descent algorithm is chosen to adapt the NN weights for its simplicity. The error goal may not be achieved for three reasons:

A risk of stacking in local minima is present, and in this case is to start over by reinitializing the weights to some new set of small random values.

The network does not have enough degrees of freedom to fit the desired input/output model. In this case, hidden nodes or layers are added and training is restarted.

There is no enough information in the training data to perform the desired mapping.

We have tested our algorithm for 2225 different position locations. The histogram on Fig.10 shows us the positioning error distribution for the case where no Pre-processing was performed, i.e. predicted RSS were used to train the NN without any correction or noise reduction, and for the case where pre-processing was performed.

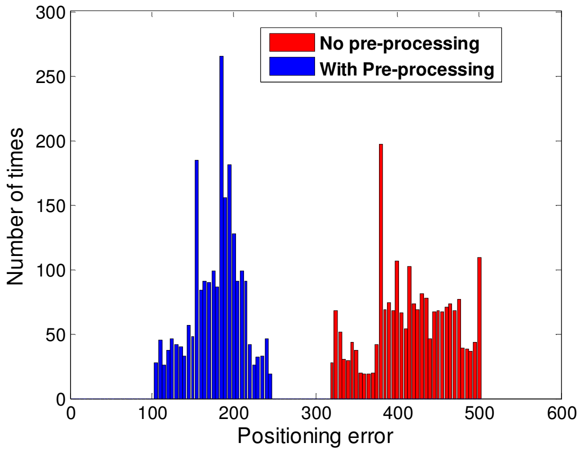

Figure 10: Histogram _Root of mean squared positioning error in meter.

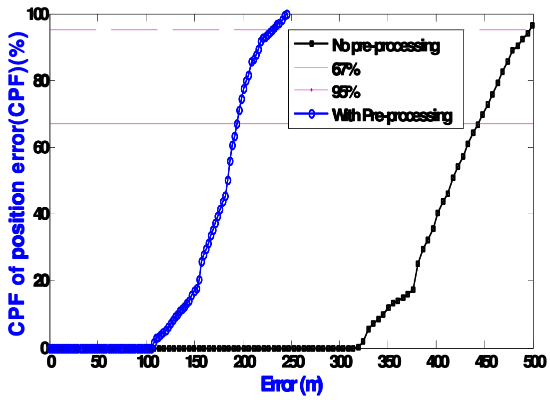

Figure 11: CPF of the positioning error. Comparison of the performances with and without pre-processing

V. Conclusion

A RSS signature-based localization has been applied in this work due to the fact that we are interested in localization within a dense urban area. Concepts solving some problems related to RSS-Signature based localization have been developed throughout this research paper:

We have used predicted RSS for training the NN, to remedy to the major problem in Signature based localization which is the big demanding effort to collect RSS-Data along the streets to realize a database.

A calibration of these predicted RSS has been performed in order to reduce the error difference between the predicted and real signal strengths.

A time delayed NN has been applied to remedy to the issue dealing with stochastic behavior of the RSS.

For a big area, one NN could be trained to cover several cells and according to the Cell Identification, the corresponding trained NN is used in the positioning algorithm.

An analysis of a suitable NN architecture and training algorithms has been undergone.

Results of Fig.10,11 show that the positioning accuracy can be improved significantly by performing the pre-processing of the RSS. A sample of collected RSS used to correct the predicted one is first filtered in order to consider the best estimated value of it. Results show that with 67% of probability, positioning error is less than 175m instead of 420m when no calibration is performed. And with 95%, the positioning error is less than 220m instead of 490m when pre-processing is not performed. Future work will be focused on tracking in case of user mobility and this way the positioning error could be further decreased.

ACKNOWLEDGMENT

We would like to thank Prof. T. Kurner and his research group from the Institute of communication Technology from Brunswick Technical University for providing the RSS prediction data of our concerned area.

REFERENCES

[1] H. Laitinen, J. Lähteenmäki, T. Nordström, "Database Correlation Method for GSM location," presented at VTC, 2001.

[2] R. Yamamoto, H. Matsutani, H. Matsuki, T. Oono, and H. Ohtsuka, "Position Location Technologies using signal strengths i cellular system," presented at VTC, 2001.

[3] C. Nerguizian, C. Despins, and S. Affes, "Indoor Geolocation with received Signal Stregth Fingerprinting Technique and Neural Networks," presented at ICT, 2004.

[4] Z. Salcic and E. Chan, "Mobile station positioning using GSM cellular phone and artificial Neural networks," Wireless personal Communications, vol. 14, pp. 235-254, 2000.

[5] K. R. Anne, K. Kyamakya, F. Erbas, C. Takenga, and J. C. Chedjou, "GSM RSSI-based positioning using extended Kalman filter for training ANN," presented at VTC, Los Angeles, 2004.

[6] C. M. Takenga, K. R. Anne, K. Kyamakya, and J. C. Chedjou, "Comparison of Gradient descent method, Kalman filtering and Decoupled Kalman in training Neural neutworks used for fingerprint-based positioning," presented at IEEE VTC Fall, Los Angeles, 2004.

[7] T. Kurner and A. Meier, "Prediction of outdoor and outdoor-to-indoor coverage in urban areas at 1.8 GHz," ieee on selected areas in communications, vol. 20, 2002.

[8] J. E. Berg, "Building Penetration Loss Along Urban Street Microcells," presented at Cost 231 Final Report, Brussels, 1999.