

POLITEHNICA UNIVERSITY TIMISOARA
ELECTRONICS AND TELECOMMUNICATIONS
FACULTY

**NEURAL NETWORK APPLICATIONS FOR
RADIOCOVERAGE STUDIES IN MOBILE
COMMUNICATION SYSTEMS**

Ph. D. Thesis

Eng. Ileana Popescu

BIBLIOTECA CENTRALA
UNIVERSITATEA "POLITEHNICA"
TIMISOARA

Supervisors: Prof. Dr. Eng. Ioan Nafornită
Prof. Dr. Philip Constantinou

2003

Acknowledgement

I am greatly indebted to the thesis supervisor Professor Ioan Nafoarniță for his encouragement and constructive criticism at all stages of this research.

I am also greatly indebted to the thesis co-supervisor Professor Philip Constantinou for his technical support and constructive criticism at all stages of this research.

I am also grateful to the Rector of University of Oradea, Professor Teodor Maghiar, for giving me the opportunity to conduct part of my research activities in the Mobile Radiocommunications Laboratory at the National Technical University of Athens, Greece.

Special thanks are extended to all my colleagues from the Mobile Radiocommunications Laboratory, NTUA, for all their support, encouragement, help and sometimes patience during the long hours spent together.

Very special thanks are also given to all my friends for keeping my moral up and being next to me during this work.

Abstract

The purpose of the thesis is the applications of Artificial Neural Networks in predicting the propagation path loss for telecommunication systems. The tremendous growth of wireless communication systems and especially mobile radio systems requires radio coverage prediction models that provide accurate results and fast processing time for several types of environments, which includes a large number of parameters describing the outdoor and indoor environment.

Neural Networks models are proposed for the prediction of propagation path loss in different environments (urban, suburban and indoor), through which some important disadvantages of both statistical and deterministic propagation models can be overcome. The proposed Neural Networks models are designed based on propagation measurement results. In order to examine the validity of the Neural Networks models, the predicted path loss by them is compared to the measured values and to the path loss obtained by applying empirical models.

Within the proposed models, environmental characteristics are considered more subtly than in standard statistical models, what usually provides greater accuracy of the model. On the other side, the Neural Network models are not computationally extensive as the deterministic models. The implementation of the proposed Neural Networks models requires a database that is easy to obtain.

Rezumat

Scopul prezentei lucrări este folosirea Rețelelor Neuronale Artificiale pentru predicția pierderilor prin propagare în sistemele de telecomunicații. Dezvoltarea puternică a sistemelor de comunicație fără fir și în special a sistemelor radio mobile necesită modele de predicție a acoperirii radio care să furnizeze rezultate precise și un timp de procesare mic, în cazul diferitelor medii de propagare a undelor radio.

Rețelele Neuronale Artificiale sunt sisteme de procesare a informației care încearcă să imite comportarea creierului uman. Aplicații ale Rețelelor Neuronale Artificiale sunt foarte numeroase. Clasificatori, procesoare de semnal, optimizatoare și regulatoare au fost deja implementate. Deși există numeroase tipuri de Rețelele Neuronale Artificiale, caracteristicile comune ale acestora sunt [Haykin, 94]:

- Imposibilitatea definirii unor formule analitice exacte
- Precizia necesară: de ordinul a câteva procente
- Cantitatea de date procesate: medie
- Adaptabilitate la mediu, care permite rețelelor să învețe de la un mediu în schimbare
- Structură paralelă, ceea ce permite rețelelor să ofere o viteză de calcul mare.

Toate aceste caracteristici ale Rețelelor Neuronale Artificiale recomandă utilizarea lor pentru predicția intensității câmpului în diverse medii de propagare. Predicția intensității câmpului poate fi descrisă ca o transformare a unui vector de intrare conținând informații topografice și morfo-grafice într-o valoare de ieșire dorită. Transformarea necunoscută este o funcție scalară de mai multe variabile (mai multe intrări și o singură ieșire). Datorită influențelor complexe ale mediului natural, funcția de transformare nu poate fi descrisă analitic. Este cunoscută doar în puncte discrete, pentru care există disponibile valori măsurate. sau în cazuri în care condițiile de propagare sunt clar definite, ceea ce permite aplicarea unor reguli simple cum ar fi, de exemplu, propagarea în spațiul liber.

Problema predicției pierderilor prin propagare între două puncte poate fi tratată ca o funcție de mai multe intrări și o singură ieșire. Intrările conțin informații despre pozițiile emițătorului și a receptorului, clădirile înconjurătoare, frecvență, etc, în timp ce ieșirea reprezintă pierderea prin propagare corespunzătoare intrărilor respective.

Din acest punct de vedere, cercetările în modelarea pierderilor prin propagare constă în găsirea intrărilor și a funcției care aproximează cel mai bine pierderea prin propagare.

Capacitatea de aproximare a funcțiilor, caracteristică rețelele neuronale artificiale, justifică utilizarea lor pentru modelarea pierderilor prin propagare.

Rețelele neuronale cu propagare înainte îndeplinesc criteriile de utilizare pentru scopuri de predicție deoarece nu permit nici o reacție dinspre ieșire (intensitatea câmpului sau pierderea prin propagare) spre intrare (date topografice sau morfografice).

Predicția intensității câmpului este o sarcină foarte complexă și dificilă. În cele mai multe cazuri nu există condiții de vizibilitate clară între emițător și receptor. Multe metode de predicție a intensității câmpului au fost propuse în literatură [COST231, 99]. Pentru a îmbunătăți precizia predicției, modelele de predicție trebuie să fie adaptabile la fiecare caz special și aceasta necesită dezvoltarea unor sisteme de predicție cu o flexibilitate structurală ridicată.

În general, modelele de predicție sunt empirice (denumite și statistice) sau teoretice (deterministice), sau o combinație a celor două. În timp ce modelele empirice se bazează pe măsurători, modelele teoretice se bazează pe principiile fundamentale ale fenomenelor de propagare a undelor radio.

Principalul avantaj al modelelor empirice este faptul că toate influențele mediului sunt luate în considerare implicit. Pe de altă parte, precizia modelelor empirice depinde nu doar de precizia măsurătorilor dar și de asemănările dintre mediul analizat și cel în care au fost efectuate măsurătorile.

Datorită faptului că modelele deterministice sunt bazate pe principiile fizicii, ele pot fi aplicate în diferite medii, fără a afecta precizia predicției. De obicei, implementarea lor necesită o bază de date mare a caracteristicilor mediului, condiție uneori imposibil de obținut. Algoritmii acestor modele sunt de obicei foarte complecși și necesită un timp mare de calcul. Datorită acestui fapt, implementarea modelelor deterministice este de obicei restrânsă la zone de dimensiuni mici ale celulelor micro- sau medii interioare. Totuși, în cazul în care modelele teoretice sunt corect implementate, precizia obținută poate fi mai mare decât în cazul modelelor empirice.

Problema principală a modelelor empirice clasice este precizia nesatisfăcătoare, în timp ce modelele deterministice necesită un timp de calcul ridicat.

Prezenta lucrare prezintă rezultatele cercetării în domeniul aplicațiilor bazate pe Rețele Neuronale Artificiale pentru predicția pierderilor prin propagare în diferite medii (urban, suburban și interior). Modelele propuse sunt următoarele:

- Modelul de predicție MLP - NN în mediul urban,
- Modelul de predicție MLP - NN în mediul suburban,
- Modele hibride de predicție MLP - NN în mediile urban și suburban,
- Modelul de predicție MLP - NN în mediul interior,
- Modelul de predicție RBF - NN în mediul urban,
- Modelul de predicție RBF - NN în mediul suburban,
- Modele hibride de predicție RBF - NN în mediile urban și suburban,
- Modelul de predicție RBF - NN în mediul interior.

În sub-capitolele 7.2.2 – 7.2.5 sunt comparate rețele neuronale de tip MLP (Perceptron Multistrat) cu diferite structuri și diferiți algoritmi de antrenare. Într-o primă fază, este investigată comportarea rețelei MLP - NN antrenată cu algoritmul Levenberg-Marquardt, cu număr diferit de neuroni în unul sau două straturi ascunse. În urma acestor simulări, este stabilită rețeaua neuronală cu configurația optimă iar în continuare, sunt investigate performanțele algoritmului Resilient Backpropagation și a versiunii Powell-Beale a algoritmului gradientului conjugat.

În sub-capitolele 7.3.1 – 7.3.4 sunt analizate modele bazate pe rețele neuronale cu funcții de bază radiale (Generalized Radial Basis Function Neural Networks RBF-NN). Performanțele tuturor modelelor RBF-NN cu diferiți parametri de intrare sunt evaluate prin compararea statisticilor erorilor de predicție, în funcție de eroarea medie absolută, deviația standard, eroarea medie pătratică și coeficientul de corelație dintre valoarea prezisă și cea măsurată.

În modelele propuse, bazate pe rețele neuronale artificiale, caracteristicile mediului sunt luate în considerare de maniera mult mai subtilă decât în modelele statistice obișnuite, ceea ce în general duce la o precizie mai mare a predicției. Pe de altă parte, modelele propuse nu necesită timp mare de calcul iar implementarea lor necesită o bază de date ușor de obținut.

În comparație cu alte modele de predicție investigate, modelele propuse, bazate pe rețele neuronale prezintă o precizie foarte bună. Avantajul principal al acestor modele constă în faptul că sunt ușor de adaptat la medii specifice și condiții de propagare complexe. În unele cazuri, precizia poate fi îmbunătățită prin antrenări suplimentare ale rețelei neuronale.

Rezultatele sunt caracterizate întotdeauna de o anumită deviație, dar precizia obținută este suficientă pentru predicție.

Rezultatele obținute de un sistem bazat pe rețeaua MLP-NN, folosit pentru predicție, sunt interesante. Dar rezultatele obținute de un sistem MLP-NN, care combină într-o abordare neliniară rezultate ale algoritmilor clasici pentru pierderea prin propagare și date fizice, deschid noi căi de investigare. Algoritmii conțin o cantitate considerabilă de cunoștințe-expert cu privire la propagarea undelor terestre; utilizarea rețelelor neuronale artificiale pentru predicția intensității câmpului permite integrarea eficientă a acestor cunoștințe precum și a altor informații, de exemplu informații topografice.

Analiza modelării hibride a predicției pierderilor prin propagare arată că rețelele neuronale pot fi utilizate în modele adaptive. Prin introducerea unor parametri aditionali, este posibilă chiar o extensie a modelelor empirice. În contrast cu algoritmi de autoregresie, rețelele neuronale oferă multiple avantaje datorită abilității lor de a reprezenta dependențe cu o neliniaritate ridicată a mai multor parametri simultan, incluzând informații care nu pot fi tratate analitic. În plus, aplicarea simultană a tuturor informațiilor disponibile conduce la rezultate bune chiar și în cazul în care se dispune de o bază de date săracă. Se arată că această abordare flexibilă și cu randament mare de calcul poate fi utilizată pentru calibrare și ca o extensie a modelelor de predicție convenționale.

Avantajul utilizării rețelelor neuronale este acela că un model de propagare particular poate fi proiectat să ia în considerare tipuri diferite de medii, pe baza măsurătorilor efectuate în mediul dorit. Rezultatele simulărilor au arătat că abordarea bazată pe utilizarea rețelelor neuronale furnizează o predicție mai precisă a intensității câmpului decât cea a modelelor empirice analizate în această lucrare.

Contents

1. Introduction	
1.1 Motivation.....	1
1.2 Thesis contribution.....	3
2. Neural Networks	
2.1. Introduction.....	6
2.1.1. Definition.....	6
2.1.2. Benefits of neural networks.....	7
2.1.3. The model of a neuron.....	9
2.1.4. Knowledge representation.....	13
2.1.5. Learning processes.....	14
2.1.6. Supervised learning.....	18
2.1.7. Unsupervised learning.....	20
2.1.8. Function approximation.....	20
2.2. The perceptron.....	22
2.2.1. Introduction.....	22
2.2.2. Perceptron convergence theorem.....	24
2.3. Multilayer perceptron.....	25
2.3.1. Introduction.....	25
2.3.2. Training algorithms for MLP.....	27
2.3.2.1. The backpropagation (BKP) algorithm.....	28
2.3.2.2. Advanced learning algorithms in MLP.....	33
A. Heuristic improvements of the BKP algorithm.....	34
B. Conjugate gradient algorithms.....	35
C. Quasi – Newton algorithms.....	36
D. Levenberg-Marquardt algorithm.....	37
2.3.3. Generalization.....	37
2.3.4. Cross – validation.....	38
2.4. Radial Basis Functions.....	40
2.4.1. Introduction.....	40
2.4.2. Structure of RBF networks.....	42
2.4.3. Radial basis functions.....	43
2.4.4. Learning strategies with RBF-NN.....	44
2.4.5. A RBF-NN algorithm.....	52
2.4.6. Issues with RBF-NN learning.....	53
2.4.7. The General Regression Neural Network.....	54
2.4.8. Comparison of RBF networks and MLP.....	56
3. Mobile radio channels	
3.1. Introduction.....	57
3.2. Representation of a mobile radio signal.....	58
3.3. Fading.....	58
3.4. Obtaining meaningful propagation loss data from measurements.....	59
3.5. Modeling requirements.....	63
4. Propagation mechanisms for mobile communication systems	
4.1. Introduction.....	64
4.2. Propagation in free space.....	66
4.3. Reflection.....	67
4.3.1. The Fresnel reflection coefficients.....	67
4.3.2. Ground reflection (2-ray) model.....	68
4.4. Diffraction over irregular terrain.....	70
4.4.1. Fresnel zone geometry.....	72
4.4.2. Diffraction losses.....	73
4.5. Scattering.....	77

4.6. Propagation mechanisms in ray theory.....	79
5. Propagation Prediction Models	
5.1. General Considerations.....	83
5.2. Propagation models for macro-cells.....	84
5.2.1. The model of Okumura.....	85
5.2.2. Hata prediction model.....	87
5.2.3. The Egli's model.....	88
5.2.4. COST 231-Hata model.....	89
5.2.5. COST 231-Walfisch-Ikegami model.....	90
5.2.6. Walfisch and Bertoni model.....	93
5.2.7. Xia model.....	93
5.2.8. Sakagami – Kuboi model.....	95
5.2.9. The Log-distance path loss model.....	96
5.2.10. Discussions.....	97
5.3. Micro – cell propagation models.....	98
5.3.1. Model 1. Two–ray model.....	98
5.3.2. Model 2.....	99
5.3.3. Model 3. Wideband PCS model.....	100
5.3.4. Model 4. Lee model.....	101
5.3.5. COST 231 models.....	104
5.3.6. Model based on UTD and Multiple Image Theory.....	105
5.3.7. Discussions.....	106
5.4. Indoor propagation models.....	108
5.4.1. Indoor radio propagation environment.....	108
5.4.2. Empirical narrowband models.....	109
5.4.2.1. Model 1.....	109
5.4.2.2. Model 2.....	110
5.4.2.3. Model 3. Floor Attenuation Factor model.....	110
5.4.2.4. Model 4. The COST 231 – Motley model.....	111
5.4.2.5. Model 5. Lafortune model.....	112
5.4.2.6. Model 6. The Multi-wall model.....	113
5.4.3. Deterministic models.....	114
5.4.3.1. Ray-launching model (RLM).....	114
5.4.3.2. Ray Tracing method.....	116
6. Neural networks applications for propagation prediction	
6.1. Outdoor environment.....	117
6.2. Indoor environment.....	123
7. Proposed neural network models	
7.1. The measurements.....	129
7.2. Proposed MLP-NN models for the prediction of propagation path loss.....	131
7.2.1. MLP implementation of Hata's formula and knife-edge diffraction model.....	133
7.2.2. Proposed MLP-NN models for propagation prediction in urban environment.....	135
7.2.3. Proposed MLP-NN models for propagation prediction in suburban environment.....	142
7.2.4. Hybrid model based on MLP-NN.....	146
7.2.5. Proposed MLP-NN models for propagation prediction in indoor environment.....	153
7.3. Proposed RBF-NN models for the prediction of propagation path loss.....	156
7.3.1. Proposed RBF-NN models for propagation prediction in urban environment.....	156
7.3.2. Proposed RBF-NN models for propagation prediction in suburban environment.....	159
7.3.3. Hybrid models based on RBF-NN.....	161

7.3.4. Proposed RBF-NN model for propagation prediction in indoor environment.....	163
7.4. Discussions.....	165
7.5. Conclusions.....	170
7.6. Further work.....	171
Appendix 1.....	172
Appendix 2.....	173
Appendix 3.....	174
Appendix 4.....	176
References.....	178
Published papers.....	186
CV.....	188

List of figures

1.1.	Propagation loss prediction as a function of several inputs.....	2
2.1.	The nonlinear model of a neuron.....	9
2.2.	Affine transformation produced by the presence of a bias.....	10
2.3.	Types of activation functions: (a) - the threshold function; (b) – the piecewise linear function; (c) - the logistic function and (d) - the hyperbolic tangent function.....	12
2.4.	Error-correction learning.....	14
2.5.	Learning with a teacher.....	19
2.6.	Unsupervised learning.....	20
2.7.	System identification.....	21
2.8.	Inverse system modeling.....	21
2.9.	The perceptron.....	23
2.10.	Decision boundary for a two classes pattern classification problem.....	24
2.11.	Multilayer perceptron with a single hidden layer.....	26
2.12.	The capability of the MLP to design complex decision boundaries: (a)-single perceptron; (b) – MLP with one hidden layer; (c) – MLP with two hidden layers.....	27
2.13.	Illustration of the early stopping rule based on cross validation.....	40
2.14.	Architecture of a RBF-NN.....	43
2.15.	The SOFM-NN with two input nodes and output nodes organized in a two-dimensional array.....	48
2.16.	General regression neural network.....	55
3.1.	Obtaining the local mean.....	60
4.1.	Propagation over a plane earth.....	68
4.2.	Two-ray ground reflection model.....	69
4.3.	Knife-edge diffraction geometry when the transmitter and the receiver are not at the same height.....	71
4.4.	The knife-edge diffraction equivalent geometry. The point T denotes the transmitter and R denotes the receiver, with an infinite knife-edge obstruction blocking the line-of-sight path.....	71
4.5.	Family of ellipsoids defining the first three Fresnel zones around the transmitter and the receiver of a radio path.....	72
4.6.	Knife – edge diffraction: (a) h and v positive, (b) h and v negative.....	74
4.7.	Knife – edge diffraction with ground reflections.....	75
4.8.	Diffraction over a cylinder.....	76
4.9.	Surface roughness criterion.....	77
5.1.	Typical propagation situation in urban areas and definition of the parameters used in the COST231-Walfisch-Ikegami model.....	90
5.2.	Definition of the street orientation.....	91
5.3.	Radio propagation paths and geometrical parameters.....	95
5.4.	Two-ray model.....	98
5.5.	The propagation mechanism of low-antenna height at the cell site.....	102
5.6.	(a) 2-D view of the reception sphere; (b)- Ray launching.....	115
5.7.	Ray tracing.....	116
7.1.	The building topology and the transmitter positions.....	131
7.2.	Comparison between predicted values achieved by neural network and knife-edge diffraction model versus the distance between the diffraction point and receiver.....	134
7.3.	Performance of MLP-NN on the test set, with different number of neurons in 1 and 2 hidden layers, 5 inputs, LOS case, urban environment.....	136
7.4.	Performance of MLP-NN on the test set, with RP and PB training algorithms, 5-10-10-1, LOS case, urban environment.....	136
7.5.	Measured and predicted path loss for LOS case in urban environment.....	137

7.6	Performance of MLP-NN on the test set, with different number of neurons in 1 and 2 hidden layers, 6 inputs, NLOS case, urban environment.....	139
7.7	Performance of MLP-NN on the test set, LM algorithm, 11 inputs, different number of neurons in 1 and 2 hidden layers, NLOS case, urban environment.....	139
7.8	Performance of MLP-NN on the test set, with RP and PB training algorithms, 6-16-16-1, NLOS case, urban environment.....	140
7.9	Performance of MLP-NN on the test set, RP and PB training algorithms, 11-16-16-1, NLOS case, urban environment.....	140
7.10	Measured and predicted path loss for NLOS case with $h_{BS} < h_{roof}$, urban case.....	141
7.11	Measured and predicted path loss for NLOS condition with $h_{BS} > h_{roof}$, urban case.....	142
7.12	Performance of MLP-NN on the test set, LM algorithm, different number of neurons in 1 and 2 hidden layers, 5 inputs, suburban environment.....	143
7.13	Performance of MLP for LM training algorithm, different number of neurons in 1 and 2 hidden layers, 7 inputs, suburban environment.....	144
7.14	Performance of MLP-NN, 5-12-12-1, RP and PB algorithms, suburban environment.....	144
7.15	Performance of MLP-NN, 7-12-12-1, RP and PB algorithms, suburban environment.....	145
7.16	Comparison between the measured and the predicted propagation path loss by the neural model and the CWI model for one particular route.....	145
7.17	The schematic diagram of the training process.....	146
7.18	The schematic diagram of the prediction.....	147
7.19	Performance of hybrid MLP-NN, LM algorithm, 1 and 2 hidden layers, 5 inputs, urban environment.....	147
7.20	Performance of hybrid MLP-NN, LM algorithm, 1 and 2 hidden layers, 6 inputs, urban environment.....	148
7.21	Performance of hybrid MLP-NN, 5-16-16-1, PB and RP training algorithms, different number of epochs.....	148
7.22	Performance of hybrid MLP-NN, 6-16-16-1, PB and RP training algorithms, urban environment.....	149
7.23	The comparison between the prediction made by the proposed error correction model, CWI model and measurements in case of a particular route, urban environment.....	150
7.24	Performance of hybrid MLP-NN, LM algorithm, 4 inputs, 1 and 2 hidden layers, suburban environment.....	151
7.25	Performance of hybrid MLP-NN, LM algorithm, 5 inputs, 1 and 2 hidden layers, suburban environment.....	151
7.26	Performance of hybrid MLP-NN, RP and PB algorithms, 4-17-17-1, suburban environment.....	151
7.27	Performance of hybrid MLP-NN, 5-17-17-1, RP and PB algorithms, suburban environment.....	152
7.28	The comparison between the prediction made by the proposed hybrid model, CWI model and measurements for a particular route, suburban case.....	152
7.29	Performance of MLP-NN, 10 inputs, LM training algorithm, 1 and 2 hidden layers, indoor environment.....	154
7.30	Performance of MLP-NN, RP and PB algorithms, 10-14-14-1, indoor environment.....	154
7.31	Comparison between predictions and measurements with the transmitter in position 2 and the receiver located in sector A, along the main corridor: a) normalized received power, b) path loss.....	156
7.32	Measured and predicted path loss for LOS case, urban environment.....	157
7.33	Measured and predicted path loss, in case of a particular route, NLOS case, urban environment.....	159
7.34	Measured and predicted propagation path loss by the proposed RBF 7 model and the CWI model, suburban environment.....	160
7.35	Prediction made by the proposed RBF11 hybrid model, CWI model and measurements, urban environment.....	162
7.36	Prediction made by the proposed hybrid model RBF13, CWI model and measurements, suburban case.....	163
7.37	Comparison between predictions and measurements with the transmitter in position 2 and the receiver located in sector A, along the main corridor: a) normalized received power; b) path loss.....	164

List of tables

3.1.	σ_m versus $2L$	61
5.1.	Definition of cell type.....	83
5.2.	Parameters for the wideband microcell model at 1900 MHz.....	101
5.3.	Wall types for the Multi-Wall model.....	114
7.1.	Comparison between the proposed MLP-NN model and the other empirical models, LOS case.....	137
7.2.	Comparison between the prediction models in LOS case, for a particular route.....	138
7.3.	Comparison between the MLP-NN approach and the other empirical models in NLOS case.....	141
7.4.	MLP-NN models and other empirical models in NLOS case, for $h_{BS} > h_{roof}$ and $h_{BS} < h_{roof}$	142
7.5.	Comparison between the proposed MLP-NN model and other empirical models, suburban environment.....	146
7.6.	Comparison between the proposed hybrid MLP-NN based approach and the CWI model in urban environment.....	149
7.7.	Comparison between the proposed hybrid MLP-NN model and CWI model, suburban environment.....	153
7.8.	Results of the prediction, MLP-NN, indoor environment.....	155
7.9.	Proposed RBF-NN model and other empirical models, LOS case, urban environment.....	157
7.10.	RBF-NN models for NLOS case, urban environment.....	158
7.11.	Proposed RBF3 model and other empirical models, NLOS case, urban environment.....	158
7.12.	Generalized RBF-NN models, suburban environment	160
7.13.	Proposed RBF7 model and the other empirical models suburban case.....	160
7.14.	Hybrid RBF-NN models for NLOS case, urban environment.....	161
7.15.	Proposed RBF11 model and the CWI model.....	161
7.16.	Hybrid RBF-NN models, suburban case.....	162
7.17.	Comparison between the proposed RBF13 hybrid model and the CWI model.....	162
7.18.	Results of the prediction, indoor environment.....	163
7.19.	Proposed NN models performance in different environments.....	166
7.20.	Background studies, outdoor environment.....	168
A.1	Summary of the LMS algorithm.....	177

1. Introduction

1.1. Motivation

The purpose of the thesis is the applications of Artificial Neural Networks (ANN) in predicting the propagation path loss for telecommunication systems. The tremendous growth of wireless communication systems and especially mobile radio systems requires radio coverage prediction models that provide accurate results and fast processing time for several types of environments, which includes a large number of parameters describing the outdoor and indoor environment.

During last years Artificial Neural Networks (ANN) have experienced a great development. Artificial neural networks are information processing systems that aim to copy the behavior of human brain. ANN applications are already very numerous. Classifiers, signal processors, optimizers and controllers have already been realized. Although there are several types of ANN's all of them share the following features [Haykin, 94]:

- Exact analytical formula impossible
- Required accuracy: some percent
- Quantity of data to process: medium
- Environment adaptation that allows them to learn from a changing environment (different terrain databases and terrain)
- Parallel structure that allows them to achieve high computation speed.

These characteristics of ANN's make them suitable for predicting field strength in different environments. The prediction of field strength can be described as the transformation of an input vector containing topographical and morphographical information (e.g. path profile) to the desired output value. The unknown transformation is a scalar function of many variables (several inputs and a single output), because a huge amount of input data has to be processed. Owing to the complexity of the influences of the natural environment, the transformation function cannot be given analytically. It is known only at discrete points where measurement data are available or in cases with clearly defined propagation conditions which allow to apply simple rules like free space propagation, etc.

The problem of predicting propagation loss between two points may be seen as a function of several inputs and a single output. The inputs contain information about the transmitter and receiver locations, surrounding buildings, frequency, etc while the output gives the propagation loss for those inputs (figure 1.1).

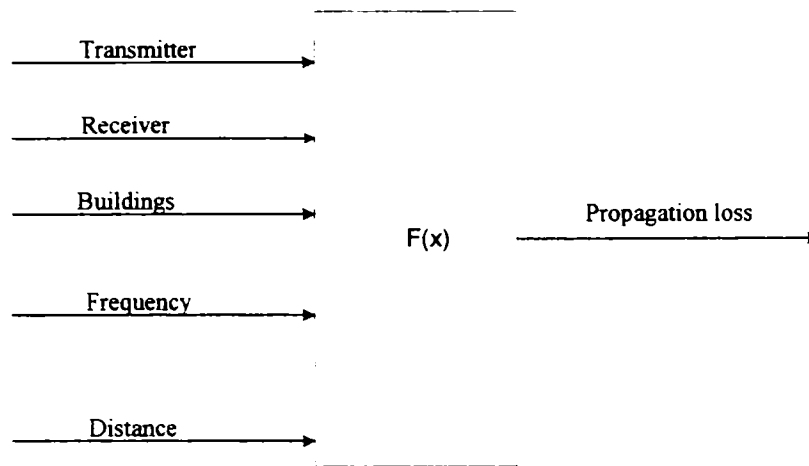


Figure 1.1. Propagation loss prediction as a function of several inputs

From this point of view, research in propagation loss modeling consists in finding both the inputs and the function $F(x)$ that best approximate the propagation loss. Given that ANN's are capable of function approximation, they are useful for the propagation loss modeling.

The feedforward neural networks are very well suited for prediction purposes because do not allow any feedback from the output (field strength or path loss) to the input (topographical and morphographical data).

The prediction of field strength level is a very complex and difficult task. In most cases, there are no clear line-of-sight (LOS) conditions between the transmitter and the receiver. Many field strength prediction methods have been proposed in the literature [COST231, 99]. Usual databases include classification of land usage and urban areas but a lot of questions still remain. Cities and open ranges are quite different in structure and may also be different in the way of database classification. Propagation models should be adapted to every special case to improve accuracy. As a consequence, the development of prediction systems with a high structural flexibility it is very desirable.

Generally, the prediction models can be either empirical (also called statistical) or theoretical (also called deterministic), or a combination of these two. While the empirical models are based on measurements, the theoretical models deal with the fundamental principles of radio wave propagation phenomena.

In the empirical models all environmental influences are implicitly taken into account regardless of whether or not they can be separately recognized. This is the main advantage of these models. On the other hand, the accuracy of the empirical models depends not only on the accuracy of the measurements but also on similarities between the environment to be analyzed and the environment where the measurements are carried out.

The deterministic models are based on the principles of physics and due to that, can be applied in different environments without affecting the accuracy. Their implementation usually requires a great database of environmental characteristics that is sometimes impossible to obtain. The algorithms of these models are usually very complex and lack in the computational efficiency. Due to that, the implementation of the deterministic models is usually restricted to the smaller areas of micro-cell or indoor environments. However, if the deterministic models are implemented correctly, greater accuracy of the prediction can be expected than in the case of empirical models.

The main problem of the classical empirical models is the unsatisfactory accuracy, while the theoretical models lack in computational efficiency. On brief, characteristics of a field strength prediction system for mobile radio can be summarized as follows:

- Exact analytical formula impossible
- Required accuracy: some percent (around 6 dB in field strength level)
- Quantity of data to process: medium
- Flexibility to adapt to different terrain databases and terrain.

1.2. Thesis contributions

This thesis presents the results of the research in the area of Neural Networks (NN) applications for prediction of propagation path loss in different environments (urban, suburban and indoor). The proposed NN models are the follows:

- MLP - NN prediction model in urban environment,
- MLP - NN prediction model in suburban environment,
- Hybrid MLP - NN models in urban and suburban environment,
- MLP - NN model for indoor environment,
- RBF - NN prediction model in urban environment,
- RBF - NN prediction model in suburban environment,
- Hybrid RBF - NN models in urban and suburban environment,

➤ RBF - NN model for indoor environment.

In section 7.2.2 – 7.2.5 a number of comparisons are made for Multilayer Perceptron Neural Networks (MLP-NN) with different architectures and different training algorithms. At a first step, the performance of the MLP-NN trained with Levenberg-Marquardt (LM) algorithm, with different numbers of neurons in one and two hidden layer was investigated. These simulations were done with the use of early stopping method. Following these simulations, the MLP-NN with the optimum configuration is established and investigations are done on the performance of two other training algorithms: the Resilient Backpropagation (RP) and the Powell-Beale (PB) version of the conjugate gradient algorithm.

In section 7.3.1 – 7.3.4 a number of Generalized Radial Basis Function Neural Networks (RBF-NN) models are studied. The performance of all RBF-NN models with different input parameters is evaluated by comparing their prediction error statistics based on absolute mean error, standard deviation, root mean square error and the correlation between predicted values and measurement data.

Within the proposed models, environmental characteristics are considered more subtly than in standard statistical models, what usually provides greater accuracy of the model. On the other side, the NN models are not computationally extensive as the deterministic models. The implementation of the proposed NN models requires a database that is easy to obtain.

In comparison with other field strength prediction models, the proposed NN models showed very good accuracy. The main advantage of the proposed NN models lays in the fact that the models should be easily adjusted to some specific environments and complex propagation condition. In more specific local cases, the accuracy can be improved by some additional NN training. Results are always connected with some uncertainty but accuracy may be sufficient for prediction purpose.

The results obtained by a pure MLP-NN system used for prediction are very interesting. But the results obtained by a MLP-NN system that combines a nonlinear NN approach, results of classical propagation loss algorithms and physical data open new ways of investigations. The algorithms carry a considerable expert knowledge on terrestrial wave propagation; the use of NN as field strength prediction model allows to efficiently integrate this knowledge as well as topographic and land cover information.

The hybrid modeling approach for the prediction of propagation path loss is studied and it is shown that NN can be used in highly adaptive models. By introducing of additional parameters during the training process even an extension of empirical models is feasible. In contrast to well-known regression algorithms, NNs offer many advantages owing their ability

to represent highly nonlinear dependencies of many parameters simultaneously, including information that cannot be treated analytically. In addition, the application of all available information at the same time is a way of getting the most even from poorly defined databases. It is shown that this flexible and computationally effective approach can be used for calibration and as an extension of conventional prediction models.

The advantage of the NN approaches is that a particular propagation model can be constructed to take account of various types of environments based on measurement data taken in the desired environment. This approach enhances the flexibility of the NN based prediction model to adapt to the terrain database of the environment. Simulation results have shown that the NN approach provides more accurate prediction of field strength loss than that of the empirical models studied in this work. This verifies the effectiveness of the best approximation capability of the NN.

2. Neural Networks

2.1. Introduction

In this chapter, the fundamental characteristics of artificial neural networks are briefly presented. The next sections are focused on the feedforward neural networks known as Multi-layer Perceptron (MLP) and Radial Basis Functions (RBF) networks.

2.1.1. Definition

Artificial neural networks (ANN), commonly referred to as neural networks (NN), can be defined as a large number of units (also called neurons) organized in different layers that are interconnected. These units are simple processors that operate only on their local data and on the inputs they receive via the connections. It is interesting to note that this model stems from the recognition that the brain operates in a completely different manner than conventional digital computers. The analogy between ANN and the human brain has been summarized in [Haykin, 94] as follows:

“A neural network is a massively parallel-distributed processor made up of simple processing units, that has a natural propensity for storing exponential knowledge and making it available for use. It resembles the brain in two aspects:

- 1. Knowledge is acquired by the network from its environment through a learning process.*
- 2. Interneuron connection strengths, known as synaptic weights, are used to store the acquired knowledge”.*

The procedure used to perform the learning process is called learning algorithm, the function of which is to modify the synaptic weights of the network in an orderly fashion to attain a desired design objective. The modification of synaptic weights provides the traditional method for the design of neural networks. It is also possible for a neural network to modify its own topology, which is motivated by the fact that neurons in the human brain can die and new synaptic connections can grow [Haykin, 99].

In [Schalkoff, 97] the following definition of artificial neural network is given:

“A structure (network) composed of a number of interconnected units (artificial neurons). Each unit has an input/output (I/O) characteristic and implements a local computation or function. The output of any unit is determined by its I/O characteristics, its

interconnection to other units, and (possibly) external inputs. Although 'hand crafting' of the network is possible, the network usually develops an overall functionality through one or more forms of training".

Neural networks are in fact a diverse family of networks. The overall function or functionality achieved is determined by the network topology, the individual neuron characteristics, the learning (or training) strategy and training data.

2.1.2. Benefits of neural networks

A neural network derives its computing power through its structure and its ability to learn, and therefore to generalize. Generalization can be defined as the ability of the trained neural network to produce reasonable outputs for inputs not encountered during the training process. The use of neural networks gives the following properties and capabilities [Haykin, 94]:

1. Nonlinearity. A neuron can be linear or nonlinear. A neural network formed by interconnections of nonlinear neurons, is itself nonlinear.
2. Input-output mapping. A popular paradigm of learning, called learning with a teacher or supervised learning, involves modification of the synaptic weights of a neural network by applying a set of labeled training examples or task examples. Each example consists of a unique input signal and a corresponding desired response. The network is presented with an example picked at random from the set and the synaptic weights (free parameters) of the network are modified to minimize the difference between the desired response and the actual response of the network produced by the input signal in accordance with an appropriate statistical criterion. The training of the network is repeated for many examples in the set until the network reaches a steady state where there are no further significant changes in the synaptic weights. The previously applied training examples may be reapplied during the training session but in a different order. Thus the network learns from the examples by constructing an input-output mapping for the problem at hand.
3. Adaptivity. Neural networks have a built-in capability to adapt their synaptic weights to changes in the surrounding environment. In particular, a neural network trained to operate in a specific environment can be easily retrained to deal with minor changes in the operating environmental conditions. Moreover, when it is operating in a non-stationary environment, a neural network can be designed to change its synaptic

weights in real time. It should be emphasized, however, that the adaptivity does not always lead to robustness; indeed, it may do the opposite.

4. Evidential response. In the context of pattern classification, a neural network can be designed to provide information not only about which particular pattern to select but also about the confidence in the decision made. This latter information may be used to reject ambiguous patterns, should they arise, and thereby improve the classification performance of the network.
5. Contextual information. Knowledge is represented by the structure and the activation state of the network. Every neuron in the network is potentially affected by the global activity of all other neurons in the network. Consequently, a neural network deals with contextual information naturally.
6. Fault tolerance. A neural network, implemented in hardware form, has the potential to be fault tolerant or capable of robust computation. For example, if a neuron or its connecting links are damaged, due to the distributed nature of information stored in the network, the damages have to be extensive before the overall response of the network is degraded seriously. Thus, in principle, a neural network exhibits a graceful degradation in performance rather than catastrophic failure. In order to be assured that the neural network is in fact fault tolerant, it may be necessary to take corrective measures in designing the algorithm used to train the network.
7. VLSI implementability. Due to its massively parallel nature, a neural network may be fast for the computation of certain task. This feature makes neural networks well suited for implementation using very-large-scale-integrated (VLSI) technology.
8. Uniformity of analysis and design. Basically, neural networks enjoy universality as information processors. Neurons, in one form or another, represent an ingredient common to all neural networks. This commonality makes it possible to share theories and learning algorithms in different applications of neural networks. Modular networks can be built through a seamless integration of modules.
9. Neurobiological analogy. The design of a neural network is motivated by the analogy with the brain. Neurobiologists look to (artificial) neural networks as a research tool for the interpretation of neurobiological phenomena. Engineers look to neurobiology for new ideas to solve problems more complex than those based on conventional hard-wired design techniques.

2.1.3. The model of a neuron

An elementary neuron with m inputs is shown in Figure 2.1. The three basic elements of the neural model are:

1. A set of synapses or connecting links, each of which is characterized by a weight of its own. A signal x_j ($j=1, \dots, m$) at the input of synapse j connected to neuron k is multiplied by the synaptic weight w_{kj} .
2. An adder for summing the input signals weighted by the respective synapses of the neuron. This operation constitutes a linear combiner.
3. An activation function for limiting the amplitude of the output of the neuron. Neurons may use any differentiable transfer function f to generate their output.

The model depicted in Figure 2.1 also includes an externally applied bias, noted b_k , that has the effect of increasing or lowering the net input of the activation function, depending on whether it is positive or negative, respectively.

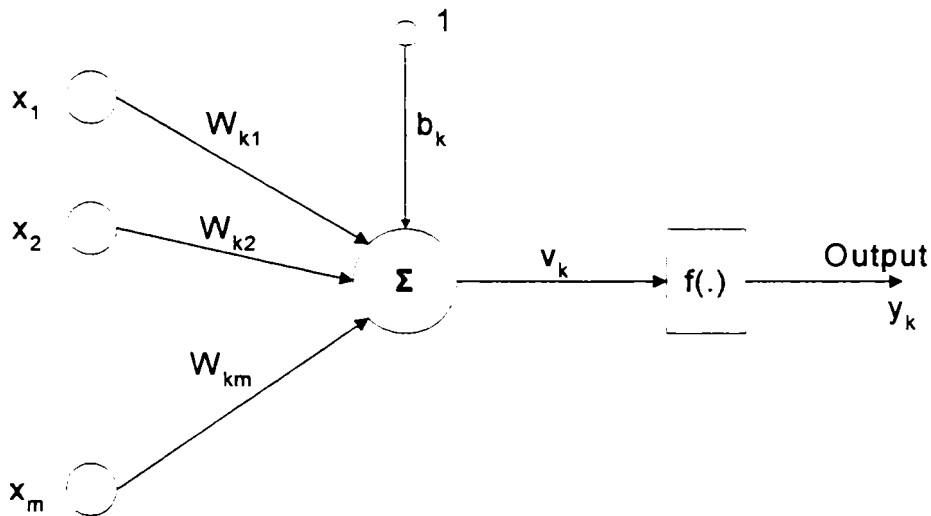


Figure 2.1. The nonlinear model of a neuron

In mathematical terms:

$$u_k = \sum_{j=1}^m w_{kj} x_j \quad (2.1)$$

and

$$y_k = f(u_k + b_k) \quad (2.2)$$

where x_j ($j = 1, 2, \dots, m$) are the input signals, w_{kj} are the synaptic weights of the neuron k , m is the number of the inputs, u_k is the linear combiner output due to the input signals, b_k is the bias, $f(\cdot)$ is the activation function and y_k is the output signal of the neuron.

The bias b_k is an external parameter of the neuron k . The use of bias b_k has the effect of applying an affine transformation to the output u_k of the linear combiner in the model of Figure 2.1, as shown by

$$v_k = u_k + b_k \tag{2.3}$$

In particular, depending on whether the bias b_k is positive or negative, the relationship between the induced local field or activation potential v_k of neuron k and the linear combiner output u_k is modified as illustrated in Figure 2.2.

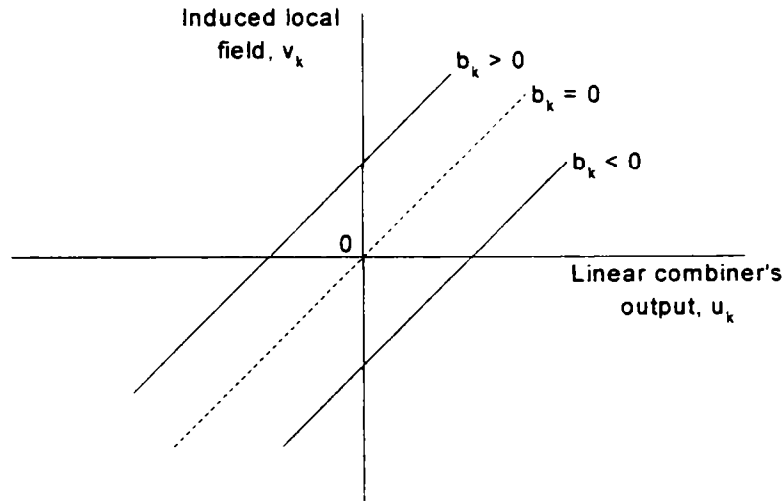


Figure 2.2. Affine transformation produced by the presence of a bias

It is possible to reformulate:

$$v_k = \sum_{j=0}^m w_{kj} x_j \tag{2.4}$$

and

$$y_k = f(v_k) \tag{2.5}$$

In equation (2.4) it was added a new synapse with the input $x_0 = 1$ and the weights $w_{k0} = b_k$.

Types of activation functions

The activation function $f(v)$ defines the output of a neuron in terms of the induced local field v . In this section are presented several basic types of activation functions.

1. The threshold function (Figure 2.3a):

$$f(v) = \begin{cases} 1 & v \geq 0 \\ 0 & v < 0 \end{cases} \tag{2.6}$$

In engineering literature this form of threshold function is commonly referred to as a Heaviside function. Such a neuron, whose activation function is the threshold function, is

referred to in the literature as the McCulloch-Pitts model. In this model, the output of the neuron takes on the value of 1 if the induced local field of that neuron is nonnegative, and 0 otherwise. This statement describes the all-or-none property of the McCulloch-Pitts model [Haykin, 99].

2. The piecewise linear functions (Figure 2.3b):

$$f(v) = \begin{cases} 1 & v \geq \frac{1}{2} \\ v & -\frac{1}{2} < v < \frac{1}{2} \\ 0 & v \leq -\frac{1}{2} \end{cases} \quad (2.7)$$

where the amplification factor inside the linear region of operation is assumed to be unity. The following two situations may be viewed as special forms of the piecewise linear function [Haykin, 99]:

- A linear combiner arises if the linear region of operation is maintained without running into saturation.
- The piecewise-linear function reduces to a threshold function if the amplification factor of the linear region is made infinitely large.

3. The sigmoid function is the most common form of activation function used in the design of artificial neural networks. Sigmoid functions are defined by the following characteristics [Haykin, 94]:

- Strictly increasing functions,
- Asymptotically limited,
- Smoothness.

Two sigmoid functions are of particular interest for neural network implementation. First, the logistic function, depicted in Figure 2.3c and defined by

$$f(v) = \frac{1}{1 + \exp(-a v)} \quad (2.8)$$

where a is the slope parameter of the sigmoid function. By varying the parameter a , sigmoid functions of different slopes are obtained. The slope at the origin equals $a/4$. In the limit, as the slope parameter approaches infinity, the sigmoid function becomes the threshold function. Whereas a threshold function assumes the value of 0 or 1, a sigmoid function assumes a

continuous range of values from 0 to 1. Note also that the sigmoid function is differentiable, whereas the threshold function is not.

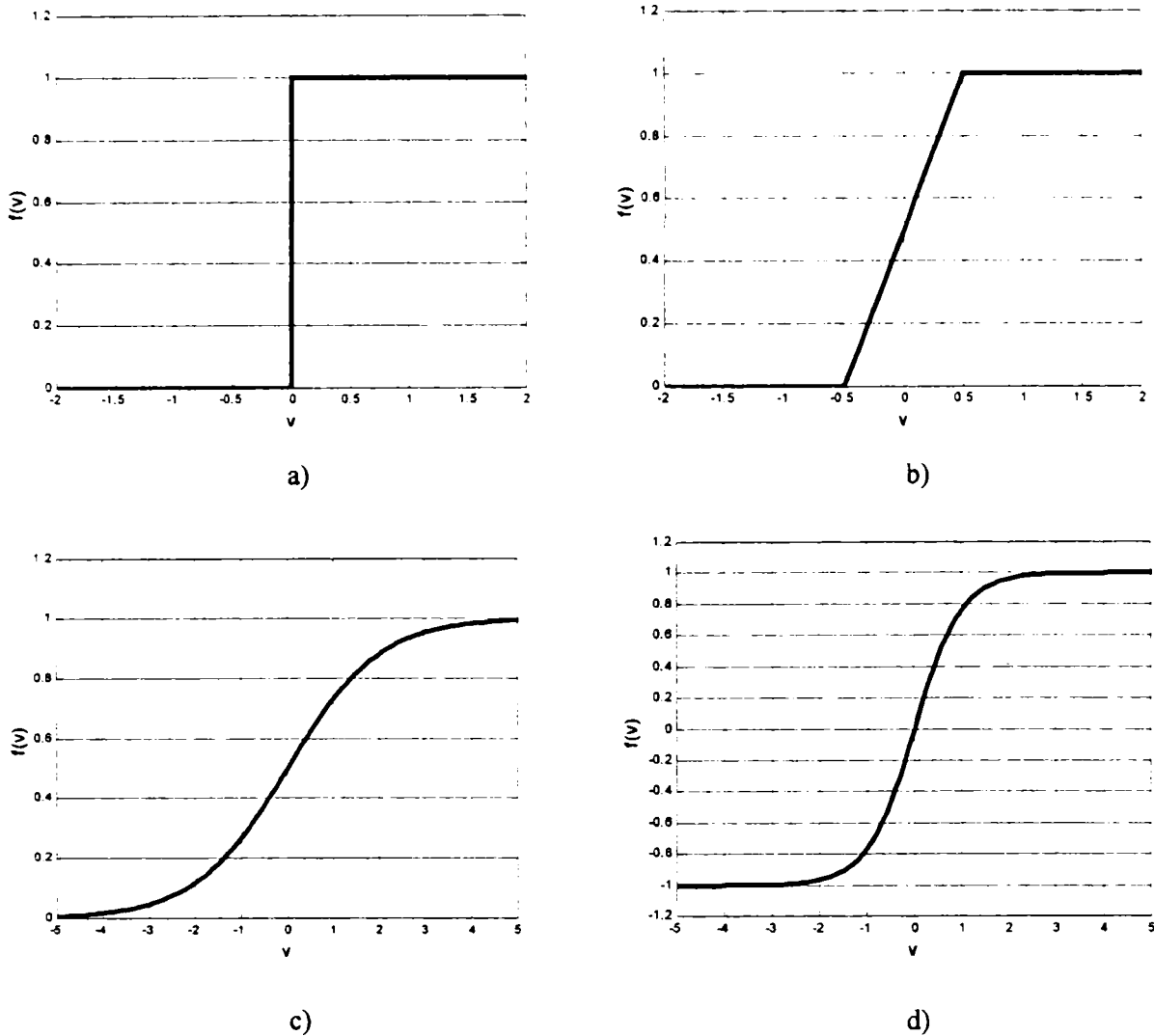


Figure 2.3: Types of activation functions: (a) - the threshold function; (b) – the piecewise linear function; (c) - the logistic function and (d) - the hyperbolic tangent function

The activation functions defined above range from 0 to 1. Sometimes it is desirable to have the activation function range from -1 to $+1$. In this case, the activation function is an odd function of the activation potential v . The corresponding form of the threshold function, which range from -1 to $+1$, is commonly referred to as the signum function. The hyperbolic tangent function, depicted in Figure 2.3d, is the corresponding form of a sigmoid function and is defined by:

$$\tanh(v) = \frac{1 - \exp(-v)}{1 + \exp(-v)} \quad (2.9)$$

2.1.4. Knowledge representation

The following generic definition for the term “knowledge” is given in [Haykin, 99]:

“Knowledge refers to stored information or models used by a person or machine to interpret, predict, and appropriately respond to the outside world”.

The challenge of the neural network is to learn the model of the environment in which operates and to maintain this model sufficiently accurately independent of any changes that this environment might undergo. [Haykin, 99] contends that knowledge in the world can be classified in two major categories:

- A prior knowledge about the environment in which the network operates. This knowledge can be communicated to the NN engineer by a subject matter expert and there are ways of incorporating this knowledge into the design of the NN.
- Observations (measurements) of the world, obtained by means of sensors designed to probe the environment in which the neural network is supposed to operate. Quite often these observations are noisy or incomplete, or both because of errors due to sensor noise and system imperfections. In any event, the observations so obtained provide the pool of information from which the examples used to train the neural network are drawn.

The training of the NN using the available observations proceeds as follows: Each example (observation) consists of an input-output pair; an input signal and the corresponding desired response for the NN. Thus, a set of examples represents knowledge about the environment of interest. This set of input-output pairs is referred to as a set of training data or training sample.

In a neural network of specified architecture, knowledge representation of the surrounding environment is defined by the values taken on by the free parameters (i.e. weights and biases) of the network. The subject of knowledge representation inside an artificial neural network is, however, very complicated, because a particular weight in a neural network is affected by many inputs to it, and the knowledge about a single input to the NN is distributed amongst many interconnection weights. Nevertheless, there are four rules for knowledge representation that are of a general commonsense nature [Haykin, 99]:

1. Similar inputs from similar classes should usually produce similar representations inside the network, and should therefore be classified as belonging to the same category.
2. Items to be categorized as separate classes should be given widely different representations in the network.

3. If a particular feature is important, then there should be a large number of neurons involved in the representation of that item in the network.
4. Prior information and invariances should be built into the design of a neural network, thereby simplifying the network design by not having to learn them.

2.1.5. Learning processes

The property that is of primary significance for a neural network is the ability of the network to learn from its environment and to improve its performance through learning. A neural network learns about its environment through an interactive process of adjustments applied to its synaptic weights and bias levels. The learning in the context of neural networks is defined as [Haykin, 99]:

“Learning is a process by which the free parameters of a neural network are adapted through a process of stimulation by the environment in which the network is embedded. The type of learning is determined by the manner in which the parameter changes take place”.

A prescribed set of well-defined rules for the solution of a learning problem is called a learning algorithm. Basically, learning algorithms differ from each other in the way in which the adjustment to a synaptic weight of a neuron is formulated. Another factor to be considered is the manner in which a neural network, made up of a set of interconnected neurons, relates to its environment.

a). Error correction learning [Haykin, 99]

Consider the simple case of a neuron k constituting the only computational node in the output layer of a feedforward neural network, as depicted in Figure 2.4.

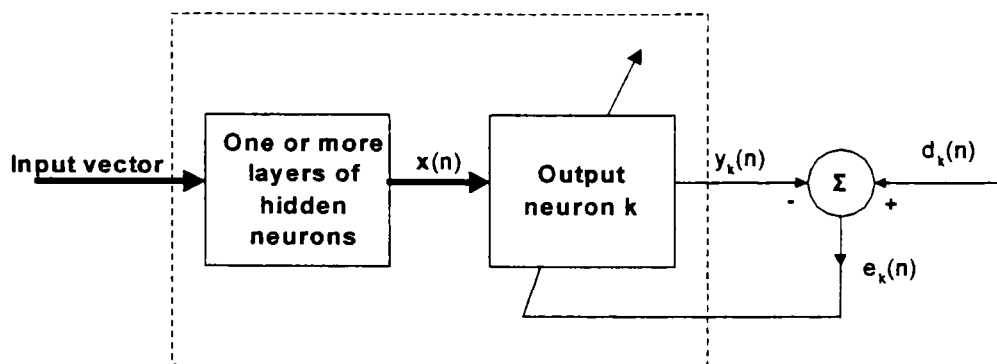


Figure 2.4. Error-correction learning

Neuron k is driven by a signal vector $x(n)$ produced by one or more layers of hidden neurons that are themselves driven by an input vector applied to the source nodes (i.e. input

layer) of the neural network. The argument n denotes the time step of an iterative process involved in adjusting the synaptic weights of neuron k . The output signal of neuron k is denoted by $y_k(n)$. This output signal, representing the only output of the neural network, is compared to a desired response, denoted by $d_k(n)$. Consequently, an error signal, denoted by $e_k(n)$, is produced. By definition, we thus have

$$e_k(n) = d_k(n) - y_k(n) \quad (2.15)$$

The error signal $e_k(n)$ actuates a control mechanism, the purpose of which is to apply a sequence of corrective adjustments to the synaptic weights of neuron k . The corrective adjustments are designed to make the output signal $y_k(n)$ come closer to the desired response $d_k(n)$ in a step-by-step manner. This objective is achieved by minimizing a cost function or index performance, $E(n)$, defined in terms of the error signal $e_k(n)$ as:

$$E(n) = \frac{1}{2} e_k^2(n) \quad (2.16)$$

That is, $E(n)$ is the instantaneous value of the error energy. The step-by-step adjustments to the synaptic weights of neuron k are continued until synaptic weights are essentially stabilized. At that point the learning process is terminated.

The learning process described herein is referred to as error-correction learning. In particular, minimization of the cost-function $E(n)$ leads to a learning rule commonly referred to as the delta-rule or Widrow-Hoff rule [Widrow, 1960]. Let $w_{kj}(n)$ denote the value of synaptic weights w_{kj} of neuron k excited by element $x_j(n)$ of the signal vector $\mathbf{x}(n)$ at time step n . According to the delta rule, the adjustment $\Delta w_{kj}(n)$ applied to the synaptic weights w_{kj} at time step n is defined:

$$\Delta w_{kj}(n) = \mu e_k(n) x_j(n) \quad (2.17)$$

where μ is the learning rate parameter (a positive constant that determines the rate of learning as the learning process proceeds from one step to another). In other words, the delta rule may be stated as [Haykin, 99]:

“The adjustment made to a synaptic weight of a neuron is proportional to the product of the error signal and the input signal of the synapse in question”.

Having computed the synaptic adjustments $\Delta w_{kj}(n)$, the update value of synaptic weight w_{kj} is determined by

$$w_{kj}(n+1) = w_{kj}(n) + \Delta w_{kj}(n) \quad (2.18)$$

In effect, $w_{kj}(n)$ and $w_{kj}(n+1)$ may be viewed as the old and new values of synaptic weight w_{kj} , respectively.

In practice, the learning rate parameter μ plays a key role in determining the performance of error-correction learning and the choice of μ also has a profound influence on the accuracy of the learning process. It is therefore important that μ is carefully selected to ensure that the stability or convergence of the iterative learning process is achieved.

b). Memory – based learning

In memory-based learning, all (or most) of the past experiences are explicitly stored in a large memory of correctly classified input-output examples: $\{(\mathbf{x}_i, d_i)\}_{i=1}^N$, where \mathbf{x}_i denotes an input vector and d_i denotes the corresponding desired response. Without loss of generality, the desired response is restricted to be a scalar. For example, in a binary pattern classification problem there are two classes, denoted by C_1 and C_2 , to be considered. In this example, the desired response d_i takes the value 0 (or -1) for class C_1 and the value 1 for class C_2 . When classification of a test vector \mathbf{x}_{test} (not seen before) is required, the algorithm responds by retrieving and analyzing the training data in a “local neighborhood” of \mathbf{x}_{test} .

All memory-based learning algorithms involve two essential ingredients:

- Criterion used for defining the local neighborhood of the test vector \mathbf{x}_{test} .
- Learning rule applied to the training examples in the local neighborhood of \mathbf{x}_{test} .

The algorithms differ from each other in the way in which these two ingredients are defined. In a simple yet effective type of memory-based learning known as the nearest neighbor rule, the local neighborhood is defined as the training example that lies in the immediate neighborhood of the test vector \mathbf{x}_{test} . In particular, the vector

$$\mathbf{x}'_N \in \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\} \quad (2.19)$$

is said to be the nearest neighbor of \mathbf{x}_{test} if

$$\min_i d(\mathbf{x}_i, \mathbf{x}_{\text{test}}) = d(\mathbf{x}'_N, \mathbf{x}_{\text{test}}) \quad (2.20)$$

where $d(\mathbf{x}_i, \mathbf{x}_{\text{test}})$ is the Euclidean distance between the vectors \mathbf{x}_i and \mathbf{x}_{test} . The class associated with the minimum distance, that is, vector \mathbf{x}'_N , is reported as the classification of \mathbf{x}_{test} . This rule is independent of the underlying distribution responsible for generating the training examples [Haykin, 99].

c). Hebbian learning

To formulate Hebbian learning in mathematical terms, consider a synaptic weight w_{kj} of neuron k with presynaptic and postsynaptic signals denoted by x_j and y_k , respectively. The adjustment applied to the synaptic weight w_{kj} at time step n is expressed in the general form

$$\Delta w_{kj}(n) = F(y_k(n), x_j(n)) \quad (2.21)$$

where $F(\cdot, \cdot)$ is a function of both postsynaptic and presynaptic signals. The signals $x_j(n)$ and $y_k(n)$ are often treated as dimensionless. The formula of equation (2.21) admits many forms, all of which qualify as Hebbian. In what follows, we consider two such forms [Haykin, 99].

Hebb's hypothesis

The simplest form of Hebbian learning is described by

$$\Delta w_{kj}(n) = \mu y_k(n) x_j(n) \quad (2.22)$$

where μ is a positive constant that determines the rate of learning. Equation (2.22) clearly emphasizes the correlational nature of a Hebbian synapse. The repeated application of the input signal (presynaptic activity) x_j leads to an increase in y_k and therefore exponential growth that finally drives the synaptic connection into saturation. At that point no information will be stored in the synapse and selectivity is lost.

Covariance hypothesis

One way of overcoming the limitation of Hebb's hypothesis is to use the covariance hypothesis introduced in [Sejnowski, 77a,b]. In this hypothesis, the presynaptic and postsynaptic signals in equation (2.22) are replaced by the departure of presynaptic and postsynaptic signals from their respective average values over a certain time interval. Let \bar{x} and \bar{y} denote the time-averaged values of the presynaptic signal x_j and postsynaptic signal y_k , respectively. According to the covariance hypothesis, the adjustment applied to the synaptic weight w_{kj} is defined by

$$\Delta w_{kj}(n) = \mu \left(x_j - \bar{x} \right) \left(y_k - \bar{y} \right) \quad (2.23)$$

where μ is the learning rate parameter. The average values constitute presynaptic and postsynaptic thresholds that determine the sign of synaptic modification.

In both cases, Hebb's hypothesis and the covariance hypothesis, the dependence of Δw_{kj} on y_k is linear; however, the intercept with the y-axis in Hebb's hypothesis is at the origin, whereas in the covariance hypothesis it is at $y_k = \bar{y}$.

The following observations can be made from equation (2.23) [Haykin, 99]:

638.001
30/12 17

- Synaptic weight w_{kj} is enhanced if there are sufficient levels of presynaptic and postsynaptic activities, that is, the conditions $x_j > \bar{x}$ and $y_k > \bar{y}$ are both satisfied.
- Synaptic weight w_{kj} is depressed if there is either
 - a presynaptic activation (i.e. $x_j > \bar{x}$) in the absence of sufficient postsynaptic activation (i.e. $y_k < \bar{y}$), or
 - a postsynaptic activation (i.e. $y_k > \bar{y}$) in the absence of sufficient presynaptic activation (i.e. $x_j < \bar{x}$)

This behavior may be regarded as a form of temporal competition between the incoming patterns [Haykin, 99].

d). Competitive learning

In competitive learning the output neurons of a neural network compete among themselves to become active. Whereas in a neural network based on Hebbian learning several output neurons may be active simultaneously, in competitive learning only a single output neuron is active at any one time.

There are three basic elements to a competitive learning rule [Rumelhart, 85]:

- A set of neurons that are all the same except for some randomly distributed synaptic weights and which therefore respond differently to a given set of input patterns.
- A limit imposed on the “strength” of each neuron.
- A mechanism that permits the neurons to compete for the right to respond to a given subset of inputs, such that only one output neuron, or only one neuron per group, is active at a time. The neuron that wins the competition is called a “winner-takes-all” neuron.

Accordingly the individual neurons of the network learn to specialize on ensembles of similar patterns; in so doing they become feature detectors for different classes of input patterns.

2.1.6. Supervised learning

An essential ingredient of supervised learning is the availability of an external teacher, as indicated in Figure 2.5. The teacher may be thought as having knowledge of the environment that is represented by a set of input-output examples. The environment is,

however, unknown to the neural network of interest. Suppose now that the teacher and the neural network are both exposed to a training vector drawn from the environment. The teacher is able to provide the neural network with a desired response for that training vector. The desired response represents the optimum action to be performed by the neural network. The network parameters are adjusted under the combined influence of the training vector and the error signal. The error signal is defined as the difference between the desired response and the actual response of the network. This adjustment is carried out iteratively in a step-by-step manner. The knowledge of the environment available to the teacher is transferred to the neural network through training as fully as possible. When this condition is reached, the teacher may be dispensed and let the neural network deal with the environment completely by itself [Haykin, 94].

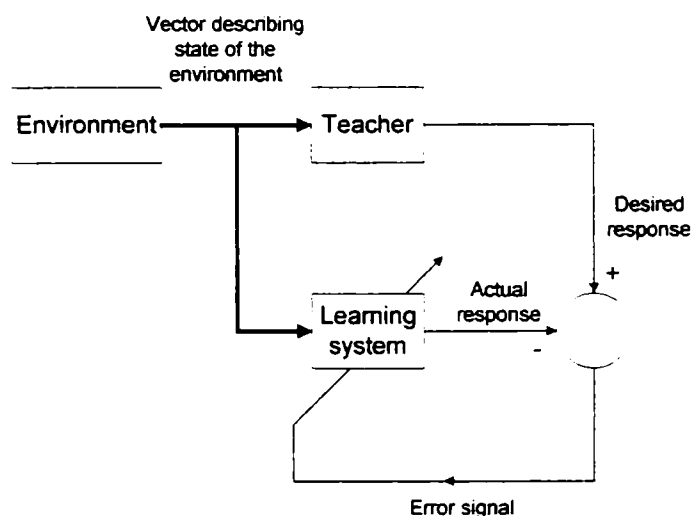


Figure 2.5. Learning with a teacher

The form of supervised learning described above is the error-correction learning discussed previously. It is a closed-loop feedback system, but the unknown environment is not included in the loop. As a performance measure for the system we may think in terms of the mean-square error or the sum of squared errors over the training samples, defined as a function of the free parameters of the system. This function may be viewed as a multidimensional error-performance surface with the free parameters as coordinates. The true error surface is averaged over all possible input-output patterns. Any given operation of the system under the teacher's supervision is represented as a point on the error surface. For the system to improve performance over time and therefore learn from the teacher the operating point has to move down successively towards a minimum point of the error surface; the minimum point may be a local minimum or a global minimum. A supervised learning system is able to do this with the useful information it has about the gradient of the error surface

corresponding to the current behavior of the system. The gradient of an error surface at any point is a vector that points in the direction of steepest descent. In fact, in the case of supervised learning from examples, the system may use an instantaneous estimate of the gradient vector, with the example indices presumed to be those of time. The use of such an estimate results in a motion of the operating point on the error surface that is typically in the form of a “random walk”. Nevertheless, given an algorithm designed to minimize the cost function, an adequate set of input-output examples and enough time permitted to do the training, a supervised learning system is usually able to perform tasks such as pattern classification and function approximation [Haykin, 99].

2.1.7. Unsupervised learning

In unsupervised learning there is no external teacher to oversee the learning process, as indicated in Figure 2.6. Suppose that the neural network is exposed to a training vector drawn from the environment. Since the teacher is absent in this setting, we are not able to provide the neural network with a desired response for the training vector.

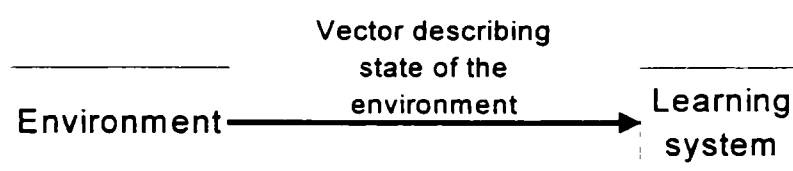


Figure 2.6. Unsupervised learning

Instead, a provision is made to identify a measure of the quality of the representation that the network is required to learn and the free parameters of the network are optimized with respect to that measure. After training is over, a grouping of the training inputs presented to the network is achieved, based on the similarity measure imposed by the network.

2.1.8. Function approximation

The choice of a particular learning algorithm is influenced by the learning task that a neural network is required to perform. In [Haykin, 99] are described six learning tasks that apply to the use of neural networks: pattern association, pattern recognition, function approximation, control, filtering and beamforming. The learning task of interest in this section is that of function approximation.

Consider a nonlinear input-output mapping described by the functional relationship [Haykin, 99]:

$$\mathbf{d} = \mathbf{f}(\mathbf{x}) \tag{2.24}$$

where the vector \mathbf{x} is the input and the vector \mathbf{d} is the output. The $\mathbf{f}(\cdot)$ is assumed to be unknown. Consider $\{(\mathbf{x}_i, \mathbf{d}_i)\}_{i=1}^N$ being a set of labeled examples. The requirement is to design a neural network that approximates the unknown function $\mathbf{f}(\cdot)$ such that the function $\mathbf{F}(\cdot)$ describing the input-output mapping actually realized by the network is closed enough to $\mathbf{f}(\cdot)$ in an Euclidean sense over all inputs, as shown by

$$\|\mathbf{F}(\mathbf{x}) - \mathbf{f}(\mathbf{x})\| < \varepsilon \quad \text{for all } \mathbf{x} \tag{2.25}$$

where ε is a small positive number. Provided that the size N of the training set is large enough and the network has an adequate number of free parameters, than the approximate error ε can be made small enough for the task.

The described approximation problem is a perfect candidate for supervised learning with \mathbf{x}_i playing the role of input vector and \mathbf{d}_i serving the role of desired response.

The ability of a neural network to approximate an unknown input-output mapping may be exploited in two important ways [Haykin, 99]:

- **System identification.** Consider that equation (2.24) describe the input-output relation of an unknown memoryless (time invariant) multiple input – multiple output (MIMO) system. The set of labeled examples $\{(\mathbf{x}_i, \mathbf{d}_i)\}_{i=1}^N$ may be used to train a neural network as a model of the system. Let y_i denote the output of the neural network produced in response to an input vector \mathbf{x}_i . The difference between \mathbf{d}_i (associated with \mathbf{x}_i) and the network output y_i provides the error signal vector \mathbf{e}_i , as depicted in Figure 2.7. This error signal is in turn used to adjust the free parameters of the network to minimize the squared difference between the outputs of the unknown system and the neural network in a statistical sense, and is computed over the entire training set.

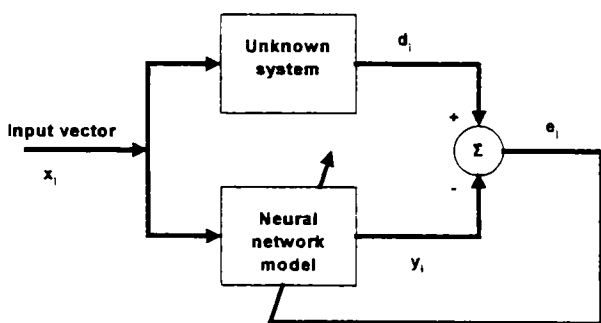


Figure 2.7. System identification

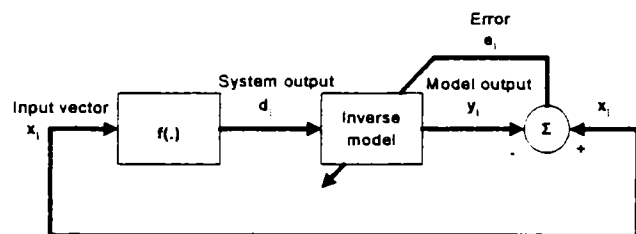


Fig. 2.8. Inverse system modeling

- Inverse system. Suppose next it is given a known memoryless MIMO system whose input-output relation is described by equation (2.24). The requirement in this case is to construct an inverse system that produces the vector \mathbf{x} in response to the vector \mathbf{d} . The inverse system may thus be described by

$$\mathbf{x} = \mathbf{f}^{-1}(\mathbf{d}) \quad (2.26)$$

where the vector-valued function $\mathbf{f}^{-1}(\cdot)$ denotes the inverse of $\mathbf{f}(\cdot)$. Note, however, that $\mathbf{f}^{-1}(\cdot)$ is not the reciprocal of $\mathbf{f}(\cdot)$; the use of superscript -1 is merely a flag to indicate an inverse. In many situations encountered in practice, the vector-valued function $\mathbf{f}(\cdot)$ is much too complex. Given the set of labeled examples $\{(\mathbf{x}_i, \mathbf{d}_i)\}_{i=1}^N$ a neural network approximation of $\mathbf{f}^{-1}(\cdot)$ may be implemented by using the scheme shown in Figure 2.8. In the situation described here, the roles of \mathbf{x}_i and \mathbf{d}_i are interchanged; the vector \mathbf{d}_i is used as the input and \mathbf{x}_i is treated as the desired response. Let the error signal vector \mathbf{e}_i denote the difference between \mathbf{x}_i and the actual output \mathbf{y}_i of the neural network produced in response to \mathbf{d}_i . As with the system identification problem, this error signal vector is used to adjust the free parameters of the neural network to minimize the squared difference between the outputs of the unknown inverse system and the neural network in a statistical sense and is computed over the complete training set.

2.2. The perceptron

The perceptron has marked an important step in the development of artificial neural network for two main reasons. First learning algorithms, allowing the training of the neural network to partition the input space into two regions, were found. Moreover, Rosenblatt [Rosenblatt, 58] proved that when the training example belong to two linear separable classes, the perceptron algorithm would always converge and drawn the decision surface in the form of a hyper-plane between the two classes.

2.2.1 Introduction

The perceptron is the simplest and the best-known model of neural network. It was proposed in 1958 in [Rosenblatt, 58] as the first model for learning with a teacher. The perceptron is built around a nonlinear neuron, namely, the McCulloch-Pitts model of a neuron

[Haykin, 94]. The model consists of a linear combiner followed by a hard limiter, as depicted in Figure 2.9:

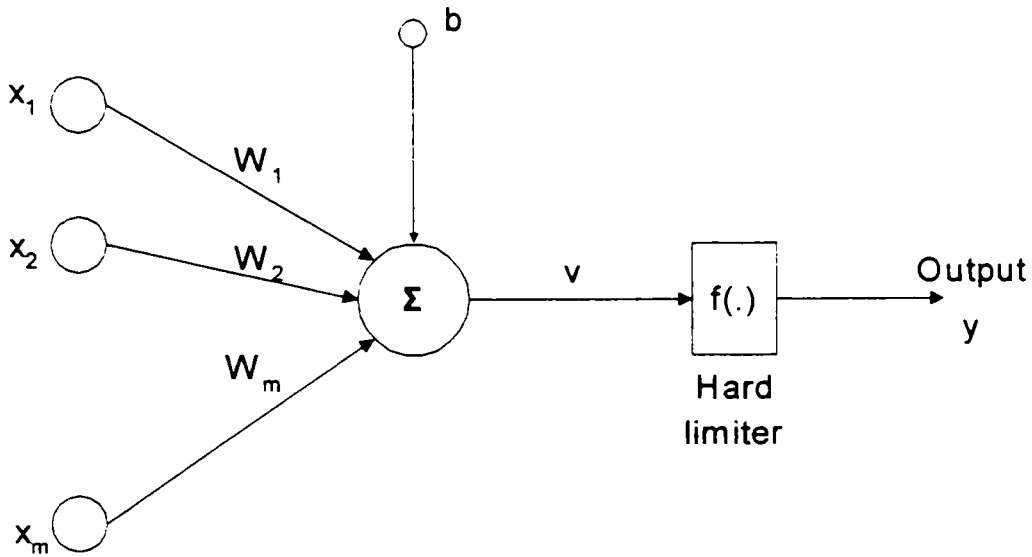


Figure 2.9. The perceptron

The hard limiter input or the activation potential v of the neuron is

$$v = \sum_{i=1}^m w_i x_i + b \quad (2.27)$$

The hard limiter performs the signum function $f(v)$

$$f(v) = \begin{cases} +1 & v \geq 0 \\ -1 & v < 0 \end{cases} \quad (2.28)$$

The goal of the perceptron is to correct classify the set of inputs x_1, x_2, \dots, x_m into one of two classes, denoted C_1 and C_2 . The decision rule for the classification is to assign the point represented by the inputs x_1, x_2, \dots, x_m to class C_1 if the perceptron output y is $+1$ and to class C_2 if it is -1 .

In the simplest form of the perceptron there are two decision regions separated by a hyperplane defined by

$$\sum_{i=1}^m w_i x_i + b = 0 \quad (3.29)$$

In Figure 2.10 it is depicted the case of two input variables x_1 and x_2 , for which the decision boundary takes the form of a straight line. A point (x_1, x_2) that lies above the boundary line is assigned to class C_1 and a point (x_1, x_2) that lies below the boundary line is assigned to class C_2 . The effect of the bias is to shift the decision boundary away from the origin.

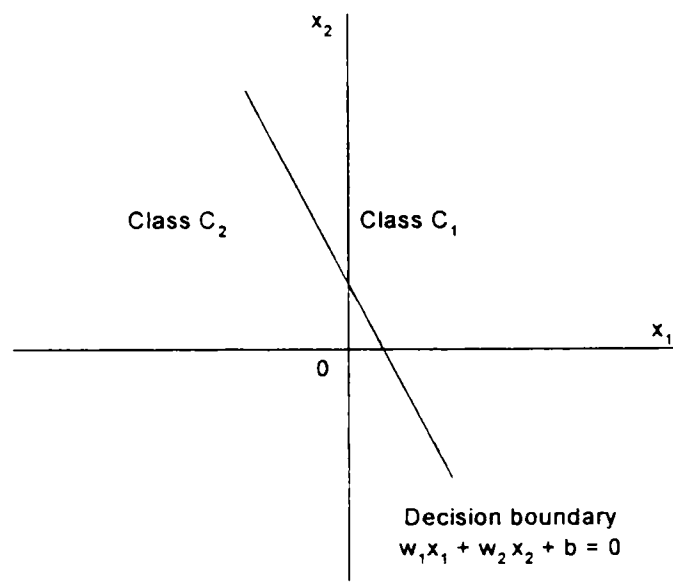


Figure 2.10. Decision boundary for a two classes pattern classification problem

The synaptic weights w_1, w_2, \dots, w_m of the perceptron can be adapted on an iteration-by-iteration basis, using an error-correction rule known as the perceptron convergence algorithm [Haykin, 99].

2.2.2. Perceptron convergence theorem

In order to derive the error-correction learning algorithm, the following variables and parameters are defined:

- n denotes the iteration step in applying the algorithm,
- $\mathbf{x}(n) = [1, x_1(n), x_2(n), \dots, x_m(n)]^T$ is the $(m+1)$ -by-1 input vector,
- $\mathbf{w}(n) = [1, w_1(n), w_2(n), \dots, w_m(n)]^T$ is the $(m+1)$ -by-1 weight vector,
- $b(n)$ is the bias treated as a synaptic weight driven by a fixed input equal to 1,
- $d(n)$ is the desired response,
- $y(n)$ is the actual response (quantized),
- μ is the learning rate parameter (a positive constant less than unity).

The linear combiner output is written in the compact form

$$v(n) = \sum_{i=0}^m w_i(n)x_i(n) = \mathbf{w}^T(n)\mathbf{x}(n) \quad (2.30)$$

where $w_0(n)$ represents the bias $b(n)$.

The perceptron convergence algorithm can be summarized as follows [Haykin, 99]:

1. Initialization. Set $\mathbf{w}(0) = 0$. Then perform the following computations for time step $n = 1, 2, \dots$

2. Activation. At time step n , activate the perceptron by applying continuous-valued input vector $\mathbf{x}(n)$ and desired response $\mathbf{d}(n)$.

3. Computation of actual response. Compute the actual response of the perceptron

$$y(n) = f[\mathbf{w}^T(n)\mathbf{x}(n)] \quad (2.31)$$

where $f(\cdot)$ is the signum function.

4. Adaptation of the weight vector. Update the weight vector of the perceptron

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu [d(n) - y(n)]\mathbf{x}(n) \quad (2.32)$$

where

$$d(v) = \begin{cases} +1 & \text{if } \mathbf{x}(n) \text{ belongs to class } C_1 \\ -1 & \text{if } \mathbf{x}(n) \text{ belongs to class } C_2 \end{cases} \quad (2.33)$$

is the quantized desired response and $d(n) - y(n)$ plays the role of an error signal.

5. Continuation. Increment time step n by one and go back to step 2.

The learning rate parameter μ is a positive constant, limited to the range $0 < \mu \leq 1$. When assigning a value to it inside this range, two conflicting requirements has to be taken into account [Lippmann, 1987]:

- Averaging of past inputs to provide stable weight estimates, which requires a small μ ,
- Fast adaptation with respect to real changes in the underlying distributions of the process responsible for the generation of the input vector \mathbf{x} , which requires a large μ .

2.3. Multilayer perceptron (MLP)

It has been shown in section 2.2 that the perceptron could only design linear decision boundaries. This might provide extremely restrictive for a wide range of problems. Some problems can involve classes that are not linearly separable. By creating a network organized in different layers, as depicted in Figure 2.11, it is possible to implement complex decision boundaries [Lippmann, 87].

2.3.1. Introduction

The multilayer perceptron (MLP) is a neural network that consists of an input layer of source nodes, one or more hidden layers of nodes and an output layer, also made up of neurons. The source nodes provide physical access point for the application at hand. The neurons in the hidden layers act physically inaccessible from the input end or output end of

the network. The neurons in the output layers present to a user the conclusions reached by the network in response to the input signals.

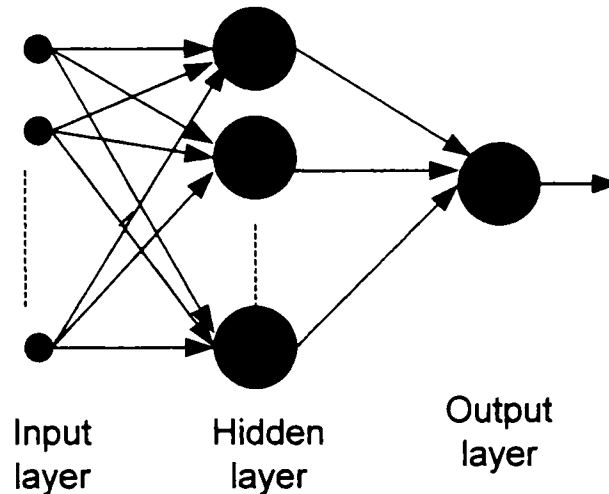


Figure 2.11. Multilayer perceptron with a single hidden layer.

Figure 2.11 depicts a multilayer perceptron with a pair of input nodes, a single hidden layer of neurons and a single output neuron. Two key characteristics of such a structure are immediately apparent from this figure [Haykin, 96]:

1. A multilayer perceptron is a feedforward network, in the sense that the input signals produce a response at the output of the network by propagating in the forward direction only. There is no feedback in the network.
2. The network may be fully connected, as shown in figure 2.11, in that each node in a layer of the network is connected to every node in the layer adjacent to it. Alternatively, the network may be partially connected in that some of the synaptic links may be missing. Locally connected networks represent an important type of partially connected networks; the term local refers to the connectivity of a neuron in a layer of the network only to a subset of possible inputs.

The number of nodes in the input layer is determined by the dimensionality of the observation space that is responsible for the generation of input signals. The number of nodes in the output layer is determined by the required dimensionality of the desired response. Thus, the design of a multilayer perceptron requires that we address three issues [Haykin, 96]:

1. The determination of the number in the hidden layers,
2. The determination of the number of neurons in each of the hidden layers,
3. The specification of the synaptic weights that interconnect the neurons in the different layers of the network.

Figure 2.12 presents the decision boundaries that can be produced using the perceptron and the MLP with one and two hidden layers and two inputs.

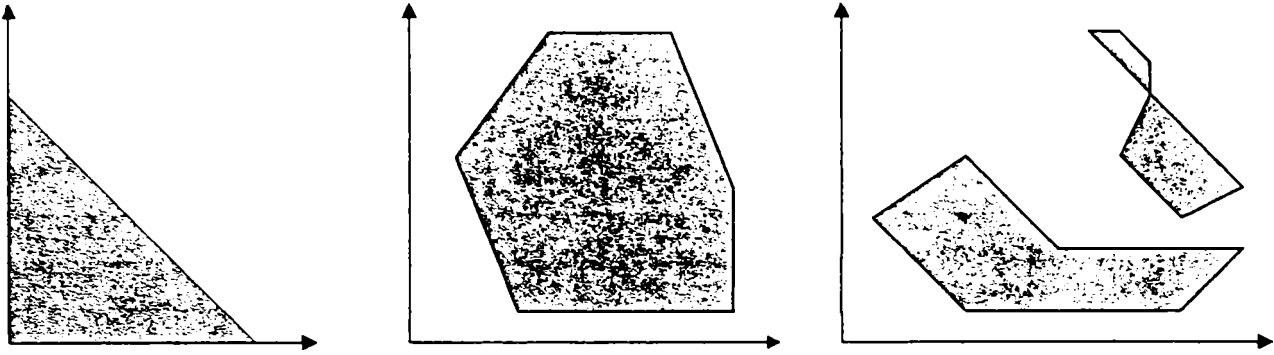


Figure 2.12. The capability of the MLP to design complex decision boundaries: (a) – single perceptron; (b) – MLP with one hidden layer; (c) – MLP with two hidden layers

Figure 2.12 indicates that the perceptron can only draw linear decision boundaries. When one hidden layer of neurons is added, it is then possible to implement arbitrary complex convex decision boundaries [Lippmann, 87]. It was shown later that neural networks with only one hidden layer are able to create regions arbitrarily close to any non-linear decision boundary [Makhoul, 89]. Finally, with two hidden layer it is also possible to design any decision boundary [Lippmann, 87]. Moreover, certain problems can be solved with a small number of neurons and two hidden layers, whereas a network with only one hidden layer would require an infinite number of neurons. The choice of the number of hidden layers in the multilayers perceptrons is generally open.

2.3.2. Training algorithms for MLP

In the previous sections, a general presentation of the artificial neural networks was given. It has been shown that artificial neural networks are adaptive algorithms and therefore require a training algorithm to adapt their synaptic weights. The derivation of the backpropagation algorithm (BKP) [Rumelhart, 86] marked an important point in the development of ANN. Using this algorithm made it possible to efficiently perform the weight adaptation of MLP and widened the range of possible areas of applications for ANN. The training algorithm is in fact performed in two separate steps. First, the derivative of the error function to be minimized is computed with respect to the weights of the neural network. This procedure corresponds to the propagation of errors backwards in the network and is the BKP algorithm itself. Then, these derivatives can be used in conjunction with some other algorithms, such as gradient descent (GD), to update the weights of the network.

2.3.2.1. The backpropagation (BKP) algorithm

First some notations are presented. A fully connected MLP consisting of N_K layers is considered. Each layer k of the neural network is made of $N_n^{(k)}$ neurons. The internal activity level and the function signal of the i^{th} neuron in the layer k are noted $y_i^{(k)}(n)$ and $x_i^{(k)}(n)$ respectively and are computed according to the following equations:

$$y_i^{(k)}(n) = \sum_{j=0}^{N_n^{(k-1)}} w_{ij}^{(k)} \cdot x_j^{(k-1)}(n) \quad (2.34)$$

$$x_i^{(k)}(n) = f(y_i^{(k)}(n)) \quad (2.35)$$

where $w_{ij}^{(k)}$ is the synaptic weight connecting the i^{th} neuron in the layer k to the j^{th} neuron in the layer $k-1$ and f represents the nonlinear activation function of the neuron that maps the internal activity level to output.

The error function to be minimized is denoted as E and it is expressed as a sum of the error functions over the number of training examples in the following form:

$$E = \sum_n E_n \quad (2.36)$$

The error function E_n is assumed to be differentiable with respect to the outputs of the network. Using activation functions that are differentiable, such as sigmoid functions, ensures the derivability of the error functions E_n with respect to the weights in the network. Using the chain rule for partial derivatives, the following equation can be written:

$$\frac{\delta E_n}{\delta w_{ij}^{(k)}} = \frac{\delta E_n}{\delta y_i^{(k)}} \cdot \frac{\delta y_i^{(k)}}{\delta w_{ij}^{(k)}} \quad (2.37)$$

Hence:

$$\frac{\delta E_n}{\delta w_{ij}^{(k)}} = \delta_i^{(k)} \cdot x_j^{(k-1)} \quad (2.38)$$

with

$$\delta_i^{(k)} = \frac{\delta E_n}{\delta y_i^{(k)}} \quad (2.39)$$

The coefficients $\delta_i^{(k)}$ are referred to as local errors and are the only parameters to be estimated in the network in order to compute the whole set of derivatives. For the output layer, the local errors can be easily computed through the following formula:

$$\delta_i^{(N_k)} = \frac{\delta E_n}{\delta y_i^{(N_k)}} = \frac{\delta E_n}{\delta x_i^{(N_k)}} \cdot f'(y_i^{(N_k)}) \quad (2.40)$$

For the first hidden layer the chain rule for the partial derivatives is used again:

$$\delta_i^{(k)} = \frac{\delta E_n}{\delta y_i^{(k)}} = \sum_{m=1}^{N_{k+1}} \frac{\delta E_n}{\delta y_m^{(k+1)}} \cdot \frac{\delta y_m^{(k+1)}}{\delta y_i^{(k)}} \quad (2.41)$$

which then leads to:

$$\delta_i^{(k)} = \sum_{m=1}^{N_{k+1}} \delta_m^{(k+1)} \cdot w_{mi}^{(k)} \cdot f'(y_i^{(k)}) \quad (2.42)$$

Equation (2.40) allows the computation of the derivatives with respect to all the weights in the neural network starting from the output layer and processing backward through hidden layers.

The BKP algorithm makes it possible to compute the derivatives of the error function to be minimized with respect to the weights for the whole network. Knowing these derivatives values, it is then possible to update the weights using a simple gradient-descent (GD) algorithm:

$$w_{ij}^{(k)}(n+1) = w_{ij}^{(k)}(n) - \mu \cdot \delta_i^{(k)}(n) \cdot x_j^{(k-1)}(n) \quad (2.43)$$

where μ is the adaptation step.

In the application of the BKP algorithm, two distinct passes of computation are distinguished: the forward pass and the backward pass. In the forward pass the synaptic weights remain unaltered throughout the network and the function signals (that propagate forward through the network) of the network are computed on a neuron-by-neuron basis. In other words, the forward phase of computation begins at the first hidden layer by presenting it with the input vector and terminates at the output layer by computing the error signal for each neuron this layer. The backward pass, on the other hand, starts at the output layer by passing the error signal leftward through the network, layer by layer, and recursively computing the local gradient for each gradient. This recursive process permits the synaptic weights of the network to undergo changes in accordance with the equation (2.43).

Sequential and batch modes of training

In practical applications of the BKP algorithm, learning results from the many presentations of a prescribed set of training examples to the multilayer perceptron. One

complete presentation of the entire set during the learning process is called an epoch. The learning process is maintained on an epoch-by-epoch basis until the synaptic weights and bias levels of the network stabilize and the average squared error over the entire training set converges to some minimum value. It is good practice to randomize the order of presentation of training examples from one epoch to the next [Haykin, 99].

In the sequential mode of BKP learning (also referred to as on-line, pattern, incremental or stochastic mode) the gradient is computed and the weights are updated after the presentation of each training example. In the batch mode of BKP learning the weights are changed after the presentation of all the training examples.

From an “on-line” operational point of view, the incremental mode of training is preferred over the batch mode because it requires less local storage for each synaptic connection and, given that the patterns are presented to the network in a random fashion, the use of pattern-by-pattern updating of weights makes the search in weight space stochastic in nature. This in turn makes it less likely for the BKP algorithm to be trapped in a local minimum. In the same way, the stochastic nature of the incremental mode makes it difficult to establish theoretical conditions for convergence of the algorithm. In contrast, the use of batch mode of training provides an accurate estimate of the gradient vector; convergence to a local minimum is thereby guaranteed under simple conditions [Haykin, 99].

When the training data are redundant (i.e. the data set contains several copies of exactly the same pattern), it was found that unlike the batch mode, the incremental mode is able to take advantage of this redundancy because the examples are presented one at a time. This is particularly when the data set is large and highly redundant [Haykin, 99].

Stopping Criteria

One of the differences between the BKP learning rule (generalized delta rule) and the perceptron learning rule is that the perceptron learning rule will converge to a solution, if such a solution exists, in a finite number of steps, while the BKP learning rule can go on forever without ever reaching a time when all the actual outputs are equal to the desired outputs. Hence, stopping criteria must be established to designate the end of the training process. It may be formulated a sensible convergence criterion for BKP learning as follows [Haykin, 99]: *“The back-propagation algorithm is considered to have converged when the Euclidean norm of the gradient vector reaches a sufficiently small gradient threshold”*.

The drawback of this convergence criterion is that, for successful trials, learning time may be long and also requires the computation of the gradient vector.

In [Haykin, 99] is suggested a different criterion of convergence:

“The back-propagation algorithm is considered to have converged when the absolute rate of change in the average squared error per epoch is sufficiently small”.

The rate of change in the averaged squared error is typically considered to be small enough if it lies in the range of 0.1 to 1 percent per epoch. Sometimes a value as small as 0.01 percent per epoch is used. Unfortunately, this criterion may result in a premature termination of the learning process [Haykin, 99].

Heuristics for making the BKP algorithm perform better

It is often said that the design of a neural network using the BKP algorithm is more of an art than a science in the sense that many of numerous factors involved in the design are the results of one's personal experience. There are methods that will significantly improve the BKP algorithm's performance [Haykin, 99].

1. Incremental versus batch update. The incremental mode is computationally faster than the batch mode, especially when the data set is large and highly redundant.
2. Maximizing information content. As a general rule, every training example presented to the BKP algorithm should be chosen on the basis that its information content is largest possible for the task at hand. Two ways of achieving this aim are: (1) the use of an example that results in the largest training error and (2) the use of an example that is radically different from all those previously used.
3. Activation function. A MLP trained with the BKP algorithm may, in general, learn faster (in terms of the number of training iterations required) when the sigmoid activation function built into the neuron model of the network is antisymmetric (i.e. odd function of its argument) than when it is nonsymmetric.
4. Desired response. It is important that the desired response be chosen within the range of the sigmoid activation function. More specifically, the desired response for a neuron in the output layer of a MLP should be offset by some amount away from the limiting value of the sigmoid activation function, depending on whether the limiting value is positive or negative. Otherwise the BKP algorithm tends to drive the free parameters of the network to infinity and thereby slow down the learning process by driving the hidden neurons into saturation.
5. Normalizing the inputs. Each input variable should be preprocessed so that its mean value, averaged over the entire training set, is close to zero, or else it is small compared to its standard deviation. In order to accelerate the BKP learning process,

the normalization of the inputs should also include two other measures: (1) the input variables contained in the training set should be uncorrelated and (2) the decorrelated input variables should be scaled so that their covariances are approximately equal.

6. Initialization. The first step in BKP learning is to initialize the network. The customary practice is to set all the free parameters (weights) in the MLP to random numbers that are uniformly distributed inside a small interval of values, symmetric around zero. The wrong choice of initial weights can lead to a phenomenon known as premature saturation. There are ways to counter this premature saturation problem. One way is to choose the weights converging to a node i in the MLP uniformly distributed over an interval of the form $[-\alpha_i/F_i, \alpha_i/F_i]$, where α_i is an appropriately chosen constant and F_i is the fan-in of the nodes that are affecting node i (or the number weights converging to node i). In this way it can be guaranteed that the output of a node in the MLP structure will not be initially saturated to an incorrect value. Another way to avoid the premature saturation problem is to design an error function whose minimization will guarantee the correct mapping for the training data [Christodoulou, 01].
7. Learning from hints. Learning from a set of training examples deals with an unknown input-output mapping function $f(\cdot)$. In effect, the learning process exploits the information contained in the examples about the function $f(\cdot)$ to infer an approximate implementation of it. The process of learning from examples may be generalized to include learning from hints, which is achieved by allowing prior information that it may exist about the function $f(\cdot)$ to be included in the learning process. Such information may include invariance properties, symmetries or any other knowledge about the function $f(\cdot)$ that may be used to accelerate the search of its approximate realization, and more importantly, to improve the quality of the final estimate.
8. Learning rates. All neurons in the MLP should ideally learn at the same rate. The last layers usually have larger local gradients than the layers at the front end of the network. Hence, the learning rate parameter should be assigned a smaller value in the last layers than in the front layers. Neurons with many inputs should have a smaller learning rate parameter than neurons with few inputs so as to maintain a similar learning time of all neurons in the network. In [LeCun, 93] it is suggested that for a given neuron, the learning rate should be inversely proportional to the square root of synaptic connections made to that neuron.

2.3.3.2. Advanced learning algorithms in MLP

The basic backpropagation algorithm described in section 2.3.2.1 is a gradient descent algorithm based on the estimation of the instantaneous sum-squared error for each layer. The simplest implementation of the BKP learning updates the network weights and biases in the direction in which the performance function decreases most rapidly – the negative of the gradient. An iteration of this algorithm can be expressed as (based on equation (2.43)) (the weight update):

$$\Delta W(n) = W(n+1) - W(n) = -\mu \cdot \nabla_w E(n) = -\mu \cdot g(n) \quad (2.44)$$

where $W(n)$ is a vector of current weights and biases (iteration n), $g(n)$ is the current gradient and μ is a positive constant called learning rate. The performance of the algorithm is very sensitive to the proper setting of the learning rate: if the learning rate is set too large the algorithm will become unstable and if the learning rate is set too small, the algorithm will take a long time to converge.

Such an algorithm is slow for three basic reasons [Haykin, 94]:

1. It uses an instantaneous sum-squared error $E(W)$ to minimize the mean squared error, denoted $J(W)$, over training epoch (iteration). The gradient of the instantaneous is not a good estimate of the gradient of the mean squared error. Therefore, satisfactory minimization of this error requires more iterations of the training process.
2. It is a first-order minimization algorithm that is based on the first-order derivatives (a gradient). Faster algorithms use also the second derivatives (the Hessian matrix).
3. The error propagation serializes computations on the layer-by-layer basis.

The mean squared error, $J(W)$, is a relatively complex surface in the weight space, possibly with many local minima, flat sections, narrow irregular valleys and saddle points. The complexity of the error surface is the main reason that the behavior of the simple steepest descent minimization algorithm can be very complex and may have oscillations around a local minimum.

The faster algorithms fall into two main categories. The first category use heuristic techniques that were developed from an analysis of the performance of the standard steepest descent algorithm. The heuristic techniques discussed are: the momentum technique, the adaptive learning rate backpropagation and the resilient backpropagation. The second category of fast algorithms uses standard numerical optimization techniques. In this chapter will be presented three types of numerical optimization techniques for neural networks: the conjugate gradient, the quasi-Newton and the Levenberg-Marquardt techniques.

In the next section we will consider improvements to the basic backpropagation algorithms based on heuristic methods. These methods aim to an improvement of the algorithm by making modifications to its parameters or to the form.

A. Heuristic improvements of the BKP algorithm

A1. The momentum term

The BKP algorithm provides an “approximation” to the trajectory in weight space computed by the method of steepest descent (appendix 1). The smaller is the learning rate parameter μ , the smaller the changes to the synaptic weights in the network will be from one iteration to the next and the smoother will be the trajectory in weight space. However, this improvement is attained at the cost of a slower rate of learning. If the learning rate parameter μ is chosen too large in order to speed up the rate of learning, the resulting large changes in the synaptic weights assume such a form that the network may become unstable (oscillatory). A simple method of increasing the rate of learning yet avoiding the danger of instability is to modify the delta rule of equation (2.43) by including a momentum term.

One simple method to avoid an error trajectory in the weight space being oscillatory is to add to the weight update a momentum term (denoted Ω) that is proportional to the weight update at the previous step. Momentum allows a network to respond not only to the local gradient, but also to recent trends in the error surface. Acting like a low pass filter, this modification to the steepest descent algorithm is able to ignore small features in the error surface. Without momentum a network may get stuck in a shallow local minima.

$$\Delta W(n) = -\mu \cdot g(n) + \Omega \cdot \Delta W(n-1) \quad (2.45)$$

A2. Adaptive learning rate

An adaptive learning rate during the training process will attempt to keep the learning step size as large as possible while keeping the learning stable. A typical strategy is based on monitoring the rate of change of the mean square error and can be described as follows:

- If the mean square error J is decreasing consistently, that is ∇J is negative for a prescribed number of steps, then the learning rate is increased linearly:

$$\mu(n+1) = \mu(n) + \alpha, \quad \alpha > 0 \quad (2.46)$$

- If the error has increased ($\nabla J > 0$), the learning rate is exponentially reduced:

$$\mu(n+1) = \beta \cdot \mu(n), \quad 0 < \beta < 1 \quad (2.47)$$

A3. Resilient backpropagation

Multilayer neural networks typically use sigmoid transfer function in the hidden layers. Sigmoid functions are characterized by the fact that their slope must approach zero, as the input gets large. This causes a problem when using steepest descent, since the gradient can have a very small magnitude and therefore causes small changes in the weights and biases. The purpose of the resilient backpropagation algorithm is to eliminate these effects of the magnitudes on the partial derivatives. Only the sign of the derivatives is used to determine the direction of the weight update, the magnitude of the derivative has no effect on the weight update. The update value for each weight and bias is increased by a factor γ whenever the derivative of the performance function with respect to that weight has the same sign for two successive iterations. The update value is decreased by a factor γ whenever the derivative with respect to that weight changes sign from the previous iteration. If the derivative is zero, there are no changes in the update value. Whenever the weights are oscillating the weight change will be reduced. If the weight continues to change in the same direction for several iterations, then the magnitude of the weight change will be increased. Generally, the resilient backpropagation algorithm converges much faster than the previous algorithms.

B. Conjugate gradient algorithms

In most of the training algorithm discussed in the previous section, a learning rate is used to determine the length of the weight update (step size). In most of the conjugate gradient algorithms the step size is adjusted at each iteration. In the conjugate gradient algorithms a search is performed along conjugate directions to determine the step size that will minimize the performance function along that line. All of the conjugate gradient algorithms start by searching in the steepest descent direction (negative of the gradient) on the first iteration (\mathbf{g} denotes the current gradient):

$$\mathbf{p}(0) = -\mathbf{g}(0) \quad (2.48)$$

A line search is then performed to determine the optimal distance to move along the search direction:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu(n)\mathbf{p}(n) \quad (2.49)$$

where next value of the weight vector $\mathbf{w}(n+1)$ is obtained from the current value of the weight vector, $\mathbf{w}(n)$, by moving it in the direction of a vector $\mathbf{p}(n)$ (n is the time step).

Then the next search direction is determined so that it is conjugate to previous search directions. The general procedure for determining the new search direction is to combine the new steepest descent direction with the previous search direction:

$$\mathbf{p}(n+1) = -\mathbf{g}(n) + \beta(n)\mathbf{p}(n) \quad (2.50)$$

where $\beta(n)$ are scaling factors to be determined and are selected so that the direction $\mathbf{p}(n+1)$ and $\mathbf{p}(n)$ are conjugate with respect to the Hessian matrix, $\nabla^2 J(\mathbf{w}) = \mathbf{H}$, that is,

$$\mathbf{p}(n+1) \cdot \mathbf{H} \cdot \mathbf{p}^T(n) = 0 \quad (2.51)$$

For all conjugate gradient algorithms, the search direction will be periodically reset to the negative of the gradient. The standard reset point occurs when the number of iterations is equal to the number of network parameters (weights and biases) but there are another reset methods that can improve the efficiency of training. In our application the neural networks is trained with the Powell-Beale version of the conjugate gradient algorithm. This method was proposed by Powell [Powell, 77], based on the earlier version proposed by Beale [Beale, 72]. For this technique the restart takes place if there is very little orthogonality left between the current gradient and the previous gradient. This is tested with the following inequality:

$$\left| \mathbf{g}^T(n-1)\mathbf{g}(n) \right| \geq 0.2 \cdot \|\mathbf{g}(n)\|^2 \quad (2.52)$$

If this condition is satisfied, the search direction is reset to the negative of the gradient.

The Fletcher-Reeves update formula is:

$$\beta(n) = \frac{\mathbf{g}(n) \mathbf{g}^T(n)}{\mathbf{g}(n-1) \mathbf{g}^T(n-1)} \quad (2.53)$$

The Polak-Ribière formula is:

$$\beta(n) = \frac{(\mathbf{g}(n) - \mathbf{g}(n-1)) \cdot \mathbf{g}^T(n)}{\mathbf{g}(n-1) \mathbf{g}^T(n-1)} \quad (2.54)$$

In summary, the conjugate gradient algorithm involves: Initial searching $\mathbf{p}(0) = -\mathbf{g}(0)$; line minimization with respect of μ ; calculation of the next search direction as in equation (2.50) and β from one of the above presented formula.

C. Quasi-Newton Algorithms

Newton's method [appendix 2] is an alternative to the conjugate gradient methods for fast optimization. The basic step of Newton's method is:

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \mathbf{H}^{-1}(n)\mathbf{g}(n) \quad (2.55)$$

where \mathbf{H} is the Hessian matrix (second derivatives) of the performance index at the current values of the weights and biases. The Hessian matrix provides additional information about

the shape of the performance index surface in the neighborhood $\mathbf{w}(n)$. Newton's method often converges faster than conjugate gradient methods. However, it requires computations of the inverse of the Hessian matrix that is relatively complex and expensive.

The quasi-Newton method (also called secant method) is based on Newton's method but do not require the calculation of the second derivatives. An approximate Hessian matrix is updated at each iteration of the algorithm. The update is computed as a function of the gradient.

D. Levenberg-Marquardt algorithm

One problem with the Newton's algorithm is that the approximate Hessian matrix may not be invertible. To overcome this problem in the Levenberg-Marquardt algorithm a small constant μ is added such as the weight update rule becomes:

$$\Delta \mathbf{w} = \mathbf{w}(n+1) - \mathbf{w}(n) = -[\mathbf{J}^T(\mathbf{w}) \cdot \mathbf{J} + \mu \cdot \mathbf{I}]^{-1} \cdot \mathbf{J}^T(\mathbf{w}) \cdot \mathbf{e}(\mathbf{w}) \quad (2.56)$$

where $\mathbf{J}(\mathbf{w})$ is the Jacobian matrix that contains the first derivatives of the network errors with respect to the weights and biases, \mathbf{I} is the identity matrix, $\mathbf{e}(\mathbf{w})$ is a vector of the network errors and μ is a small constant. For large values of μ the $\mathbf{J}^T(\mathbf{w}) \mathbf{J}(\mathbf{w})$ terms become negligible and learning progresses according to $\mu^{-1} \mathbf{J}^T(\mathbf{w}) \mathbf{e}(\mathbf{w})$, which is gradient descent. Whenever a step is taken and error increases, μ is increased until a step can be taken without increasing error. However, if μ becomes too large, no learning process takes place (i.e. $\mu^{-1} \mathbf{J}^T(\mathbf{w}) \mathbf{e}(\mathbf{w})$ approaches zero). This occurs when an error minima has been found. For small value of μ , the above expression becomes the Gauss-Newton method [appendix 3].

2.3.3. Generalization

After the presentation of the training set to the neural network, it is hoped that the weights have converged to a point allowing a good operation when the data set is presented. The ability of the network to operate on unknown data (test data never used in training the network) is referred to as generalization. The generalization mainly depends on three parameters [Haykin, 94]:

1. The training process (number of training examples and the extent to which they represent the classes to be classified),
2. The network configuration (number of hidden layers and neurons),
3. The complexity of the problem to be solved.

Clearly, there is no control over the latter. In the context of the other two factors, the issue of generalization may be viewed from two different perspectives [Haykin, 99]:

- The architecture of the network is fixed (hopefully in accordance with the physical complexity of the underlying problem) and the issue to be resolved is that of determining the size of the training set needed for a good generalization to occur.
- The size of the training set is fixed and the issue of interest is that of determining the best architecture of network for achieving good generalization.

The choice of the network architecture impacts on the training procedure. In [Hush, 89] it is shown that a small number of training examples would imply better performance when only one hidden layer is considered rather than two. This is due to the high flexibility and the large number of free parameters associated with three layer neural networks, which then call for a long training procedure in order to converge. The influence of the number of neurons in the network and the number of training examples on the generalization capabilities of the network are closely linked. [Huang, 91] has proved that, for a given size of training examples, the number of neurons in the network necessary to implement the training data is of the order of the number of training examples. If the network is oversized, the training data will be memorized and the generalization will not be possible. Hence, the neural network should be complex enough in order to be able to draw decisions boundaries complex enough to solve the problem at hand. Nevertheless, as the number of neurons increases, the length of the required training sequence will increase. It is therefore important to keep the size of the network as low as possible in order to reduce the transmission overhead induced by the training sequence.

2.3.4. Cross - validation

The network selection problem may be viewed as choosing, within a set of candidate model structures, the “best” one according to a certain criterion. Cross-validation is a standard tool used in statistical prediction and model selection in control theory. First, the available data set is randomly partitioned into a training set and a test set. The training set is further partitioned into two disjoint sets: estimation subset (used to select the model) and validation subset (used to test or validate the model).

The motivation here is to validate the model on a data set different from the one used for parameter estimation. In this way, the training set may be used to assess the performance of various candidate models, and thereby choose the “best” one. There is, however, a distinct

possibility that the model with the best-performing parameter values so selected may end up overfitting the validation subset. To guard against this possibility, the generalization performance of the selected model is measured on the test set, which is different from the validation subset. The use of cross-validation is appealing particularly when we have to design a large neural network with good generalization as the goal. For example, cross-validation may be used to determine the multilayer perceptron with the best number of hidden neurons and when it is best to stop training.

On the basis of the results reported in [Kearns, 96], 80 percent of the training set could be assigned to the estimation subset and the remaining 20 percent are assigned to the validation subset.

Early stopping method of training

A multilayer perceptron trained with the backpropagation algorithm learns in stages and during the training process the mean square error decreases with an increasing number of epochs: it starts from off at a large value, decreases rapidly and then continues to decrease slowly as the network makes its way to a local minimum on the error surface. With good generalization as the goal, it is very difficult to figure out when is the best to stop training. In particular, it is possible for the network to end up overfitting the training data if the training session is not stopped at the right point.

In the early stopping method of training the estimation subset of examples is used to train the network, in the usual way, with the observation that the training session is stopped periodically and the network is tested on the validation subset after each period of training. More specifically, the periodic estimation-followed-by-validation process proceeds as follows [Haykin, 99]:

- After a period of estimation (training), the synaptic weights and bias levels of the multilayer perceptron are all fixed, and the network is operated in its forward mode. The validation error is thus measured for each example in the validation subset.
- When the validation phase is completed, the estimation (training) is resumed for another period, and the process is repeated.

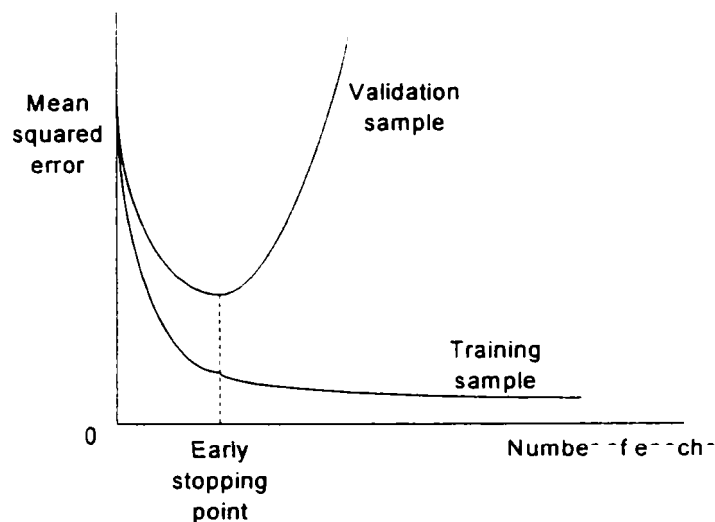


Figure 2.13. Illustration of the early stopping rule based on cross validation

From Figure 2.13 it can be noticed that the model does not as well on the validation subset as it does on the estimation subset. The estimation learning curve decreases monotonically for an increasing number of epochs and the validation learning curve decreases monotonically to a minimum, it then starts to increase as the training continues. What the network is learning beyond the minimum point on the validation learning curve is essentially noise contained in the training data. This heuristic suggests that the minimum point on the validation learning curve be used as a sensible criterion for stopping the training session.

If the training data are noise free, both of the estimation and validation errors cannot be simultaneously driven to zero and this implies that the network does not have the capacity to model the function exactly. The best that can be done in that situation is to try to minimize, for example, the integrated squared error that is (roughly) equivalent to minimizing the usual global mean-square error with a uniform input density.

2.4. Radial Basis Function networks

2.4.1. Introduction

The backpropagation algorithm for the design of a multilayer perceptron as described in the previous section may be viewed as the application of a recursive technique known in statistics as stochastic approximation. A completely different approach is considered in this section by viewing the design of a neural network as a curves fitting (approximation) problem in a high dimensional space. According to this viewpoint, learning is equivalent to finding a

surface in a multidimensional space that provides a best fit to the training data, with the criterion for “best fit” being measured in some statistical sense. Correspondingly, generalization is equivalent to the use of this multidimensional surface to interpolate the test data.

The construction of a Radial Basis Function (RBF) Neural Network (or RBF-NN) involves three layers of nodes with entirely different roles.

1. The input layer, made up of source nodes, where the inputs are applied
2. The hidden layer, where radial basis functions are applied on the input data; this layer applies a nonlinear transformation from the input space to the hidden space; in most applications the hidden space is of high dimensionality.
3. The output layer, where the outputs are produced.

Radial basis function (RBF) Neural Networks can solve any approximation problem. Park and Sandberg [Park, 93] proved that RBF neural networks (RBF-NN) are capable of universal approximation. Broomhead and Lowe in 1988 [Broomhead, 88] were the first to explore the use of RBFs in the design of NNs and to show how RBF-NNs model nonlinear relationships and implement interpolation. Mitchell [Mitchelli, 86] showed that RBF-NNs can produce an interpolating surface that exactly passes through all the pairs of the training set. However, in applications the exact fit is neither useful nor desirable since it may produce anomalous interpolating surfaces. Poggio and Girosi viewed the learning process as an ill-posed problem, in the sense that the information in the training data is not sufficient to uniquely reconstruct the mapping in regions where the data are not available.

A popular choice of the radial basis functions at the hidden layer is Gaussian functions (i.e. they resemble multidimensional Gaussian probability density functions) with appropriate centers (means) and autocovariance matrices. One of the major differences between RBF neural networks and MLP neural networks (with one hidden layer) is that RBFs (which can be thought of as activation functions) have localized centers. That is, they provide a nonzero output for portions of the input space that is closely concentrated around the center of the RBF. This is not true for the activation functions used in the hidden layers of the MLPs. Another major difference is that if we choose the parameters of the RBFs a priori (e.g., centers of Gaussian and autovariances of Gaussians), then the learning of weights can focus only on the weight parameters converging to the output layer of the RBF network. Hence, in this case, the learning process of an RBF neural network becomes equivalent to the learning process of a single layer perceptron. Since the learning of the weights of a single layer perceptron is much faster process than the learning of weights of an MLP, the convergence to

a solution for such an RBF network can be orders of magnitude faster than the convergence to a solution for the MLP. Of course, the problem remains how to choose the centers and the autocovariance matrices of the Gaussian functions. One of the most straightforward approaches to making this choice was proposed in [Moody, 89]. In [Moody, 89], Moody and Darken proposed a K-means clustering procedure to choose the means of the Gaussian functions and a P-nearest neighbor heuristic to determine the diagonal elements of the autocovariance matrices; the nondiagonal elements of the autocovariance matrices; the nondiagonal elements of the autocovariances matrices are arbitrarily chosen to be equal to zero. The weights converging to the output layer are updated according to a supervised least mean square procedure (e.g., delta rule of a single layer perceptron). This learning approach [Moody, 89] is a classic example of hybrid learning, where unsupervised methods are used to find the parameters (weights) associated with the hidden layer, while a supervised procedure is utilized in learning the weights converging to the output layer. Another way of finding the centers of the Gaussian in an RBF-NN is by utilizing the self-organizing-feature map (SOFM) neural networks (SOFM-NN), introduced by Kohonen [Kohonen, 90]. The K-means procedure and the SOFM procedure to find the centers of the Gaussian in an RBF-NN will be discussed later. Since [Moody, 89], a variety of supervised approaches to learn the parameters of the hidden layer in the RBF-NN have been proposed. The supervised procedures to learn the centers and autocovariance matrices tend to make the training process more time-consuming. On the other hand, supervised procedures to find centers and autocovariances of Gaussian leads us to trained RBF neural networks that tend to generalize better.

2.4.2. Structure of RBF networks

Figure 2.14 depicts the block diagram of a RBF-NN with M input nodes, K hidden nodes (plus the bias node 0) and one output node. The input-output mapping performed by the RBF-NN may be expressed as:

$$y = \sum_{k=1}^K w_k \varphi(\mathbf{x}; \mathbf{v}_k) + w_0 \quad (2.57)$$

The term $\varphi(\mathbf{x}; \mathbf{v}_k)$ is the k^{th} radial basis function that computes the “distance” between an input vector \mathbf{x} and its own center \mathbf{v}_k ; the output signal produced by the k^{th} hidden node is a nonlinear function of the distance. The scaling factor w_k in equation (2.57) represents the weight that connects the k^{th} hidden node to the output node of the network. The constant term w_0 represents the bias.

The input-output mapping performed by the RBF-NN is accomplished in two stages:

- A nonlinear transformation that maps the input space onto an intermediate space
- A linear transformation that maps the intermediate space onto the output space.

The nonlinear transformation is defined by the set of radial basis functions φ_k and the linear transformation is defined by the set of weights $w_k, k = 1, 2, \dots, K$.

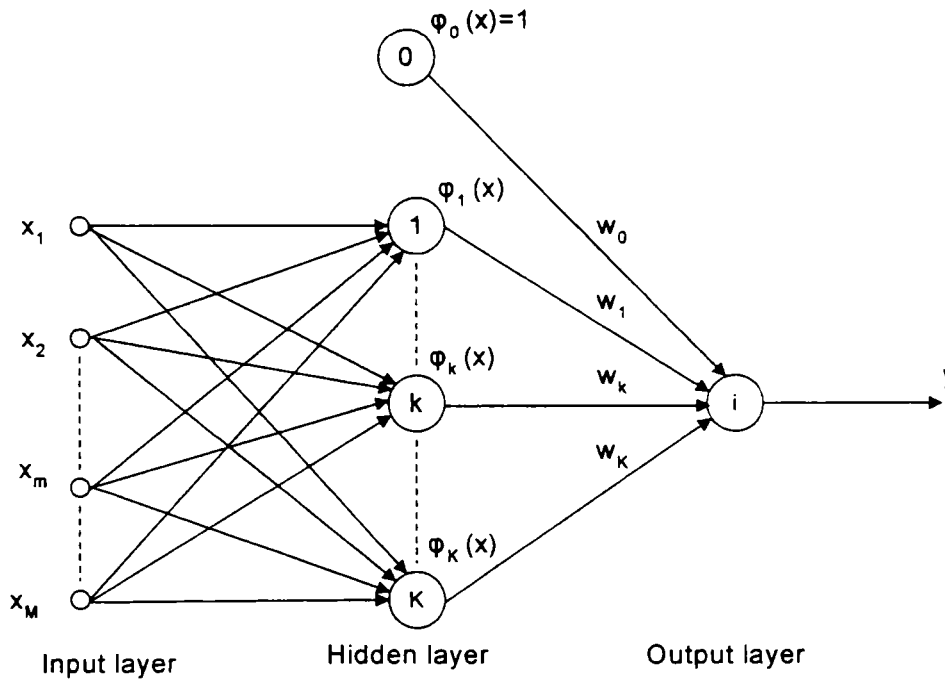


Figure 2.14. Architecture of a RBF-NN

2.4.3. Radial basis functions

At the heart of RBF network is the hidden layer that is defined by a set of radial basis functions from which the network derives its name. The following functions are of particular interest in the study of RBF networks [Haykin, 99]:

1. Gaussian,

$$\varphi(r) = \exp\left(-\frac{r^2}{\sigma^2}\right) \quad (2.58)$$

2. Multiquadratic function

$$\varphi(r) = (r^2 + \sigma^2)^{1/2} \quad (2.59)$$

3. Inverse – multiquadratic function

$$\varphi(r) = \frac{1}{(r^2 + \sigma^2)^{1/4}} \quad (2.60)$$

4. Piecewise linear approximation

$$\varphi(r) = r \quad (2.61)$$

5. Cubic approximation

$$\varphi(r) = r^3 \quad (2.62)$$

6. Thin plate spline

$$\varphi(r) = \left(\frac{r}{\sigma}\right)^2 \ln\left(\frac{r}{\sigma}\right) \quad \text{for some } \sigma > 0 \text{ and } r \geq 0 \quad (2.63)$$

Of these examples, the Gaussian function is the most commonly used in practice. Given an input vector \mathbf{x} , the k^{th} Gaussian radial basis function of the RBF network is defined as follows:

$$\varphi(\mathbf{x}; \mathbf{v}_k) = \exp\left(-\frac{1}{\sigma_k^2} \|\mathbf{x} - \mathbf{v}_k\|^2\right), \quad k = 1, 2, \dots, K \quad (2.64)$$

where \mathbf{v}_k is the center, σ_k is the width and $\|\mathbf{x} - \mathbf{v}_k\|$ denotes the Euclidean distance between \mathbf{x} and \mathbf{v}_k .

Substituting equation (2.64) in (2.57), the input-output mapping realized by a Gaussian RBF network may be reformulated as follows:

$$y = \sum_{k=1}^K w_k \exp\left(-\frac{1}{\sigma_k^2} \|\mathbf{x} - \mathbf{v}_k\|^2\right) \quad (2.65)$$

From a design point of view, the requirement is to select suitable values for the parameters of the K Gaussian radial basis functions, namely σ_k and \mathbf{v}_k , $k = 1, 2, \dots, K$, and solve for the weights of the output layer.

2.4.4. Learning strategies with RBF-NN

There are a number of choices for these functions φ , and all of these choices guarantee that the resulting RBF-NN structure can implement any continuous mapping from an input space of arbitrary dimensionality to an output space of arbitrary dimensionality. The most popular choice for the function φ is a multivariate Gaussian function with an appropriate mean and autocovariance matrix. That is,

$$\varphi_k(\mathbf{x}) = \exp -\frac{1}{2} (\mathbf{x} - \mathbf{v}_k)^T \Sigma_k^{-1} (\mathbf{x} - \mathbf{v}_k) \quad (2.66)$$

where \mathbf{v}_k is the mean vector and Σ_k is the autocovariance matrix of the multivariate Gaussian function corresponding to hidden node k . Given the above expression for the functions φ involved in the hidden layer of the RBF-NN structure, we can see that we have at our disposal a lot of parameters that can be modified to achieve our objective. These parameters are the

mean vectors and the autocovariance matrices of each Gaussian function in the hidden layer and the interconnection weights from the hidden to the output layer. There are four primary learning strategies that have been proposed in the literature for changing the parameters of an RBF-NN.

A. Fixed centers selected at random

In this learning strategy, the means (centers) of the Gaussian functions are chosen randomly from the training data set. In other words, each \mathbf{v}_k is chosen to be equal to one of the training input patterns \mathbf{x} , selected randomly from the training data. In effect, the standard deviation (i.e. width) σ_k for each Gaussian radial basis functions is fixed at the common value:

$$\sigma = \frac{d_{\max}}{\sqrt{2K}} \quad (2.67)$$

where K is the number of centers and d_{\max} is the maximum distance between the chosen centers. This formula ensures that the individual radial basis functions are not too peaked or too flat; both of these extreme should be avoided [Haykin, 99].

The only parameters that would need to be learned in this approach are the linear interconnection weights from the hidden layer to the output layer. These interconnection weights are chosen in a way that minimizes the error function

$$E(\mathbf{w}_i) = \sum_{p=1}^P E^p(\mathbf{w}_i) \quad (2.68)$$

where P represents the number of the training patterns. One way of finding the weights that minimize the aforementioned error function is by following the gradient descent procedure that modifies the weights by an amount proportional to the negative gradient of $E(\mathbf{w}_i)$. The minimum of the above error function is zero if the transformed input patterns at the hidden layer are linearly independent. If an exact solution does not exist, an approximate solution can be found by using the pseudoinverse of a matrix. The constraints can be put into a matrix form

$$\Phi W = D \quad (2.69)$$

where Φ is the $N \times (K+1)$ matrix of the transformed input patterns, the W matrix is a $(K+1) \times 1$ vector of interconnection weights from the hidden-to-output layer, and D is the $N \times 1$ vector of desired outputs. The pseudoinverse approach gives a solution for (2.69) that is of the form

$$W = (\Phi^T \Phi)^{-1} \Phi^T D \quad (2.70)$$

In (2.70) it is assumed that the G matrix has $(K+1)$ linearly independent columns. In the case where the RBF-NN has many output nodes (i.e., I output nodes), the solution will be given by

(2.70), where W is now a matrix $(K+1)*I$ of interconnection weights, and Φ is a matrix $N*I$ of desired outputs [Christodoulou, 01].

B. Self-organized selection of centers

The main problem with the method described above is the fact that the random selection of centers is arbitrary and it might lead to poor performance of the network if the centers are not chosen properly. One way of overcoming this limitation is to use some kind of clustering procedure to define the centers of the Gaussian functions in the RBF-NN. Popular clustering procedures that have been proposed include the K-means algorithm [Tou, 74] and the SOFM [Kohonen, 90]. A clustering approach involves two steps:

- The unsupervised learning algorithm for the selection of the centers of the Gaussian radial basis functions.
- The supervised learning algorithm for the computation of the interconnection weights from the hidden to the output layer.

K-means clustering procedure

The K-means clustering algorithm [Duda, 73] places the centers of the Radial Basis Functions in only those regions of the input space where significant data are present. A description of the K-means algorithm may be found in [Haykin, 99]:

1. Initialization. Choose random values for the initial centers $\mathbf{v}_k(0)$; the only restriction is that these initial values be different. It may also be desirable to keep the Euclidean norm of the centers small.
2. Sampling. Draw a sample vector \mathbf{x} from the input space with a certain probability. The vector \mathbf{x} is input into the algorithm at iteration n .
3. Similarity matching. Let $k(\mathbf{x})$ denote the index of the best-matching (winning) center for input vector \mathbf{x} . Find $k(\mathbf{x})$ at iteration n by using the minimum distance Euclidean criterion:

$$k(\mathbf{x}) = \arg \min_k \|\mathbf{x}(n) - \mathbf{v}_k(n)\| \quad k = 1, 2, \dots, K \quad (2.71)$$

where $\mathbf{v}_k(n)$ is the center of the k^{th} radial basis function at iteration n .

4. Updating. Adjust the centers of the radial basis functions using the update rule:

$$\mathbf{v}_k(n+1) = \begin{cases} \mathbf{v}_k(n) + \mu [\mathbf{x}(n) - \mathbf{v}_k(n)], & k = k(\mathbf{x}) \\ \mathbf{v}_k(n), & \text{otherwise} \end{cases} \quad (2.72)$$

where μ is a learning rate parameter that lies in the range $0 < \mu < 1$.

5. Continuation. Increment n by 1, go back to step 2, and continue the procedure until no noticeable changes are observed in the centers \mathbf{v}_k .

Once the centers are identified by the K-means clustering algorithm, the variances of the Gaussians are chosen to be equal to the mean distance of every Gaussian center from its neighboring Gaussian centers.

A limitation of the K-means clustering algorithm is that it can only achieve a local optimum solution that depends on the initial choice of cluster centers. Consequently, computing resources may be wasted in that some initial centers get stuck in regions of the input space with a scarcity of data points and may therefore never have the chance to move to new locations where they are needed. The network result in possible an unnecessarily large network. To overcome this limitation of the conventional K-means clustering algorithm, [Chen, 95] proposed the use of an enhanced K-means clustering algorithm due to [Chinunrueng, 94] that is based on a cluster variation weighted measure that enables the algorithm to converge to an optimum or near optimum configuration, independent of the initial center locations [Haykin, 99].

Having identified the individual centers of the Gaussian radial basis functions and their common widths using the K-means clustering algorithm or its enhanced version, the next stage is to estimate the interconnection weights from the hidden to the output layer. A simple method for this estimation is the least-mean-square (LMS) algorithm (appendix 4). The vector of output signals produced by the hidden units constitutes the input vector to the LMS algorithm. Note also that the K-means clustering algorithm for the hidden units and the LMS algorithm for the output unit(s) may proceed with their own individual computations in a concurrent fashion, thereby accelerating the training process.

The Self-Organizing Feature Map (SOFM) Clustering Procedure

The SOFM-NN consists of an input layer of nodes, where the inputs to the NN are applied, and an output layer of nodes, where the groupings of the inputs are formed. Most often, the nodes in the output layer of an SOFM-NN are organized in a two-dimensional array (Figure 2.15, where an SOFM-NN, with many output nodes, organized in a two-dimensional array, and two input nodes, is shown).

Each input is fully connected to every output node, and a weight is assigned to each connection. Training is performed in an unsupervised way using the Kohonen learning algorithm [Kohonen, 90]. The Kohonen learning algorithm belongs to the broader class of competitive algorithms, which can be viewed as a procedure that learns to group input

patterns in clusters in a way inherent to the data. To train the SOFM-NN, continuous valued input vectors are presented in random sequence to the network. The mapping from the external input patterns to the network's activity patterns is realized by correlating the input patterns with the connection weights. After enough input patterns have been presented, weights converging to output nodes of the SOFM-NN specify cluster centers that represents the input patterns.

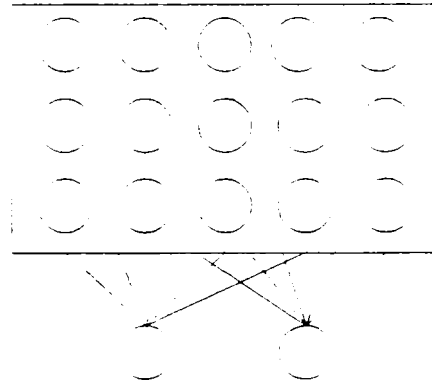


Figure 2.15. The SOFM-NN with two input nodes and output nodes organized in a two-dimensional array

The two most central issues to the Kohonen learning algorithm are the weight adaptation process and the idea of a topographical neighborhood of nodes. The network operates in two phases: the similarity matching phase and the weight adaptation phase. The SOFM-NN can be described in a number of easy to implement steps [Christodoulou, 01]:

1. Initializing the network. Define v_{km} , $1 \leq k \leq K$ and $1 \leq m \leq M$, to be the weight from input node m to output node k , where M is the dimensionality of the input patterns and K is the number of nodes in the network. Choose the number of training iterations equal to n_{max} .
2. Similarity matching phase. Present an input pattern $\mathbf{x}(p)$ from the training collection of P input patterns and compute the Euclidean distance of this input pattern from each weight vector associated with the output nodes.

$$d_k = \sum_{m=1}^M (x_m(p) - v_{km})^2 \quad (2.73)$$

3. Selecting the minimum distance. Find the index k_{max} of the output node that minimizes d_k . That is

$$d_{k_{max}} = \min_k d_k \quad (2.74)$$

4. Weight adaptation phase. Update the weights converging to node k_{max} and the weights converging to all the other nodes j that are in the neighborhood $N_{k_{max}}$ of the winning node k_{max} . Specifically,

$$v_{km}(n+1) = v_{km}(n) + \mu(t)h(k, k_{\max})(n) [x_m(p) - v_{km}(n)]$$

$$\text{for } k \in N_{k_{\max}(n)}, 1 \leq m \leq M \quad (2.75)$$

where

$$h(k, k_{\max}) = \exp\left(\frac{-|r_k - r_{k_{\max}}|^2}{2\sqrt{\sigma(n)}}\right) \quad (2.76)$$

where $r_k - r_{k_{\max}}$ is the distance between the nodes k and k_{\max} . All the weights converging to nodes that are not the winning node or the nodes in the neighborhood of the winning node remain unaltered.

5. Checking the stopping criterion. If $n = n_{\max}$, stop. Otherwise, go to step 2 and present the next in sequence input pattern.

C. Supervised selection of centers

Although learning strategies A and B are simple to implement and converge on a solution relatively quickly, they have been criticized because of their heuristic way of choosing the centers and covariance matrices of the Gaussian functions at the hidden layer. Heuristic approaches of choosing the centers and covariances of the Gaussian functions lead to RBF-NNs that are suboptimal (i.e., they do not generalize well when exposed to data with which they have not been trained). One way around this problem is to apply a gradient descent procedure to choose the centers of the Gaussians or the centers and the covariances of the Gaussians. In the following are demonstrated the equations that pertain to the changes of weights and the centers of the Gaussians in the case where the covariance matrices are assumed to be diagonal of constant variance.

The first step in the development of this supervised learning procedure is to define the error function associated with the RBF-NN, when p^{th} input pattern is presented at the input layer of the RBF-NN

$$E^p(\mathbf{w}) = \frac{1}{2} \sum_{i=1}^I [d_i^2(p) - y_i^2(p)]^2 \quad (2.77)$$

The gradient descent procedure tells us that the change of w_{ik} should be proportional (constant of proportionality designated as μ) to the negative gradient. That is,

$$\Delta w_{ik} = -\mu \frac{\delta E^p(\mathbf{w})}{\delta w_{ik}} \quad (2.78)$$

It can be easily obtained that

$$\Delta w_{ij} = \mu [d_i^2(p) - y_i^2(p)] \phi_k[\mathbf{x}(p)] \quad (2.79)$$

As in the MLP case, it can be seen that a weight from a hidden to an output nodes needs to be changed by an amount that is proportional to the error of the node to which the weight converges and proportional to the output of the node from which the weight emanates.

To calculate the amount of change required by each one of the centers of the Gaussian functions, we apply again the gradient descent rule to determine that

$$\Delta \mathbf{v}_k = -\mu \frac{\delta E^p(\mathbf{w})}{\delta \mathbf{v}_k} \quad (2.80)$$

and finally,

$$\Delta \mathbf{v}_k = \mu \sum_{i=1}^I [d_i^2(p) - y_i^2(p)] w_{ik} \phi_k[\mathbf{x}(p)] \Sigma_k^{-1}(\mathbf{x}(p) - \mathbf{v}_k) \quad (2.81)$$

In the special case where the variances of the Gaussian functions in the hidden nodes are equal, equation (2.81) simplifies to

$$\Delta \mathbf{v}_k = \mu \phi_k[\mathbf{x}(p)] \sum_{i=1}^I [d_i^2(p) - y_i^2(p)] w_{ik} (\mathbf{x}(p) - \mathbf{v}_k) \quad (2.82)$$

It is not difficult to see that the weight change equation, (2.82) resembles the weight change equations obtained for the MLP case. The error terms for the output layer nodes and the hidden layer nodes of the RBF-NN are identical with the error terms produced in the MLP case. The only term that is missing, compared to the MLP expression, is the output of the node with which the corresponding \mathbf{v} component is associated.

The error function that we focused on above is the error associated with a particular pattern presentation (input pattern p). In this case, we apply weight changes after every pattern presentation in a continuous update or pattern-by-pattern update. If instead of considering the error function of (2.77), the cumulative error function (the sum of squared errors over all the output nodes and input patterns) defined as follows

$$E(\mathbf{w}) = \sum_{p=1}^P E^p(\mathbf{w}) = \sum_{p=1}^P \sum_{i=1}^I [d_i^2(p) - y_i^2(p)]^2 \quad (2.83)$$

is considered, then weight updates are applied after the presentation of all the input patterns in the training list in a batch update. The corresponding change of the weights and centers equations are provided as follows:

$$\Delta w_{ik} = \mu \sum_{p=1}^P [d_i^2(p) - y_i^2(p)] \phi_k[\mathbf{x}(p)] \quad (2.84)$$

$$\Delta \mathbf{v}_k = \mu \sum_{p=1}^P \sum_{i=1}^I [d_i^2(p) - y_i^2(p)] w_{ik} \phi_k[\mathbf{x}(p)] \Sigma_k^{-1}(\mathbf{x}(p) - \mathbf{v}_k) \quad (2.85)$$

for Gaussians with unequal variances, and

$$\Delta \mathbf{v}_k = \mu \sum_{p=1}^P \sum_{i=1}^I [d_i^2(\mathbf{p}) - y_i^2(\mathbf{p})] w_{ik} \phi_k[\mathbf{x}(\mathbf{p})] (\mathbf{x}(\mathbf{p}) - \mathbf{v}_k) \quad (2.86)$$

for Gaussians with equal variances.

D. Supervised selection of centers and variances

In the previous learning strategy, it is described a procedure that finds the centers of the Gaussian through a supervised procedure. It is possible to extend this procedure to the case where not only the centers of the Gaussians but also the covariances matrices of the Gaussians are chosen in a supervised way. Once more, the objective is to minimize the sum of the squared differences between the actual and the desired outputs, defined in equation (2.77) where

$$y_i(\mathbf{p}) = \sum_{k=1}^J w_{ik} \phi_k[\mathbf{x}(\mathbf{p})] \quad (2.87)$$

If the gradient descent procedure is applied on the error function with independent parameters w_{ik} , \mathbf{v}_k and Σ_k we obtain the following equations for the change of these parameters:

$$\Delta w_{ik} = \mu_w [d_i^2(\mathbf{p}) - y_i^2(\mathbf{p})] \phi_k[\mathbf{x}(\mathbf{p})] \quad (2.88)$$

$$\Delta \mathbf{v}_k = \mu_v \sum_{i=1}^I [d_i^2(\mathbf{p}) - y_i^2(\mathbf{p})] w_{ik} \phi_k[\mathbf{x}(\mathbf{p})] \Sigma_k^{-1} (\mathbf{x}(\mathbf{p}) - \mathbf{v}_k) \quad (2.89)$$

$$\Delta \Sigma_k^{-1} = \mu_\Sigma \sum_{i=1}^I [d_i^2(\mathbf{p}) - y_i^2(\mathbf{p})] w_{ik} \phi_k[\mathbf{x}(\mathbf{p})] (\mathbf{x}(\mathbf{p}) - \mathbf{v}_k) (\mathbf{x}(\mathbf{p}) - \mathbf{v}_k)^T \quad (2.90)$$

where in the above equations μ_w , μ_v and μ_Σ are the corresponding learning rates of the parameters w , v and Σ , respectively. If the error function of interest is the cumulative error function $E(\mathbf{w})$ of equation (2.83), we first calculate the error changes of the weights and network parameters (centers, covariance functions of Gaussians) for every pattern presentation and then we sum up these changes to obtain the required weight/parameter change. The weight/parameter change for every pattern presentation is provided by equations (2.88)-(2.90).

E. Comparison of RBF-NN learning strategies

The methods that first choose the centers and the variances of the Gaussians through some kind of heuristic procedure tend to converge to a solution faster than the methods that employ supervised means of changing centers and variances of the Gaussians. On the other hand, heuristic procedures to choose the centers and variances of the Gaussians sometimes

leads to trained networks that do not generalize very well. In [Wettschereck, 92] was experimented a number of approaches to improve the performance of a RBF-NN. The conclusions of this study were that

- Supervised selection of the centers of the Gaussians improved the performance of the network considerably, compared to heuristic techniques of choosing the Gaussian centers,
- Simultaneous supervised learning of the centers and variances of the Gaussians exhibited an inferior performance compared to the network where only supervised selection of centers was implemented.

However, it was admitted that these results might be biased because of the fact that only a specific problem was tested.

2.4.5. A RBF-NN algorithm

In this section, a step-by-step complete training process corresponding to the supervised selection of centers is presented [Christodoulou, 01].

1. Select initial values for the weights from hidden to output layers. These weights are chosen to be small random values. Select initial values for the centers of the Gaussian in the hidden layer. These centers are randomly chosen from the training data. Select initial values for the diagonal elements of the covariances of the Gaussian functions. These variances are all chosen to be equal to some constant. The off-diagonal elements of the covariances of the Gaussians are chosen to be equal to zero.
2. Present the p^{th} input pattern at the input layer of the RBF-NN.
3. Calculate the outputs of the nodes in the hidden and output layers of the RBF-NN, according to (2.91) and (2.92).

$$\varphi_k[\mathbf{x}(p)] = \exp\left\{-\frac{1}{2}(\mathbf{x}(p) - \mathbf{v}_k)\Sigma_k^{-1}(\mathbf{x}(p) - \mathbf{v}_k)^T\right\} \quad (2.91)$$

$$y_i^2(\mathbf{x}(p)) = \sum_{k=0}^K w_{ik} \varphi_k[\mathbf{x}(p)] \quad (2.92)$$

4. Compare the actual outputs at the output layer and the desired outputs. If $y_i^2(p) = d_i^2(p)$ for $1 \leq i \leq I$, go to step 5. If $y_i^2(p) \neq d_i^2(p)$ for some i , proceed to change the weight/parameter values as follows:

$$\Delta w_{ik} = \mu [d_i^2(p) - y_i^2(p)] \varphi_k[\mathbf{x}(p)] \quad (2.93)$$

$$\Delta \mathbf{v}_k = \mu \phi_k[\mathbf{x}(p)] \sum_{i=1}^I [d_i^2(p) - y_i^2(p)] \mathbf{w}_{ik} (\mathbf{x}(p) - \mathbf{v}_k) \quad (2.94)$$

5. If $p = P$ and the cumulative error function $E(\mathbf{w})$ is smaller than a prespecified threshold, we consider the training completed. If $p = P$ and $E(\mathbf{w})$ is larger than the prespecified threshold, then we return to step 2 starting with the first input pattern of index $p = 1$. If $p \neq PT$, we return to step 2, by increasing the pattern index p by one.

Note that the above algorithm procedure is the continuous update version of the algorithm. The periodic update is very similar, with the only difference being that after every pattern presentation we do not apply weight and parameter changes; instead we wait until all the parameters are presented to implement a cumulative weight/parameter change according to (2.84) and (2.86). The algorithmic procedure, when both the centers and the covariances of the Gaussian are changing, is also very similar with the only difference being that in step 4 we need to apply the changes to the covariance matrices designated by (2.90). Also, if the unsupervised selection of the parameters is chosen, the algorithmic procedure is not that different either. In this case, the centers and the covariances matrices will be chosen in step 1 and stay fixed thereafter. Beyond this point only weight changes will be enforced according to (2.88) [Christodoulou, 01].

2.4.6. Issues with RBF-NN learning

The weights between the hidden and the output layer are, in most cases, initialized to random numbers that are uniformly distributed in a small interval of values, symmetric around zero. The centers of the Gaussians in learning strategies C and D could initially be chosen to be the cluster centers of a K-means clustering applied on the input patterns, or K randomly chosen input patterns. Once the centers of the Gaussian are chosen, a number of heuristics can be applied to find the initial elements of the covariance matrices of the Gaussians. The premature saturation problem will occur if the activation functions at the output layer of the RBF-NN are of the sigmoid or hyperbolic tangent type. These types of activation functions are most often used when the types of problems under consideration are classification problems. All the equations presented so far for the RBF-NN learning strategies assume that the activation functions at the output layer are linear. Linear activation functions do not suffer from the premature saturation problem. The RBF-NN can operate in the continuous update training mode (where changes of the weights/parameters are applied after every pattern presentation), or in the periodic update training mode (where changes of the

weights/parameters occur after the presentation of all the patterns), or finally in the hybrid training mode (where changes of the weights parameters occur after a fixed number of training pattern presentation; this fixed number is larger than one and smaller than the number of patterns of P). The stopping criteria for RBF-NN learning are identical with the stopping criteria mentioned for the backpropagation learning method. The number of hidden layers is not an issue with the RBF-NN since we always have one hidden layer of nodes. The number of nodes in the hidden layer should be chosen as large as possible to take advantage of the increased dimensionality of the transformed space compared with the dimensionality of the input space. At the same time, the number of nodes in the hidden layer should be as small as possible if we are committed to designing the smallest possible NN structure. All the variations of the backpropagation algorithm discussed in this chapter, can also be applied to the RBF-NN algorithm, since they both are gradient descent procedures applied to same error function [Christodoulou, 01].

2.4.7. The General Regression Neural Network

The general regression neural network (GRNN) is an NN architecture that shares a lot of similarities with the RBF-NN. In this section we discuss the necessary background information and the specifics of this architecture. The information included below is obtained from [Christodoulou, 01].

Regression is the least-mean-squares estimation of the value of a variable of interest based on observations of other variables that are related with the variable of interest. The term general regression implies that the regression surface is not restricted by being linear. If the variable of interest is the future value of an observed variable, the GRNN is a predictor. If they are dependent variables related to input variables related to input variables in a process, plant, or system, the GRNN can be used to model the process, plant or system.

Figure 2.16 is the overall network topology implementing the GRNN. As it can be seen from the figure, the NN consists of two layers of nodes (excluding the input layer where the input data are applied). The hidden layer units are very similar with the hidden layer units of the RBF-NN discussed so far. Hence, the outputs of these units are of the form

$$\varphi_k[\mathbf{x}] = \exp\left[-\frac{(\mathbf{x} - \mathbf{v}_k^x)^T(\mathbf{x} - \mathbf{v}_k^x)}{2\sigma^2}\right] \quad (2.95)$$

when \mathbf{v}_k^x are the corresponding clusters for the inputs and \mathbf{v}_k^y are the corresponding clusters for the outputs obtained by applying a clustering technique of the input/output data that produces K cluster centers.

v_k^y is defined as

$$v_k^y = \sum_{y(p) \in \text{cluster } k} y(p) \quad (2.96)$$

N_k is the number of input data in the cluster center k , and

$$d(\mathbf{x}, \mathbf{v}_k^x) = (\mathbf{x} - \mathbf{v}_k^x)^T (\mathbf{x} - \mathbf{v}_k^x) \quad (2.97)$$

with

$$v_k^x = \sum_{\mathbf{x}(p) \in \text{cluster } k} \mathbf{x}(p) \quad (2.98)$$

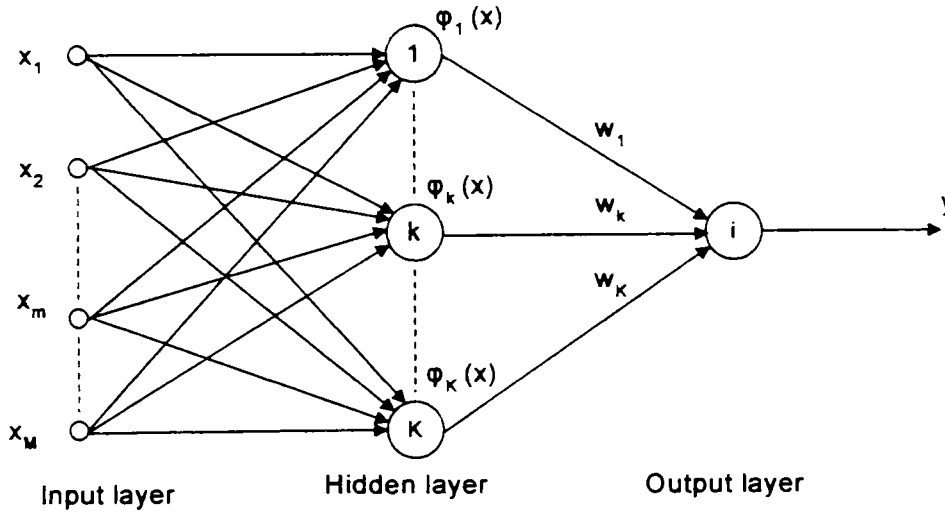


Figure 2.16. General regression neural network

The outputs of the hidden layer nodes are multiplied with appropriate interconnection weights to produce the output of the GRNN. The weights for the hidden node k (i.e., w_k) is equal to

$$w_k = \frac{v_k^y}{\sum_{k=1}^K N_k \exp\left[-\frac{d(\mathbf{x}, \mathbf{v}_k^x)^2}{2\sigma^2}\right]} \quad (2.99)$$

In the case where we have to estimate a vector \mathbf{y} instead of a scalar y , the output layer consists as many nodes as the number of components of the vector \mathbf{y} and the weights from the hidden layer nodes to output nodes are chosen according to (2.99), where now v_k^y depends on the component of the output vector that is estimated.

2.4.8. Comparison of RBF and MLP neural networks

RBF networks and MLP are examples of nonlinear layered feedforward networks. They are both universal approximators. These two networks differ from each other in several important respects.

1. An RBF neural network has a single hidden layer, whereas an MLP may have one or more hidden layers.
2. The kernel characterizing a hidden unit of the RBF network is defined, for example by the Gaussian function

$$\varphi(\mathbf{x}; \mathbf{x}_k) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}_k\|^2}{2\sigma^2}\right), \quad k = 1, 2, \dots, K, \quad (2.100)$$

where \mathbf{x} is the input vector, \mathbf{x}_k is the center of the k^{th} unit and σ^2 is a common bandwidth. The input-output map realized by the RBF network with K hidden units is then defined by

$$y = \sum_{k=1}^K w_k \varphi(\mathbf{x}; \mathbf{x}_k) \quad (2.101)$$

By contrast, the kernel characterizing a hidden unit of the multilayer perceptron is defined, for example, by the logistic function

$$\varphi(\mathbf{x}; \mathbf{x}_k) = \frac{1}{1 + \exp(-\mathbf{x}_k^T \mathbf{x})}, \quad k = 1, 2, \dots, K. \quad (2.102)$$

The Radial Basis Function neural network is a local approximator, whereas the Multilayer Perceptron is a global approximator.

3. Mobile radio channels

3.1. Introduction

The mobile radio channel places fundamental limitation on the performance of wireless communication systems. The transmission between the transmitter and the receiver can vary from simple line-of-sight to one that is severely obstructed by buildings, mountains and foliage. Unlike wired channels that are stationary and predictable, radio channels are extremely random and do not offer easy analysis. Even the speed of motion impacts how rapidly the signal level fades as a mobile terminal moves in space [Rappaport, 96].

In the mobile radio environment, a part of the electromagnetic energy radiated by the transmitting antenna reaches the receiver antenna by propagating through different paths. Along these paths, interactions that are commonly referred to as propagation mechanisms may occur between the electromagnetic field and various objects. Possible interactions are specular reflection on large plane surfaces, diffuse scattering from surfaces exhibiting small irregularities or from objects of small size, transmission through dense material like walls or floors, shadowing by obstacles like trees, etc. The attributes small and large are to be understood here with respect to the wavelength. A detailed description of these propagation mechanisms is given in Chapter 4.

The propagation of electromagnetic waves either near the ground or inside a building due to diffraction, scattering, reflection and absorption of the incoming signal is broken into several components that are attenuated and delayed differently. The signal at the receiver antenna is thus composed of a direct component and a delayed, scattered component. The direct path can be obstructed, depending on the antenna location and shadowing conditions. The degree of shadowing varies very strongly with the movement of the mobile antenna, leading to equivalent time fluctuations of the received power of the direct ray and delayed components.

As a result of the multipath propagation, the received signal presents rapid fluctuations that are characterized as fast fading or Rayleigh fading. In fast fading, the received signal power may vary by as much as three or four orders of magnitude (30 or 40 dB) when the receiver is moved by only a fraction of a wavelength. The median value of the received signal strength also fluctuates due to large-scale variations along the path. The median value fadings

are defined as slow fades or log-normal fading. Typically, the local average received power is computed by averaging signal measurements over a measurement track of 5λ to 40λ . When the received signal also includes the line-of-sight component, the envelope is Rice-distributed.

3.2. Representation of a mobile radio signal

The field strength can be represented as a function of distance in space (the spatial domain) or as a function of time (the time domain). The received field strength (the envelope $r(x)$ of a received signal $s(x)$ along x-axis in space) show severe fluctuation when the mobile unit is away from the base station. Field strengths $r(x)$ can be studied either by associating them with geographical locations or by averaging a length of field strength data to obtain a so-called local mean at each corresponding point. The speed of the mobile unit must remain constant while the data are measured. Since the speed is kept constant, the time axis can be converted to spatial axis. Both field strength representations are useful. The representation $r(t)$ in the time domain is used to study the signal fading phenomenon. The representation $r(x)$ in the spatial domain is used to generate the propagation path loss curve.

The mobile radio signal is received while the mobile unit is in motion. In this situation the field strength (also called the fading signal) of a received signal with respect to time t , or space x , is observed. When the operating frequency becomes higher, the fading signal becomes more severe. The average signal level of the fading signal $\bar{r}(x)$ or $\bar{r}(t)$ decreases as the mobile unit moves away from the base station transmitter.

3.3. Fadings

The received signal strength $r(t)$ can be artificially separated into two parts by cause: long-term fading $m(t)$ and short-term fading $r_0(t)$ as

$$r(t) = m(t) r_0(t) \quad \text{or} \quad r(x) = m(x) r_0(x)$$

Long-term fading is the average or envelope of the received fading signal. It is also called a local mean since along the long-term fading each value corresponds to the mean average of the field strength at each location point. The estimated local mean $\hat{m}(x_1)$ at point x_1 along x-axis can be expressed mathematically as

$$\hat{m}(x) = \frac{1}{2L} \int_{x_1-L}^{x_1+L} r(x) dx = \frac{1}{2L} \int_{x_1-L}^{x_1+L} m(x) r_0(x) dx \quad (3.1)$$

Assume that $m(x_1)$ is the true local mean, then at point x_1

$$m(x = x_1) = \hat{m}(x = x_1) \quad x_1 - L < x < x_1 + L \quad (3.2)$$

when L is properly chosen and the estimated local mean $\hat{m}(x_1)$ becomes

$$\hat{m}(x_1) = m(x_1) \frac{1}{2L} \int_{x_1-L}^{x_1+L} r_0(x) dx \quad (3.3)$$

To let $\hat{m}(x_1)$ approach $m(x_1)$, the following relationship holds

$$\frac{1}{2L} \int_{x_1-L}^{x_1+L} r_0(x) dx \rightarrow 1 \quad (3.4)$$

The length L will be determined in section 3.4. The long-term signal fading $m(x)$ is mainly caused by terrain configuration and the built environment between the base station and the mobile unit. The terrain configuration causes local mean (long-term fading) attenuation and fluctuation, whereas the human-made environment also causes short-term fluctuation (fading) in signal reception. Under certain circumstances the fluctuation of a long-term fade caused by the terrain configuration can form a log-normal distribution because of the statistical nature of the fluctuation. We must differentiate between the terms “radio path” and “mobile path”. The former is the path that the radio wave travels and the other is the path that the mobile unit travels. In [Lee, 93] are considered two cases: one is when mobile unit is circling around the base station and the other is when the mobile unit is moving away from the base station. In the first case the radio path does not correspond to the mobile path. In the second case the fluctuation of the long-term fading received is affected by the radial terrain contour where the mobile travels in a certain direction. The radio path corresponds to the mobile path and the terrain contour where the mobile unit travels has a strong correlation with the received signal.

Short-term fading is mainly caused by multipath reflections of a transmitted wave by local scatterers such as houses, buildings and other human-built structures or by natural obstacles such as forest surrounding a mobile unit. It is not caused by a natural obstruction such as mountain or hill located between the transmitting site and the receiving site.

3.4. Obtaining meaningful propagation loss data from measurements

In a mobile radio environment the irregular configuration of the natural terrain, the various shapes of architectural structures, changes in weather and changes in foliage conditions make the predicting of propagation loss very difficult. In addition the signal is

received while the mobile unit is in motion. There is no easy analytic solution to this problem. Combining both statistics and electromagnetic theory helps to predict the propagation loss with greater accuracy.

The local mean can be obtained by averaging a suitable spatial length L over a piece of raw data as shown in Figure 3.1.

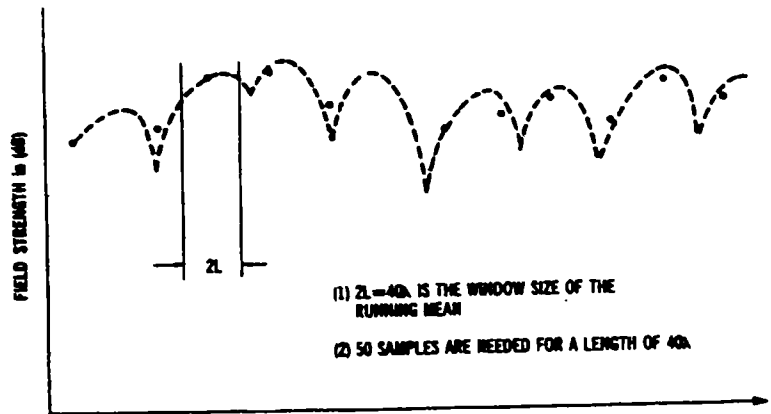


Figure 3.1. Obtaining the local mean

The length L can be treated as an average window over a long piece of raw data. If the length L is too short, the short-term variation cannot be smoothed out and will affect the local mean. If the length L is too long, the averaged output cannot represent the local mean since it washes out the detailed signal changes due to terrain variation. Therefore it is essential that the suitable length L to be determined.

Determining the length L [Lee, 93]

Let the short-term fading r_0 be a Rayleigh fading. Then

$$\langle \hat{m}(x) \rangle = \frac{\sqrt{\pi}}{2} (\sqrt{\bar{r}^2}) \quad (3.5)$$

where \bar{r}^2 is the average power of the short-term fading or $\sqrt{\bar{r}^2}$ is an RMS value of r . This equation shows that the true mean equals the mean of the sample mean; $m(x) = \langle \hat{m}(x) \rangle$.

$$\sigma_m^2 = \langle \hat{m}^2(x) \rangle - \langle \hat{m}(x) \rangle^2 = \frac{1}{4L} \int_0^{2L} \left(1 - \frac{y}{2L}\right) J_0^2(\beta y) dy \quad (3.6)$$

The $1 \sigma_m$ spread is defined as

$$1 \sigma_m \text{ spread} = 10 \log \frac{1 + \sigma_m}{1 - \sigma_m} [\text{dB}] \quad (3.7)$$

The computed results of equations (3.6) and (3.7) are given in Table 3.1.

Table 3.1. σ_m versus $2L$

$2L$	σ_m	$1 \sigma_m$ Spread [dB]
5	0.165	3
10	0.122	2.1
20	0.09	1.56
40	0.06	1

As can be seen from Table 3.1, the length $2L = 40\lambda$ is desirable because its $1 \sigma_m$ spread for 40λ approaches 1 dB. The 40λ is considered to be the proper length to use in smoothing out the Rayleigh fading. If the length $2L$ is shorter than 40λ , the average output would retain only a weaker portion of Rayleigh fading. If the length $2L$ is greater than 40λ , the excessive length of averaging would smooth out the local mean information, which it is not supposed to do. Therefore, $2L = 40\lambda$ is considered to be the appropriate length. However, in practice L in the 20λ to 40λ range is acceptable.

Determining the number of sample points required over 40λ [Lee, 93]

Since most data processing is done digitally, what is the proper number of samples required for a piece of analog data? Experimental autocorrelation has shown that a separation of 0.8λ is required for a correlation coefficient below 0.2 between two adjacent samples. Then 50 weak-correlated samples would be needed to represent a length of 40λ in digital form. It must be determined whether 50 samples are enough to obtain an average value over a length of 40λ with great confidence. From the ensemble average \bar{r}_j of a set of N variables, r_i along a piece of M -sample data is shown as

$$\bar{r}_j = \frac{\sum_{i=(j-1)N+1}^{jN} r_i}{N} \quad 1 \leq j \leq \frac{M}{N} \quad (3.8)$$

We define \hat{m} and $\hat{\sigma}$ as the mean and standard deviation of \bar{r}_j , respectively. \bar{r}_j is always a Gaussian variable if all N variables r_i are added in linear scales. Since r_i itself is a Rayleigh with both mean m and standard deviation σ_r expressed in linear values, it can be shown that

$$\hat{m} = \langle \bar{r}_j \rangle = m \quad (3.9)$$

$$\hat{\sigma} = \left(\langle \bar{r}_j^2 \rangle - \langle \bar{r}_j \rangle^2 \right)^{1/2} = \frac{\sigma_r}{\sqrt{N}} \quad (3.10)$$

Applying the confidence interval of 90%

$$P\left(-1.65 \leq \frac{\bar{r}_j - \hat{m}}{\hat{\sigma}} \leq 1.65\right) = 90\% \quad (3.11)$$

Equation (3.11) can be restated

$$P(\hat{m} - 1.65\hat{\sigma} \leq \bar{r}_j \leq \hat{m} + 1.65\hat{\sigma}) = 90\% \quad (3.12)$$

Equation (3.11) shows the 90% confidence interval CI of \bar{r}_j is within $\hat{m} \pm 1.65\hat{\sigma} - \bar{r}_j$ approaches \hat{m} if $\hat{\sigma}$ becomes smaller.

Inserting equation (3.9) and (3.10) into equation (3.36) yields

$$P\left(m_r - 1.65 \frac{\sigma_r}{\sqrt{N}} \leq \bar{r}_j \leq m_r + 1.65 \frac{\sigma_r}{\sqrt{N}}\right) = 90\%$$

or

$$P\left[\left(1 - \frac{1.65}{\sqrt{N}} \cdot \frac{\sigma_r}{m}\right)m \leq \bar{r}_j \leq \left(1 + \frac{1.65}{\sqrt{N}} \cdot \frac{\sigma_r}{m}\right)m\right] = 90\% \quad (3.13)$$

Inserting the values of m and σ_r of Rayleigh distribution:

$$m = \frac{\sqrt{\pi}}{2}(\sqrt{\bar{r}^2}) \text{ and } \sigma_r = \frac{\sqrt{2-\pi}}{2}(\sqrt{\bar{r}^2})$$

in equation (3.13) yields

$$P\left[\left(1 - \frac{1.65}{\sqrt{N}} \cdot \sqrt{\frac{4-\pi}{\pi}}\right)m \leq \bar{r}_j \leq \left(1 + \frac{1.65}{\sqrt{N}} \cdot \sqrt{\frac{4-\pi}{\pi}}\right)m\right] = 90\%$$

Simplifying the above equation we obtain the following expression

$$P\left[\left(1 - \frac{0.8625}{\sqrt{N}}\right)m \leq \bar{r}_j \leq \left(1 + \frac{0.8625}{\sqrt{N}}\right)m\right] = 90\% \quad (3.14)$$

The 90% confidence interval (CI) expressed in dB is

$$CI = 20 \log \left[\frac{\left(1 + \frac{0.8625}{\sqrt{N}}\right)}{m} \right] = 20 \log \left(1 + \frac{0.8625}{\sqrt{N}} \right) \quad (3.15)$$

Let $N = 50$; equation (3.39) becomes

$$90\% \text{ CI} = 1 \text{ dB} \quad (3.16)$$

The estimated value of \bar{r}_j with $N = 50$ and $2L = 40\lambda$ for a 90% confidence interval is within 1 dB of its true mean value. If N is reduced to 36, the 90% confidence interval increases to 1.17 dB of its mean.

Perhaps using 36 or up to 50 samples in an interval of 40λ wavelengths is an adequate averaging process for obtaining the local means. A simpler way of obtaining local means is to use a running mean with a 40λ window. For a low frequency operation, we may have to take an interval of 20λ for obtaining local means. The reason is that the terrain contour may change at a distance greater than 20λ when the wavelength becomes longer [Lee, 93].

3.5. Modeling requirements

To characterize the mobile channels requires a complete knowledge of the propagation parameters mentioned above for all environments where the system will operate. Conducting measurements to obtain all propagation parameters for all possible environments is an impossible task and for a limited number of environments a time-consuming exercise. In addition, testing a new system requires repetition of the measurements with the same propagation medium (i.e. a stable propagation environment). Therefore a propagation model is required that provides all the parameters which characterize the mobile channel.

The equipment and the mobile radio systems design engineers require complete channel characterization. Propagation models that apply to a wide variety of locations, but in a limited frequency band and for limited distances, are needed for general system design, such as when systems are being developed that will operate in many locations. When a given performance objective is not to be met in a known location, the specific system design requires a propagation model that accounts for relevant environmental and topographical information.

4. Propagation mechanisms for mobile communication systems

4.1. Introduction

The purpose of this chapter is to introduce the propagation mechanisms for mobile radio systems in order to be used in ANN based models.

This chapter describes the radio propagation mechanisms. The propagation mechanisms are examined in order to help the development of propagation prediction models and to enhance the understanding of electromagnetic wave propagation phenomena involved when dealing with radio transmission in mobile communication environments.

The radio propagation phenomena do not depend on the environment considered. Considering all existing radio propagation phenomena, the most important one must be identified and investigated in order to improve the modeling of the mobile radio communication channel or of the prediction of radio coverage and signal quality in radio communication systems. The radio propagation phenomena to be identified as the most important depend on the environment and differ whether we consider flat terrain covered with grass, or brick houses in a suburban area, or buildings in a modern city center, etc. Propagation models are more efficient when only the dominant phenomena are taken into account. Which radio propagation phenomena need to be taken into account and in how much details do they have to be considered will differ if we are interested in modeling the average signal strength, delay spread or any other characteristics.

The mobile radio environment causes some special difficulties to the investigation of the propagation phenomena [COST231, 99]:

1. The distance between a base station and a mobile unit ranges from some meters to several kilometers.
2. Man-made structures and natural features have size ranging from smaller too much larger than a wavelength and affect the radio wave propagation.
3. The description of the environment is not usually available in very much detail.

The main propagation mechanisms usually taken into account when modelling the radio propagation in macro – cell, micro – cell and indoor environments are reflection,

diffraction, scattering, absorption and guide wave. For different propagation mechanisms the range dependence of the field strength is given in the following:

- For specular reflection the field is proportional to $(d_1+d_2)^{-1}$ (d_1 and d_2 are the distances from the reflection point to the transmitter and receiver, respectively);
- For single diffraction, the field is proportional to $(d_1 / d_2 (d_1+d_2))^{-0.5}$ (d_1 and d_2 are the distances from the diffraction point to the transmitter and receiver, respectively);
- For multiple diffraction and for a source illuminating all edges, the field is proportional to $d^{-1.9}$ (d is the distance between transmitter and receiver);
- For volume scattering and rough surface scattering, the field is proportional to $(d_1 d_2)^{-1}$ (d_1 and d_2 are the distances from the scattering object to the transmitter and receiver, respectively);
- For penetration and absorption, the field is mainly attenuated by a constant;
- For the wave guiding phenomena, the logarithm of the field is proportional to d (d is the distance between transmitter and receiver).

In macro - cells

Forward propagation including multiple diffractions over terrain and buildings is used in most propagation prediction models for macro – cells. Scattering or reflection from large buildings, hills, mountains are modeled to improve the prediction quality and especially to characterize the time dispersion of the radio channel.

In micro – cells

Most models rely on specular reflection and diffraction phenomena. Some empirical formulations use guided wave (Telekom model and Uni – Karlsruhe 2D – URBAN – PICO model) or virtual source at intersections that can be viewed as a way to model the combined effects of diffraction and scattering (Ericsson model). Scattering effects from walls and trees as well as from individual scatterer such as balcony, lamppost, windows, cars, etc. remain to be carefully examined. Contributions from over rooftop propagation are usually modeled using models similar to the ones for macro – cells.

Indoor

Mainly reflection from and transmission through walls, partitions windows, floors and ceilings are used to predict propagation within buildings. Wave guiding in corridors or in hallways are more difficult to model and thus are usually not considered. Although diffraction

effects have been sometimes identified, diffraction at edges from walls or windows is usually not taken into account due to the difficulties related to the requirement on the input database and due to the resulting large computation time.

4.2. Propagation in free space

The available power at a receiver antenna that is separated from a radiating transmitter antenna by a distance d is given by Friis free space equation:

$$P_r = \frac{P_t G_t G_r \lambda^2}{(4\pi)^2 d^2 L} \quad (4.1)$$

where

P_t = transmitted power,

P_r = received power,

G_t = transmitter antenna gain,

G_r = receiver antenna gain,

L = system loss factor not related to propagation ($L \geq 1$)

λ = wavelength in meters.

The losses L are usually due to the transmission line attenuation, filter losses and antenna losses in the communication systems. A value of $L = 1$ indicates no loss in the system hardware.

The propagation loss (or path loss), which represents signal attenuation as a positive quantity measured in dB, is defined as the difference (in dB) between the effective transmitted power and the received power.

$$L[\text{dB}] = 10 \cdot \log \frac{P_t}{P_r} = -10 \log G_t - 10 \log G_r + 20 \log f - 20 \log d - k \quad (4.2)$$

where

$$k = 20 \cdot \log \frac{c}{4\pi} = 147.6 \quad (4.3)$$

It is often useful to compare path loss with the basic path loss between isotropic antennas, that is:

$$L_B[\text{dB}] = -32.44 - 20 \log f - 20 \log d \quad (4.4)$$

Equation (4.1) shows that the received power obeys an inverse square with range d , so that it falls by 6 dB when the range is doubled (or reduces by 20 dB/decade). The Friis free

space model is only valid for values of d that are in the far field of transmitting antenna. The far-field (or Fraunhofer region) of a transmitting antenna is defined as the region beyond the distance d_f that is related to the largest linear dimension of the transmitting antenna aperture and the wavelength by:

$$d_f = \frac{2D^2}{\lambda} \quad (4.5)$$

where d is the largest physical dimension of the antenna. Additionally, to be in the far –field region, d_f must satisfy:

$$d_f \gg D \text{ and } d_f \gg \lambda \quad (4.6)$$

4.3. Reflection

Reflection occurs when a propagating electromagnetic wave impinges upon an object that has very large dimensions when compared to the wavelength of the propagating wave. Reflections occur from the surface of the earth and from buildings and walls.

4.3.1. The Fresnel reflection coefficients

Figure 4.1 shows the case of propagation over a plane earth, the distance between the two antennas being small enough to neglect curvature and to assume the reflecting surface to be flat. In Figure 4.1 the subscripts i and r refer to the incident and reflected E-fields, respectively. The amplitude and phase of ground-reflected wave depend on the reflection coefficient of the earth at the point of reflection and differs from vertical and horizontal polarization. In practice, the earth is neither a perfect conductor nor a perfect dielectric so the reflection coefficient depends on the ground constants, in particular and dielectric constant ϵ_c and the conductivity σ . The most common expression for the reflection is the Fresnel reflection coefficient that is valid for an infinite boundary between two media:

$$R_h = \frac{\sin \theta - (\epsilon_c - \cos^2 \theta)^{1/2}}{\sin \theta + (\epsilon_c - \cos^2 \theta)^{1/2}} \quad (\text{horizontal incidence}) \quad (4.7a)$$

$$R_v = \frac{\epsilon_c \sin \theta - (\epsilon_c - \cos^2 \theta)^{1/2}}{\epsilon_c \sin \theta + (\epsilon_c - \cos^2 \theta)^{1/2}} \quad (\text{vertical incidence}) \quad (4.7b)$$

where R_h and R_v are the complex reflection coefficients, θ is the angle of incidence, λ is the wavelength of the incident wave field and ϵ_c is the relative dielectric constant $\epsilon_c = \epsilon_r - j60\sigma\lambda$, ϵ_r is the relative permittivity and σ is the special conductivity of the reflecting surface.

The reflection coefficients are complex and the reflected wave will therefore differ in both magnitude and phase from the incident wave. For horizontal polarization the relative phase of the incident and reflected waves is nearly 180° for all angles of incidence. For very small values of c (near-grazing incidence) equation (4.7a) shows that the reflected wave is equal in amplitude and 180° out of phase with the incident wave for all frequencies and all ground conductivities. As the angle of incidence is increased, the amplitude and the phase of the reflected wave change, but only by relative small amounts. This change is greatest at higher frequencies and when the ground conductivity is poor. At grazing incidence there is no difference between horizontal and vertical polarization.

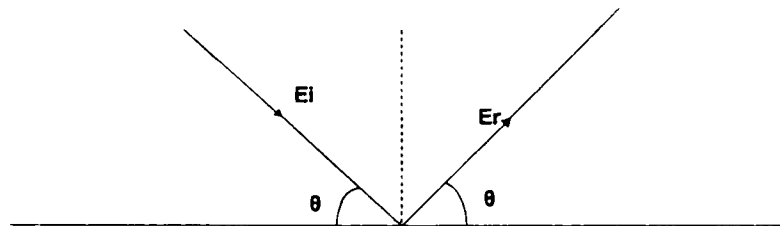


Fig. 4.1. Propagation over a plane earth

The Brewster angle is the angle at which no reflection occurs in the medium of the origin. It occurs when the incident angle θ_B is such that the reflection coefficient R_v is equal to zero. For the case when the first medium is the free space and the second medium has a relative permittivity ϵ_r the Brewster angle is given by the value of θ_B that satisfies:

$$\sin(\theta_B) = \sqrt{\frac{\epsilon_r - 1}{\epsilon_r^2 - 1}} \tag{4.8}$$

Note that the Brewster angle occurs only for vertical polarization.

4.3.2. Ground reflection (2-ray) model

For distances less than a few tens of kilometres, it is permissible to neglect the earth curvature and assume the surface to be smooth and flat. The 2-ray ground reflection model

represented in Fig. 4.2 is based on geometric optics and considers both the direct path and a ground reflected path between transmitter and receiver.

This model has been found to be reasonable accurate for predicting the large scale signal strength over distances of several kilometres for mobile radio systems that use tall towers (heights which exceed 50 m), as well as for LOS micro-cell channels in urban environment.

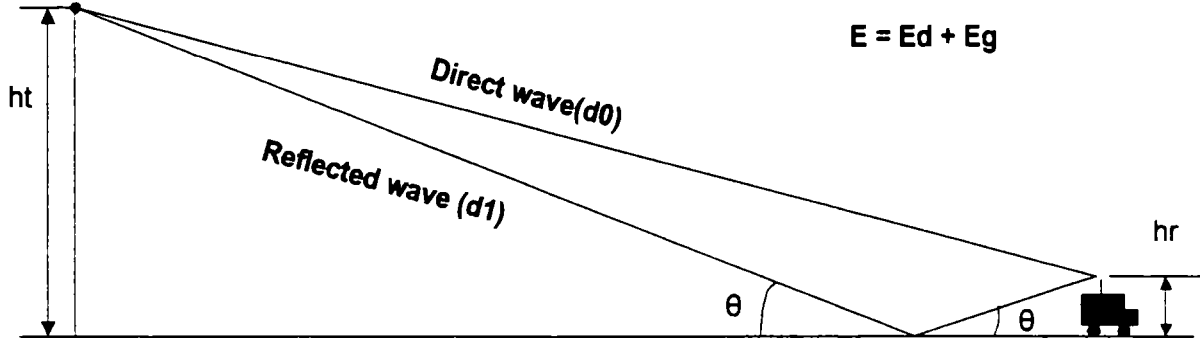


Figure 4.2. Two-ray ground reflection model

The total received E-field strength, E , is then a result of the direct line-of-sight component, E_d , and the ground reflected component, E_g :

$$E = E_d [1 - R_v \exp(-j\Delta\phi)] \quad (4.9)$$

where R_v represents the reflection coefficient and $\Delta\phi$ is the phase difference between the two rays. This phase difference can be expressed as

$$\Delta\phi = \beta \cdot \Delta d = \frac{2\pi}{\lambda} \Delta d \quad (4.10)$$

where β is the wave number ($\beta=2\pi/\lambda$) and Δd is the difference between the two radio paths ($\Delta d=d_1-d_0$).

Since in the mobile radio environment R_v is always approximately equal to -1 and $\Delta\phi$ is much less than one radian, the received power can be expressed [Lee, 93]

$$P_r = P_t G_t G_r \left(\frac{\lambda}{4\pi d} \right)^2 (\Delta\phi)^2 \quad (4.11)$$

where, in case of $d \gg h_t + h_r$

$$\Delta\phi \cong \frac{4\pi h_1 h_2}{\lambda d} \quad (4.12)$$

The plane earth propagation equation becomes:

$$P_r = P_t G_t G_r \left(\frac{h_t h_r}{d^2} \right)^2 \quad (4.13)$$

This equation is an imperfect formula since it does not take into account the wavelength and it shows an inverse fourth-power law with range rather than the inverse square law of the free space formula. This means a more rapid decrease in received power with range, 12 dB for each doubling of distance in this case. This equation applies only ranges where the assumption $d \gg h_t + h_r$ is valid.

4.4. Diffraction over irregular terrain

Diffraction occurs when the radio path between the transmitter and receiver is obstructed by a surface that has sharp irregularities (edges). The secondary waves resulting from the obstructing surface are present throughout the space and even behind the obstacles, giving rise to a bending of waves around the obstacle, even when a line-of-sight path exists between transmitter and receiver. At high frequencies, diffraction, like reflection, depends on the geometry of the object, as well as the amplitude, phase and polarization of the incident wave at the point of diffraction. The phenomena of diffraction can be explained by Huygens's principle. Huygens assumed that every point on a wave front might be regarded as a source of spherical wavelets the envelope of which is the position of the wave front at a later time. Huygens was thus able to account for rectilinear propagation and for the laws of reflection and refraction. Fresnel added the hypothesis that the wavelets can interfere, and this led to a theory of diffraction. The field strength of a diffracted wave in the shadowed region is the vector sum of the electric field components of all secondary wavelets in the space around the obstacle.

Consider a transmitter and a receiver separated in free space as shown in Figure 4.3. Also consider an obstructing screen of height h with infinite width placed between them at a distance d_1 from the transmitter and a distance d_2 from the receiver. The ways propagating from the transmitter to the receiver via the top of the screen traverse a longer path than if a direct line-of-sight path (through the screen) exists.

In terms of the geometry of Fig. 4.3 and Fig. 4.4 the "excess path length" is given by

$$\Delta \approx \frac{h^2}{2} \frac{d_1 + d_2}{d_1 d_2} \quad (4.14)$$

assuming $h \ll d_1, d_2$ and $h \gg \lambda$. The corresponding phase difference is

$$\phi = \frac{2\pi\Delta}{\lambda} \approx \frac{2\pi h^2}{\lambda} \frac{d_1 + d_2}{2 d_1 d_2} \quad (4.15)$$

The dimensionless Fresnel-Kirchoff diffraction parameter v that is given by:

$$v = h \sqrt{\frac{2(d_1 + d_2)}{\lambda d_1 d_2}} \quad (4.16)$$

and from equation the phase difference can be expressed as

$$\phi = \frac{\pi}{2} v^2 \quad (4.17)$$

Alternatively, using the same approximations, it can be written:

$$\phi = \frac{\pi \alpha^2}{\lambda} \frac{d_1 d_2}{d_1 + d_2}$$

and

$$v = \alpha \sqrt{\frac{2 d_1 d_2}{\lambda (d_1 + d_2)}}$$

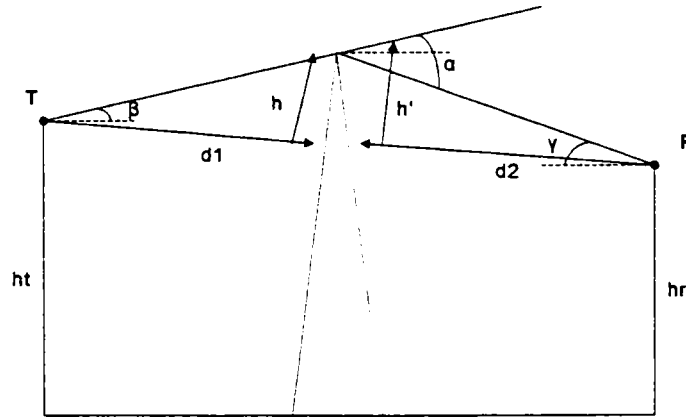


Figure 4.3. Knife-edge diffraction geometry when the transmitter and the receiver are not at the same height.

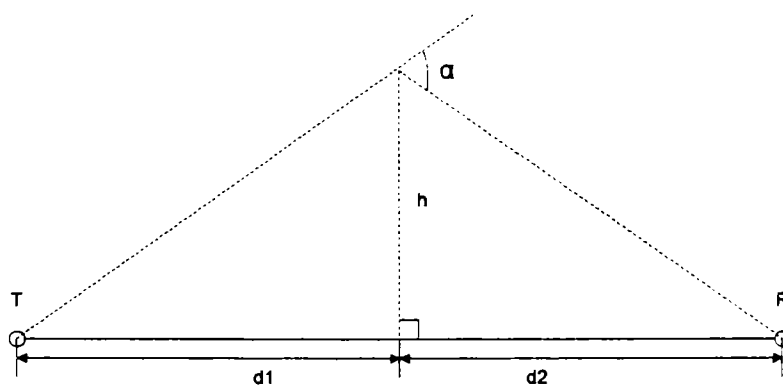


Figure 4.4. The knife-edge diffraction equivalent geometry. The point T denotes the transmitter and R denotes the receiver, with an infinite knife-edge obstruction blocking the line-of-sight path.

In practice, in order to obtain transmission under free space conditions, usually involves raising the antenna height until the necessary clearance over terrain obstacles is

obtained. If the terminals of a radio link path for which line-of-sight clearance over obstacles exists are low enough for the direct path to pass close to the surface of the earth at some intermediate point, then there may well be a path loss considerably in excess of the free space loss even though the path is not actually blocked. It is needed a quantitative measure of the required clearance over any terrain obstruction and this may be obtained in terms of Fresnel–zone ellipsoids drawn around the path terminals.

4.4.1. Fresnel zone geometry

The Fresnel zones explain the concept of diffraction loss as a function of the path difference around an obstruction. The Fresnel zones represent successive regions where the secondary waves have a path length from the transmitter to the receiver which are $n\lambda/2$ greater than the total path length of a line-of-sight path. In Fig. 4.5 is represented a family of ellipsoids defining the first three Fresnel zones around the terminals of a radio path.

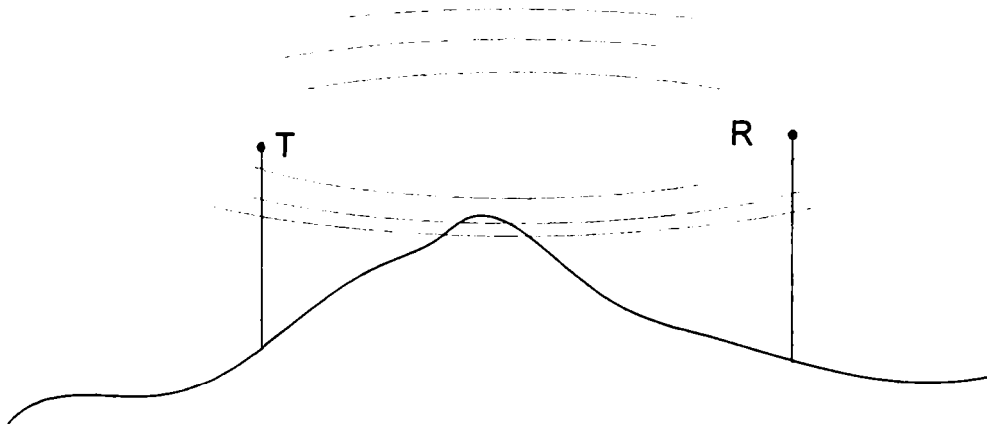


Figure 4.5. Family of ellipsoids defining the first three Fresnel zones around the transmitter and the receiver of a radio path

The successive Fresnel zones have the effect of alternatively providing constructive and destructive interference to the total received signal. The radius of the n^{th} Fresnel zone circle can be expressed as

$$r_n = \sqrt{\frac{n\lambda d_1 d_2}{d_1 + d_2}} \quad (4.18)$$

This is an approximation that is valid for $d_1, d_2 \gg r_n$ and is, therefore, realistic except in the immediate vicinity of the terminals. The volume enclosed by the ellipsoid defined by $n = 1$ is known as the first Fresnel zone, the volume between this one and the ellipsoid defined by $n = 2$ is the second Fresnel zone, etc.

Contributions to the field at the receiving point from successive Fresnel zones tend to be in phase opposition and therefore interfere destructively rather than constructively. If an obstructed screen was actually placed at a point between transmitter and receiver, then if the radius of the aperture was increased from that corresponding to the first Fresnel zone to that defining the limit of the second Fresnel zone, then to the third Fresnel zone, etc., then the field at the receiver would oscillate. The amplitude of the oscillations would gradually decrease since smaller amounts of energy propagate via the outer zone.

In mobile communication systems, diffraction losses occur from the blockage of the secondary waves such that only a portion of the energy is diffracted around an obstacle. That is, an obstruction causes a blockage of energy from some of the Fresnel zones, thus allowing only some of the transmitted energy to reach the receiver. Depending of the geometry of the obstruction, the received energy will be a vector sum of the energy contributions from all unobstructed Fresnel zones.

4.4.2. Diffraction losses

Estimating the signal attenuation caused by diffraction of radio waves over hills and buildings is essential in predicting the field strength in a given service area. Generally, it is impossible to make very precise estimates of the diffraction losses, and in practice prediction is a process of theoretical approximation modified by necessary empirical corrections. Though the calculation of diffraction losses over complex and irregular terrain is a mathematically difficult problem, expressions for diffraction losses for many simple cases have been derived. As a starting point, the limiting case of propagation over a knife-edge gives good insight into the order of magnitude of diffraction loss.

If an ideal, straight, perfectly absorbing screen is interposed between T and R it will have little effect when the top of the screen is well below the line-of-sight path. The field at the receiver will be then the free space value E_0 . It will begin to oscillate as the height of the screen is increased, blocking more of the Fresnel zones below the LOS path. The amplitude of the oscillations increase until the obstruction edge is just in line with T and R at which point the field strength is exactly half the unobstructed value i.e. the loss is 6 dB. As the height is increased above this value the oscillations cease and the field strength decrease steadily.

In order to express this in a quantitative way, any obstruction between transmitter and receiver is replaced by an absorbing plane placed at the same position and the classical

diffraction theory is used. The plane is normal to the direct path and extends to infinity in all directions except vertically where it stops at the height of the original obstruction.

Considering a receiver point R, located in the shadow region (also called the diffraction zone) (Figure 4.3), the field strength at receiver is determined as the sum of all secondary Huygens sources in the plane above the obstruction and can be expressed as

$$\frac{E}{E_0} = F(v) = \frac{1+j}{2} \int_v^{\infty} \exp\left(-j\frac{\pi}{2}t^2\right) dt \quad (4.19)$$

where $F(v)$ is known as the complex Fresnel integral with v being the value given by equation (4.16) for the height of the obstruction under consideration. If the obstruction lies below the line-of-sight then h and v are negative. If the path is actually obstructed then h and v are positive as depicted in Figure 4.6.

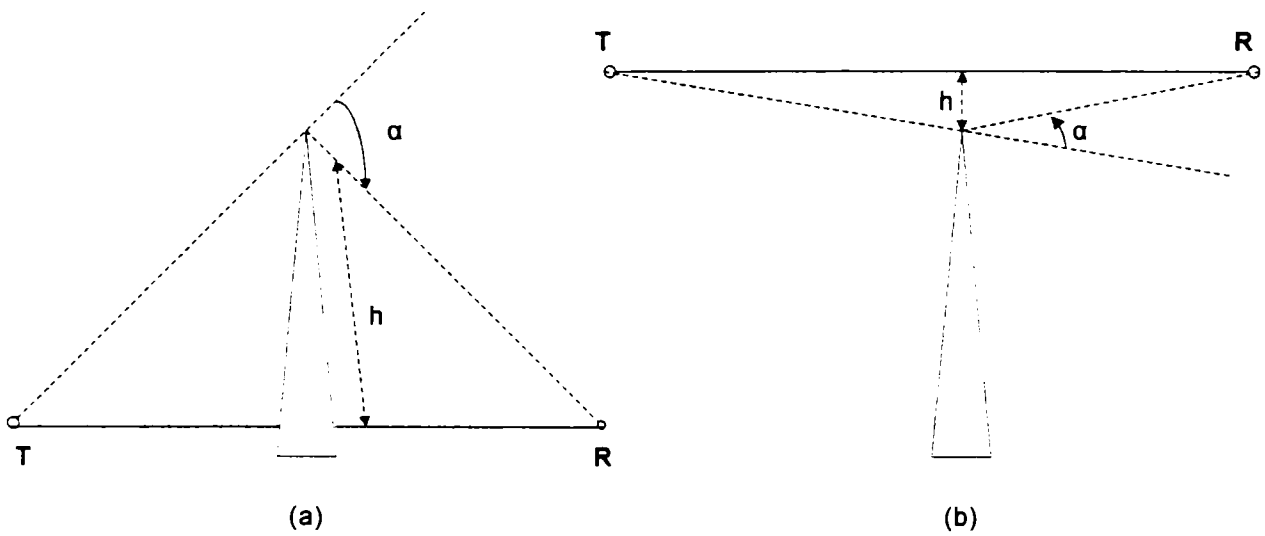


Figure 4.6. Knife – edge diffraction: (a) h and v positive, (b) h and v negative

In practice, the Fresnel integral is commonly evaluated using tables or graphs for given values of v . The diffraction gain due to the presence of a knife-edge as compared to the free space E_0 , is given by

$$G_d[\text{dB}] = 20 \log|F(v)| \quad (4.20)$$

In practice, graphical and numerical solutions are relied upon to compute diffraction gain. The approximate solution provided by Lee [Lee, 93] is expressed as follows:

$$L(v) = 0 \quad v \leq -1 \quad (4.21.a)$$

$$L(v) = 20 \log(0.5 - 0.62v) \quad -1 \leq v \leq 0 \quad (4.21.b)$$

$$L(v) = 20 \log(0.5 \exp(-0.95v)) \quad 0 \leq v \leq 1 \quad (4.21.c)$$

$$L(v) = 20 \log \left(0.4 - \sqrt{0.1184 - (0.38 - 0.1v)^2} \right) \quad 1 \leq v \leq 2.4 \quad (4.21.d)$$

$$L(v) = 20 \log \left(\frac{0.225}{v} \right) \quad v > 2.4 \quad (4.21.e)$$

Ground reflections

The previous analysis has ignored the possibility of ground reflections either side of the terrain obstacle. This possibility is expressed in figure 4.7 and then, four paths have to be taken into account in computing the field at the receiver point [Anderson, 58]. It should be noted that the four rays depicted in Figure 4.7 travel different distances and therefore have different phases at the receiver. In addition, the Fresnel v parameter is different in each case so the field at the receiver has to be computed from:

$$E = E_0 \sum_{k=1}^4 L(v_k) \exp(j\phi_k) \quad (4.22)$$

In any particular situation a ground reflection may exist only on the transmitter or the receiver side of the obstacle in which case only three rays exist.

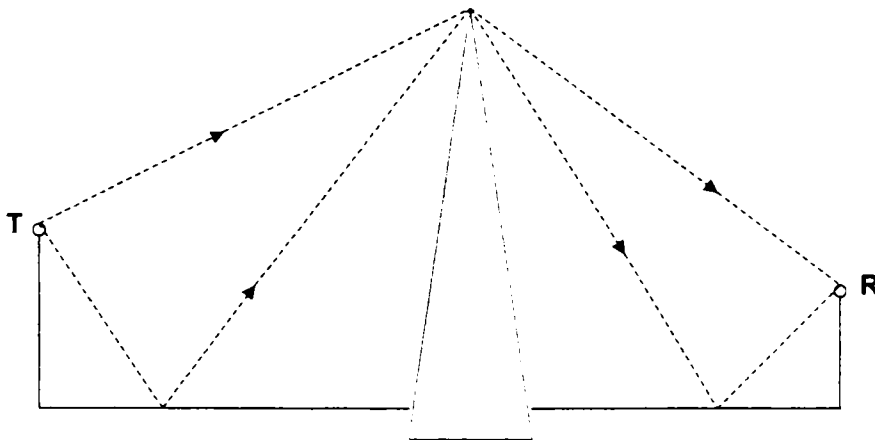


Figure 4.7. Knife – edge diffraction with ground reflections

Diffraction over rounded obstacles

Objects encountered in the physical world often have dimensions that are large compared with the wavelength of transmission. In [Hacking, 68] it has been showed that the loss due to rounded obstacles exceeds the knife-edge diffraction loss. If a rounded hilltop as depicted in Figure 4.8 is replaced by a cylinder of radius r equal to that of the crest, then the cylinder supports reflections either side of the hypothetical knife-edge that coincides with the peak, and the Huygens wave front above that point is therefore modified. The mechanism is similar to that in the “four – ray” situation described above. An excess loss can be added to

the knife – edge loss to account for this, the value being given by [Hacking, 68]:

$$L_{ex}[\text{dB}] = 11.7 \left(\frac{\pi r}{\lambda} \right)^2 \alpha \quad (4.23)$$

If the hilltop is rough, due of the presence of trees, then the diffraction loss is about 65% of the value given above.

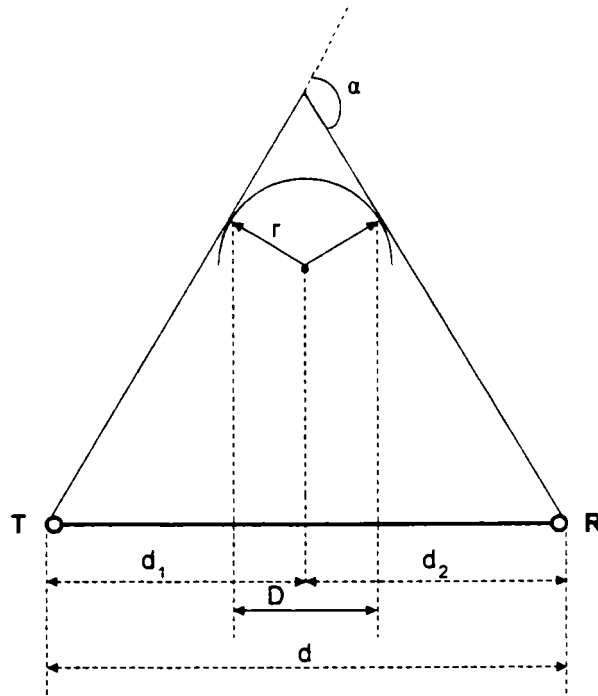


Figure 4.8. Diffraction over a cylinder

An alternative solution given in [Dougherty, 64] is available through a dimensionless parameter ρ defined as:

$$\rho = \left(\frac{\lambda}{\pi} \right)^{1/6} r^{1/3} \left(\frac{d_1 + d_2}{d_1 d_2} \right)^{1/2} \quad (4.24)$$

The diffraction loss can then be represented by the quantity $A(v, \rho)$, normally expressed in dB. It is related to the ideal knife – edge loss $A(v, \rho)$ by

$$A(v, \rho) = A(v, 0) + A(0, \rho) + U(v\rho) \quad (4.25)$$

where $U(v\rho)$ is a correction factor. Approximations are available for $A(0, \rho)$ and $U(v\rho)$ [Causebrook, 71] as:

$$A(0, \rho) = 6 + 7.19\rho - 2.02\rho^2 + 3.63\rho^3 - 0.75\rho^4 \quad \rho < 1.4 \quad (4.26)$$

$$U(v\rho) = \begin{cases} (43.6 + 23.5v\rho) \log(1 + v\rho) - 6 - 6.7v\rho & v\rho < 2 \\ 22v\rho - 20 \log(v\rho) - 14.13 & v\rho \geq 2 \end{cases} \quad (4.27)$$

Both the above methods are strictly applicable only to horizontally - polarized signals, but measurements [Hacking, 68] have shown that at VHF and UHF they can be applied to vertical polarization with reasonable accuracy.

The radius of a hill crest (Figure 4.8) may be estimated as:

$$r = \frac{2Dd_1d_2}{\alpha(d_1^2 + d_2^2)} \tag{4.28}$$

4.5. Scattering

Rough surfaces and finite surfaces scatter the incident wave in all directions with a radiation diagram that depends on the roughness and size of the surface or volume. The dispersion of energy through scattering means a decrease of the energy reflected in the direction of the receiving antenna so the ground reflected wave might therefore make a negligible contribution to the received signal.

A measure of the surface roughness is given by the Rayleigh criterion that is set at

$$\Delta\phi = \frac{\pi}{2} \tag{4.29}$$

where $\Delta\phi$ is the phase difference between the two rays shown in Fig. 4.9.

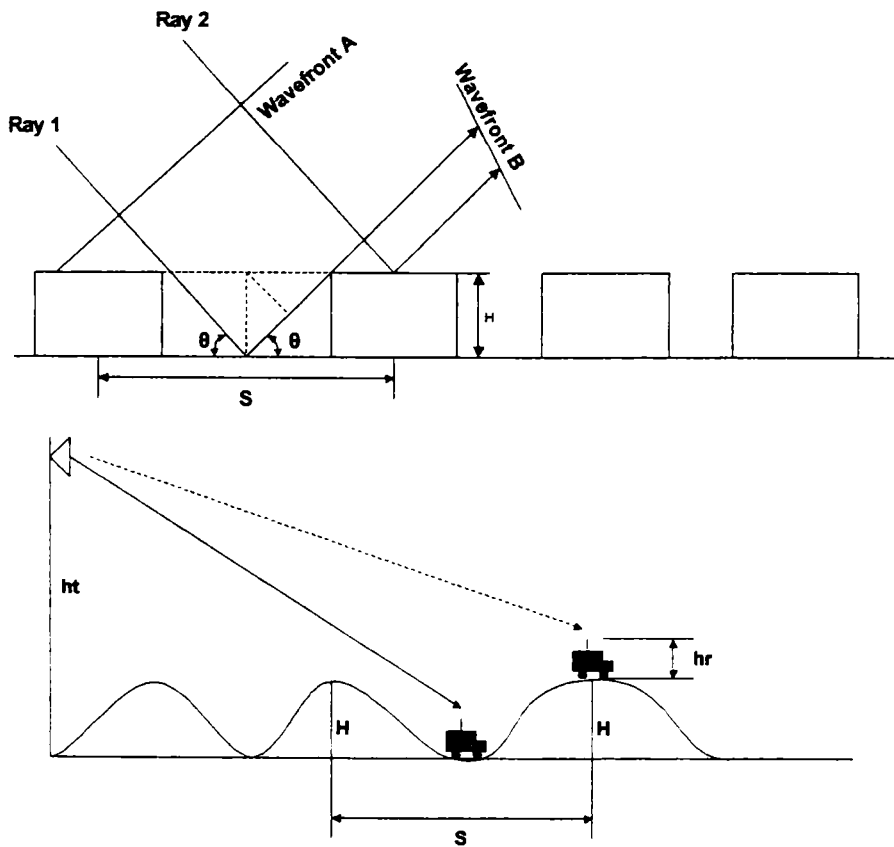


Fig. 4.9. Surface roughness criterion

For a small angle of incident, the Rayleigh height H_R and the minimum spacing S_R between protuberances is expressed as:

$$H_R = \frac{\lambda}{8\theta}; \quad S_R = \frac{\lambda}{4\theta^2} \quad (4.30)$$

According to the Rayleigh criteria, if the undulation surface height is greater than H_R under the constraint that the spacing between two noticeable humps is greater than S_R , then it is a rough terrain. In a mobile radio environment we would assume different criteria for different propagation path lengths. The reason is that the mobile antenna is usually close to the ground so that the reception of both the direct ray and specular reflected wave is weak. In this case, even if the phase difference $\Delta\phi$ between direct and specular reflected wave is about $\pi/4$ at a path distance larger than 9.7 km, unexpected reflected waves often received and will further weaken the resultant.

Therefore, the following criteria are suggested for different propagation path lengths [Rappaport, 96]:

$$\begin{aligned} \Delta\phi = \frac{\pi}{2} \text{ (less than 3.2km),} \quad H_R = \frac{\lambda}{8\theta}; \quad S_R = \frac{\lambda}{4\theta^2} \\ \Delta\phi = \frac{\pi}{3} \text{ (from 3.2 to 9.6 km)} \quad H_R = \frac{\lambda}{12\theta}; \quad S_R = \frac{\lambda}{6\theta^2} \\ \Delta\phi = \frac{\pi}{4} \text{ (9.6 km and up)} \quad H_R = \frac{\lambda}{16\theta}; \quad S_R = \frac{\lambda}{8\theta^2} \end{aligned} \quad (4.31)$$

The parameters H_R and S_R are functions of the incident angle θ that can be computed from the following expression:

$$\theta = \frac{H + h_t + h_r}{d} \quad (4.32)$$

where h_t and h_r are the antenna heights of the base and mobile unit, respectively, and H is the difference in elevation between the high and low spots around the mobile unit.

The roughness is determined by the frequency, the incident angle and the terrain irregularity heights and spacing as it is expressed in equation (4.31). Therefore, at one frequency a surface can be considered rough but not at another frequency. This statement also applies to different incident angles.

In the mobile radio environment we use the following criteria to determine the terrain roughness. If H is the difference of two adjacent extremes in elevation and $H > H_R$ in the vicinity of $\Delta x = 1/2 S_R$ at the mobile site, then the terrain is rough. Based on the location of the mobile unit, which is on the hump or in the valley, the incident angle θ is within

$$\frac{h_t + h_r}{d} \leq \theta \leq \frac{H + h_t + h_r}{d} \quad (4.33)$$

Usually, the base station antenna height is greater than the mobile antenna height so the reflection point is closer to the mobile unit. The distance $\Delta x = 1/2S_R$ from the mobile unit towards the base station only needs to be searched to find out if any variation of elevation is greater than H_R . If one variation of elevation H within Δx is greater than H_R , the terrain is called a rough terrain.

4.6. Propagation mechanism in ray theory

The main propagation mechanisms defined by the ray theory are on brief explained below. A radio ray is assumed to propagate along a straight line bent only by refraction, reflection, diffraction or scattering. These are the concepts of Geometrical Optics. There is no transversal dimension associated to the rays. However, the finite size of the wavelength at radio frequencies leads to hinder in some ways the assumption of infinitely thin rays. Related to the “thickness” of a radio ray is the concept of Fresnel zones.

Specular reflection

The specular reflection phenomenon is the mechanism by which a ray is reflected at an angle equal to the incident angle. The reflected wave fields are related to the incident wave fields through the Fresnel reflection coefficient. The application of the Fresnel coefficient formulas is very popular in ray tracing software tools. Specular reflection are mainly used to model reflection from the ground surface and from building walls. The mechanism of specular reflections have been used to interpret measurements in some particular environments such as high rise city center, micro – cells, indoor and down in a street canyon illuminated from over the roof. Whether scattering ($1/d_1d_2$ dependence) or truly specular reflection ($1/(d_1+d_2)$ dependence) is the proper propagation phenomena was not mentioned and cannot be readily determined since the two phenomena are usually involved simultaneously. The reflection from a finite surface can be seen either as the sum of the two phenomena specular reflection and edge diffraction, or as a scattering process.

Diffraction

The diffraction process in ray theory is the propagation phenomenon that explains the transition from the lit region to the shadow regions behind the corner of a building or over the rooftops. Diffraction by a single wedge can be solved in various ways: empirical formulas [Kreuzgruber, 94], [Jakoby, 95]; Perfectly Absorbent Wedge (PAW) [Wagen, 93]; Geometrical Theory of Diffraction (GTD) [Wagen, 93], [Keller, 62]; Uniform Theory of Diffraction (UTD) [McNamara, 90] or even more exact formulation. The advantages and disadvantages of using either one formulation is difficult to address since it may not be independent on the environments under investigations. Indeed, reasonable results are claimed for each formulation. The various expressions differ mainly from the approximations being made on the surface boundaries of the wedge considered. One major difficulty is to express and use the proper boundaries in the derivation of the diffraction formulas. Another problem is the existence of wedges in real environments: the complexity of a real building corner or of the building roofs clearly illustrates the modeling difficulties. Despite these difficulties, diffraction around a corner or over a rooftop are commonly modeled using heuristic UTD formulas since they are fairly easy to program, are well behaved in the lit/shadow transition region, and account for the polarization and the wedge material.

Multiple diffraction

For the case of multiple diffractions the complexity increases dramatically. In the case of propagation over rooftop the results of Walfisch and Bertoni [Walfisch, 88] have been used to produce the COST – Walfisch – Ikegami model (chapter 5). The approximate procedure of Giovanelli or Deygout have been revisited by some authors. The limitations of these approximations lead several researchers to more accurate methods that are numerical schemes to compute the multiple diffractions and apart from the last contribution they do not give a clear physical understanding of the multiple diffraction processes, at least not yet [COST231, 99].

One method frequently applied to multiple diffraction problems is UTD. The main problem with straightforward applications of the UTD is that in many cases one edge is in the transition zones of the previous edges. Strictly speaking this forbids the application of ray techniques but in the spirit of UTD the principle of local illumination of an edge should be valid. At least within some approximate degrees, a solution can be obtained. In [Andersen, 96] a solution is shown that is quite accurate in most cases of practical interest. The key point in the theory is to include slope diffraction that is usually neglected as a higher order term in

an asymptotic expansion, but in transition zone diffraction the term is of the same order as the ordinary amplitude diffraction term [Andersen, 94]. Another key element in the method is automatic enforcement of continuity of amplitude and slope at each point.

For the case of diffraction over multiple screens of arbitrary heights and spacing a solution is obtained within the frame of UTD. This solution agrees to a good approximation within the known results for constant spacing and numerical results using Vogler's solution [Vogler, 82]. The limitation of the method is that it is not applicable when one spacing becomes very small relative to other spacings. Thus the method cannot predict the collapse of two screens into one.

In ITU – R 526 – 2 [ITU, 94] equations are given to compute effects of multiple diffractions around curved cylinders. In [Li, 96] an investigation of this method revealed that a modification of the ITU equations yields good results even for multiple knife – edge diffractions. The diffraction losses for the single obstacles are replaced by the diffraction losses for single knife – edges; furthermore, a modified correction factor has to be used.

The ITU equations are simply to apply and cannot account for knife – edges with unequal heights and separations. In the special case of grazing incidence over a series of equal distance and equal height knife – edges, the modified ITU equations yield the correct analytical results. This holds true even for large number of edges.

Scattering

Rough surfaces and finite surfaces scatter the incident energy in all directions with a radiation diagram that depends on the roughness and size of the surface or volume. The dispersion of energy through scattering means a decrease of the energy reflected in the specular direction. This simple view leads to account for the scattering process only by decreasing the reflection coefficient and thus, only by multiplying the reflection coefficient with a factor smaller than one which depends exponentially on the standard deviation of the surface roughness according to the Rayleigh theory [COST231, 99]. This description does not take into account the true dispersion of radio energy in various directions but accounts for the reduction of energy in the specular direction due to diffuse components scattered in all other directions.

More realistic scattering processes have been investigated within COST 231. Most investigations deals with the application of the bistatic radar equation to account for the scattering from hills or mountain slopes. A preliminary study investigated the scattering pattern from large irregularities on a building wall [Rizk, 94]. The concept promoted in that

study was to model the scattering by equivalent sources of scattering located at the buildings corners.

Penetration and absorption

Penetration loss due to building walls have been investigated and found very dependent on the particular situation. Absorption due to trees or the body absorption are also propagation mechanisms difficult to quantify with precision.

Another absorption mechanism is the one due to atmospheric effects. These effects are usually neglected in propagation models for mobile communication applications at radio frequencies but are important when higher frequencies (e.g. 60 GHz) are used.

Atmospheric effects

Atmospheric effects are usually not taken into account for mobile radio applications at UHF frequencies, although empirical correction factors can be incorporated in some coverage prediction tools to handle seasonal variations.

Guided wave

Wave guiding can be viewed as a particular propagation mechanism to describe the propagation in street canyons (Telekom model), in corridors or tunnels. The wave guiding phenomena can be explained based on multiple reflections or propagation models.

5. Propagation Prediction Models

5.1. General considerations

Propagation prediction models are required in order to compare classical propagation prediction methods with the ones obtained by neural network applications. The phenomena that influence radio wave propagation can generally be described by four basic mechanisms: reflection, penetration, diffraction and scattering, described in chapter 4. For the practical prediction of propagation in a real environment these mechanisms must be described by approximations. This requires a three-stage modeling process. In the first step the real (analogue) terrain has to be digitized yielding digital terrain database. The information includes terrain height information, land usage data, building shape and height information and building surface characteristics. The second modeling step includes the definition of mathematical approximations for the physical propagation mechanisms. Based on the solutions for the basic problems both deterministic and empirical approaches have been developed for the various environments, which is the third modeling step. In the different environments distinctions of the models are required both in terms of the dominant physical phenomena and the specification of the digital terrain data. In section 5.2 are treated models dedicated for macro-cells. As the definition of cell type is not unique in the literature, the cell type definition used in this work is explained more detailed in [COST231, 99]. The summary of the different cell types is shown in Table 5.1. [COST231, 99].

Table 5.1. Definition of cell type

Cell type	Typical cell radius	Typical position of base station antenna
Macro-cell (large cell)	1 km to 30 km	Outdoor; mounted above medium rooftop level, heights of all surrounding are below base station antenna height
Small macro-cell	0.5 km to 3 km	Outdoor; mounted above medium rooftop level, heights of some surrounding buildings are above base station antenna height
Micro-cell	Up to 1 km	Outdoor; mounted below medium rooftop
Pico-cell/in house	Up to 0.5 km	Indoor or outdoor (mounted below rooftop level)

In large cells and small cells the base station antenna is installed above rooftops. In this case the path loss is determined mainly by diffraction and scattering at rooftops in the

vicinity of the mobile, i.e. the main rays propagate above the rooftops. In micro-cells, the base station antennas are mounted generally below rooftops. Wave propagation is determined by diffraction and scattering around buildings, i.e. the main rays propagate in street canyons somehow like in grooved waveguides. Pico-cells are applied to cover mainly indoor or very small outdoor areas. In any case the base station antenna of a pico-cell is mounted inside a building or fairly below rooftop level in outdoors.

5.2. Propagation models for macro-cells

Macro-cell design philosophy is based on the assumption of high radiation centerlines, usually placed above the surroundings, transmitter powers of several tens of Watts and large cells (Table 5.1). Under these assumptions, LOS conditions are usually not satisfied and the signal from the transmitter to the receiver propagates by means of the diffraction and the reflection. Also, for large cells the effects of refraction are very important. All of these factors make the problem of field strength prediction very difficult. In the next sections, only a few very popular methods are presented.

Path loss models for urban areas often comprise two components that correspond to the dominant mechanism of propagation: an expression relevant for the propagation in the vertical plane (over rooftops) and another one for the horizontal plane (along street canyons). The former will be more dominant far away from the base station, whereas the latter can be expected to play the dominant role in the vicinity of the base station. There are, of course, cases when both of them are important. Much depends on the base station height relative to the surrounding buildings. Most of the models of rooftop propagation are based on various approximate solutions to the problem of diffraction by multiple (infinitely thin) screens. They may be augmented by empirical correction factors derived from measurement data.

The propagation of radio waves in built-up areas has been found to be strongly influenced by the nature of the environment, in particular the size and density of buildings. Generally, a description of the environment is often employed using terms such as rural, urban and suburban.

Usually the propagation loss model is used to express the measured propagation loss as a function of the variables associated with the environment and terrain of the mobile, base station antenna height h_{BS} , frequency f , mobile antenna height h_M and the distance between transmitting and receiving antennas, d . The degree of terrain undulation is given by a

parameter known as intercede range Δh [Jakes, 74]. The value of Δh depends on the terrain topography. For instance, the value of Δh falls within an interval (20m, 40m) for quasismooth terrain or (40m, 80m) for rolling terrain. Thus, a simple mathematical expression for the propagation loss, L_p , in a specific type of built-up environment and size of city is represented by

$$L_p = f(h_{BS}, h_M, f, d, \Delta h) \quad (5.1)$$

5.2.1. The model of Okumura

Okumura's empirical model is derived from an extensive series of measurements made in a variety of environments in and around Tokyo, Japan, at frequencies up to 1920 MHz. There are published as a number of curves [Okumura, 68] giving the median attenuation relative to free space (A_{mu}) in an urban area over a quasi-smooth terrain with a base station effective antenna height (h_{BS_e}) of 200 m and a mobile antenna height (h_M) of 3 m. These curves were developed from extensive measurements using vertical omni-directional antennas at both the base and mobile and are expressed as a function of frequency (in the range 100-3000 MHz) and distance from the base station (in the range 1-100 km). In order to determine path loss using Okumura's method, the free space path loss between the points of interest is first determined, and then the value of $A_{mu}(f, d)$ (as read from the curves) is added to it along with correction factors to account for the type of terrain. The model can be expressed as follows:

$$L_{50}(\text{dB}) = L_{FS} + A_{mu}(f, d) - G(h_{BS_e}) - G(h_M) - G_{AREA} \quad (5.2)$$

where:

L_{50} = the 50th percentile (i.e. median) value of propagation path loss,

L_{FS} = the free space propagation loss,

A_{mu} = the median attenuation relative to free space,

$G(h_{BS_e})$ = the base station antenna height gain factor,

$G(h_M)$ = the mobile antenna height gain factor,

G_{AREA} = the gain due to the type of environment.

Okumura found that $G(h_{BS_e})$ varies at a rate of 20 dB/decade and $G(h_M)$ varies at a rate of 10 dB/decade for heights less than 3 m.

$$G(h_{BS_e}) = 20 \log \left(\frac{h_{BS_e}}{200} \right) \quad 30 \text{ m} < h_{BS} < 1 \text{ km} \quad (5.3)$$

$$G(h_M) = 10 \log\left(\frac{h_M}{3}\right) \quad h_M < 3 \text{ m} \quad (5.4)$$

$$G(h_M) = 20 \log\left(\frac{h_M}{3}\right) \quad 3 \text{ m} < h_M < 10 \text{ m} \quad (5.5)$$

Further corrections are also provided, in graphical form, to allow for street orientation, for transmission in suburban and open (rural) areas and over irregular terrain. These must be added or subtracted as appropriate. Irregular terrain is further divided into rolling hilly terrain, isolated mountain, general sloping terrain and mixed land – sea path. The terrain related parameters that have to be evaluated in order to determine the various correction factors are the terrain undulation height (Δh), the effective base station antenna height (h_{Bse}), isolated ridge height, average slope of the terrain and the mixed land-sea path parameter [Parsons, 92].

Okumura's model is based on measured data and does not provide any analytical explanation. For many situations, extrapolations of the derived curves can be made to obtain values outside the measurement range, although the validity of such extrapolations depends on the circumstances and the smoothness of the curve in question.

In practice, the Okumura technique produces predictions that correlate reasonably well with measurements, although by its nature it tends to average over some of the extreme situations and not respond sufficiently quickly to rapid changes in the radio path profile. Common standard deviations between predicted and measured path loss values are around 10 dB to 14 dB [Rappaport, 96]. Allsebrook found that an extended version of the Okumura technique produced prediction errors comparable to those of his own method. The comparison presented in [Delisle, 85] and [Aurand, 85] also showed the Okumura technique to be amongst the better models for accuracy although it was rated as "rather complex". Generally, the technique is quite good in urban and suburban areas, but not in rural areas over irregular terrain. There is a tendency for the predictions to be optimistic, i.e. suggesting a lower path loss than the actually measured [Parsons, 92].

A number of authors have fitted equations to these curves, the best known being Hata, 1980, and Reudink, 1974. Hata's equations include the effects of distance, frequency, mobile station antenna height (h_M) and base station antenna height (h_{BS}). The application of Hata's formulas is restricted to urban, suburban and open areas and to distances up to 20 km. However, measurements performed in (Low, 1986) have shown that Hata's formulas can also be used in forest and for distances up to 40 km [COST 207].

Akeyama's modification

The Okumura's technique adopts curves for urban areas as the datum from which other predictions are obtained. Cautions must be exercised in applying the environmental definitions as described by Okumura to locations in countries other than Japan. Okumura's definition for urban, for example, is based on the type and density of buildings in Tokyo and it may not be directly transferable to cities in North America or Europe. Indeed, experience with measurements in United States has shown that the typical US urban environment lies somewhere between Okumura's definition of suburban and open areas. Since the CCIR have adopted the Okumura urban curve as its basic model for 900 MHz propagation, it is also prudent to exercise caution when using these curves [Parsons, 92].

One other problem that is encountered in the use of Okumura model is that correction factor that accounts for environments other than urban (suburban, quasi-open and open) is a function only of the buildings in the immediate vicinity of the mobile. The factor is often more than 20 dB, is discrete and cannot be objectively related to the height and density of the buildings. It is uncertain how the factors suggested by Okumura can be applied to cities other than Tokyo, particularly those where the architectural style and construction materials are quite different. Some attempts have been made to extend the concept of "degree of urbanization". A ground cover (degree of urbanization) factor has been proposed by Akeyama [Akeyama, 82] to account for values α less than 50% in a continuous way. The value of S (the deviation from Okumura's reference median field strength due to buildings surrounding the mobile station curve at 450 MHz) is given by [Parsons, 92]:

$$S(\text{dB}) = \begin{cases} 30 - 25 \log \alpha & 5\% < \alpha < 50\% \\ 20 + 0.19 \log \alpha - 15.6 (\log \alpha)^2 & 1\% < \alpha \leq 5\% \\ 20 & \alpha < 1\% \end{cases} \quad (5.6)$$

where α is the percentage of the area covered by buildings.

5.2.2. Hata prediction model

Okumura [Okumura, 68] published an empirical path loss prediction model based on field measurements taken in Tokyo area. It provides an initial path loss estimate for the urban area with a quasi-smooth terrain ($\Delta h \cong 20\text{m}$). In addition, some correction factors must be used to adapt to the results in some other conditions, for example, the type of propagation environment and the size of city. However, Okumura's method cannot be easily automated,

because it involves various curves. In an attempt to make Okumura method easy to apply, Hata [Hata, 80] established empirical mathematical relationships to describe the graphical information given by Okumura. Hata's formulation is limited to certain ranges of input parameters and is applicable only over quasi-smooth terrain. The mathematical expression for path loss L_p (in dB) for an urban area is:

$$L_p[\text{dB}] = 69.55 + 26.16 \log f - 13.82 \log h_{BS} - a(h_M) + (44.9 - 6.55 \log h_{BS}) \log d \quad (5.7)$$

where, for medium or small size city:

$$a(h_M) = (1.1 \cdot \log f - 0.7) h_M - (1.56 \cdot \log f - 0.8) \quad (5.8)$$

for large city:

$$a(h_M) = 8.29 (\log 1.54 h_M)^2 - 1.1 \quad \text{for } f \leq 200 \text{ MHz} \quad (5.9)$$

$$a(h_M) = 3.2 (\log 11.75 h_M)^2 - 4.97 \quad \text{for } f \geq 400 \text{ MHz} \quad (5.10)$$

For suburban area:

$$L_{ps}[\text{dB}] = L_p(\text{urban}) - 2 \left[\log (f/28)^2 \right] - 5.4 \quad (5.11)$$

For open area:

$$L_{po}[\text{dB}] = L_p(\text{urban}) - 4.78 (\log f)^2 + 18.33 \log f - 40.94 \quad (5.12)$$

The model is restricted to the following range of parameters:

f: 150...1500 MHz

h_{BS} : 30...200 m

h_M : 1...10 m

d: 1...20 km

Although Hata's model does not have any of the path specific corrections that are available in Okumura's model, the above expressions have significant practical values. The prediction of the Hata model compare very closely with the original Okumura model, as long as d exceeds 1 km [Rappaport, 96]. This model is well suited for large cell mobile systems, but not personal communication systems (PCS) that have cells on the order of 1 km radius.

5.2.3. The Egli's model

One of the best-known statistical models for predicting propagation loss in the urban or rural environment is due to Egli [Egli, 57]. As in the case of Okumura-Hata, it does not include diffraction losses caused by propagation over irregular terrain; however, Okumura

implicitly takes into account the effect of buildings, which is not the case for Egli. An initial comparison between two models can then be made for open (rural) areas where both models neglect diffraction losses.

According to Egli, propagation losses are expressed as:

$$L = 20 \log f + 40 \log d - 20 \log h_{BS} + \begin{cases} 76.3 - 10 \log h_M, & h_M \leq 10 \text{ m} \\ 85.9 - 20 \log h_M, & h_M \geq 10 \text{ m} \end{cases} \quad (5.13)$$

where:

L = total propagation loss, in dB,

f = frequency, in MHz

d = distance between base and mobile station, in km,

h_{BS}, h_M = base and mobile station effective antenna height, in m.

A comparison between the models of Hata and Egli is made in [Delisle, 85] for open regions where both models are applicable and neglect diffraction losses due to irregular terrain. One point to note is that while Egli predicts the average signal strength will decrease with distance at the rate of 40 dB/decade, Hata uses a rate depending on the base station effective antenna height, i.e. $44.9 - 6.55 \log h_{BS}$; a 30 m height leads to a loss a 35.2 dB/decade and 70 m to a calculated loss of 32.8 dB/decade.

5.2.4. COST 231-Hata model

COST 231 [COST231, 99] has extended Hata's model to the frequency band 1500...2000 MHz by analyzing Okumura's propagation curve in the upper frequency band.

The mathematical expression of the path loss for suburban areas is:

$$L_p[\text{dB}] = 46.3 + 33.9 \log f - 13.82 \log h_{BS} - a(h_M) + (44.9 - 6.55 \log h_{BS}) \log d + C_m \quad (5.14)$$

where $a(h_M)$ is defined in equation (5.3) and

$$C_m = \begin{cases} 0 \text{ dB for medium sized city and suburban centres with medium tree density} \\ 3 \text{ dB for metroplian centres} \end{cases} \quad (5.15)$$

The application of the COST 213 - Hata model is restricted to large and small micro-cells, i.e. base station antenna heights above rooftop levels adjacent to the base station. Hata's formulation and its modifications must not be used for micro-cells.

5.2.5. The COST 231-Walfisch-Ikegami model

COST231 [COST231, 99] proposed a combination of the Walfisch [Walfisch, 88] and Ikegami [Ikegami, 84] models. The model allows for improved path loss estimation by consideration of more data to describe the character of the urban environment, namely:

- Heights of the buildings h_{roof} ,
- Widths of the roads w ,
- Building separation b ,
- Road orientation with respect to the direct radio path.

The parameters are defined in figures 5.1 and 5.2. However this model is more statistical and not deterministic because only characteristics values can be inserted and no topographical database of the building is considered.

The model distinguished between line-of-sight (LOS) and non-line-of-sight (NLOS) situations. In the LOS case, between base station and mobile antennas within a street canyon, a simple propagation loss formula different from free space loss is applied.

$$L[\text{dB}] = 42.6 + 26 \cdot \log d + 20 \cdot \log f \quad \text{for } d \geq 20 \text{ m} \quad (5.16)$$

where the first constant is determined in such a way that L is equal to free space loss for $d = 20$ m. In the NLOS case the basic transmission loss is composed of the terms free space loss L_0 , multiple screen diffraction loss L_{msd} and rooftop-to-street diffraction and scatter loss L_{rts} :

$$L = \begin{cases} L_0 + L_{\text{rts}} + L_{\text{msd}} & \text{for } L_{\text{rts}} + L_{\text{msd}} > 0 \\ L_0 & \text{for } L_{\text{rts}} + L_{\text{msd}} \leq 0 \end{cases} \quad (5.17)$$

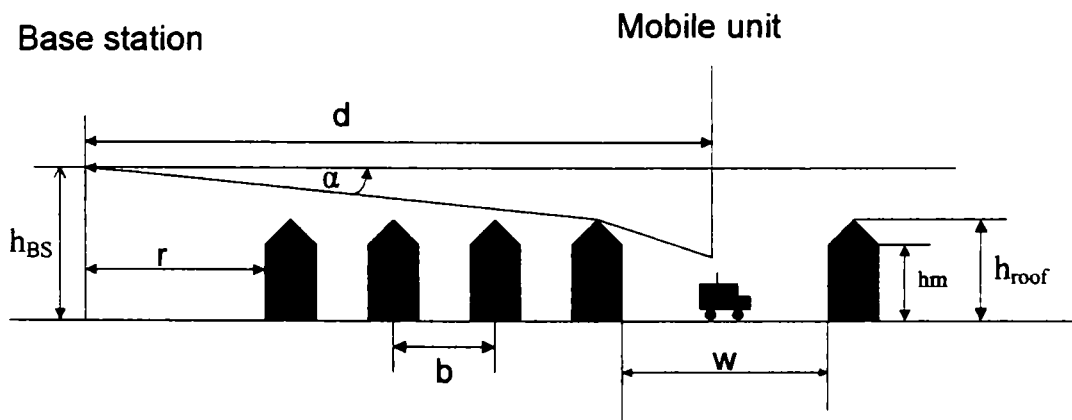


Figure 5.1. Typical propagation situation in urban areas and definition of the parameters used in the COST 231-Walfisch-Ikegami model

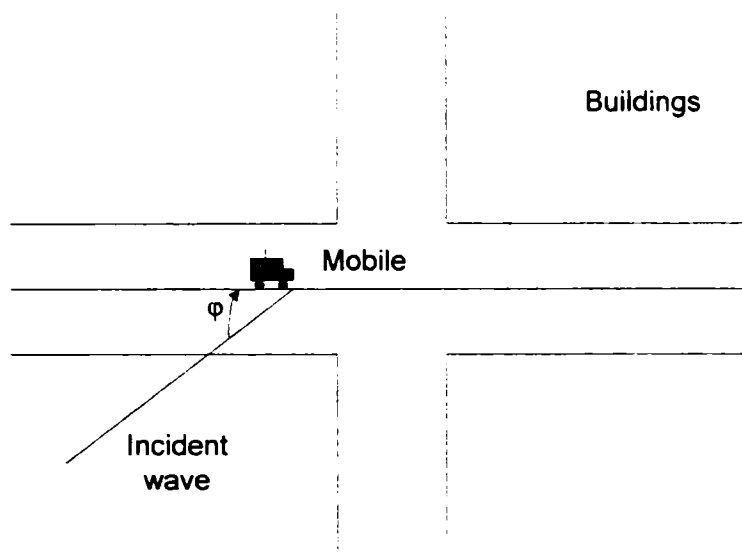


Figure 5.2. Definition of the street orientation

The free space loss is given by:

$$L_0[\text{dB}] = 32.4 + 20 \cdot \log d + 20 \cdot \log f \quad (5.18)$$

The term L_{rts} describes the coupling of the wave propagating along the multiple-screen path into the street where the mobile station is located. The determination of L_{rts} is mainly based on Ikegami's model. It takes into account the width of the street and the street orientation. COST 213 has applied another street-orientation function than Ikegami:

$$L_{\text{rts}} = -16.9 - 10 \cdot \log w + 10 \cdot \log f + 20 \cdot \log \Delta h_M + L_{\text{ori}} \quad (5.19)$$

$$L_{\text{ori}} = \begin{cases} -10 + 0.354\phi & \text{for } 0^\circ \leq \phi < 35^\circ \\ 2.5 + 0.075(\phi - 35) & \text{for } 35^\circ \leq \phi < 55^\circ \\ 4.0 - 0.114(\phi - 55) & \text{for } 55^\circ \leq \phi < 90^\circ \end{cases} \quad (5.20)$$

$$\Delta h_M = h_{\text{roof}} - h_M \quad (5.21)$$

$$\Delta h_{\text{BS}} = h_{\text{BS}} - h_{\text{roof}} \quad (5.22)$$

Scalar electromagnetic formulation of multi-screen diffraction results in an integral for which Walfisch and Bertoni published an approximate solution in the case of base station antenna located above rooftops. COST 231 extends this model for base station antenna heights below the rooftop levels using an empirical function based on measurements. The heights of buildings and their spatial separation along the direct radio path are modeled by absorbing screens for the determination of L_{msd} :

$$L_{\text{msd}} = L_{\text{bsh}} + k_a + k_d \cdot \log d + k_f \cdot \log f - 9 \cdot \log b \quad (5.23)$$

where:

$$L_{\text{bsh}} = \begin{cases} -18 \cdot \log(1 + \Delta h_{\text{BS}}) & \text{for } h_{\text{BS}} > h_{\text{roof}} \\ 0 & \text{for } h_{\text{BS}} \leq h_{\text{roof}} \end{cases} \quad (5.24)$$

$$k_a = \begin{cases} 54 & \text{for } h_{\text{BS}} > h_{\text{roof}} \\ 54 - 0.8 \cdot \Delta h_{\text{BS}} & \text{for } d \geq 0.5 \text{ km and } h_{\text{BS}} \leq h_{\text{roof}} \\ 54 - 0.8 \cdot \Delta h_{\text{BS}} \cdot \frac{d}{0.5} & \text{for } d < 0.5 \text{ km and } h_{\text{BS}} \leq h_{\text{roof}} \end{cases} \quad (5.25)$$

$$k_d = \begin{cases} 18 & \text{for } h_{\text{BS}} > h_{\text{roof}} \\ 18 - 15 \cdot \frac{\Delta h_{\text{BS}}}{h_{\text{roof}}} & \text{for } h_{\text{BS}} \leq h_{\text{roof}} \end{cases} \quad (5.26)$$

$$k_f = -4 + \begin{cases} 0.7 \cdot \left(\frac{f}{925} - 1 \right) & \text{for medium sized city and suburban} \\ & \text{centres with medium tree density} \\ 1.5 \cdot \left(\frac{f}{925} - 1 \right) & \text{for metropoli tan centres} \end{cases} \quad (5.27)$$

The term k_a represents the increase of the path loss for base station antennas below the rooftops of the adjacent buildings. The terms k_d and k_f control the dependence of multi-screen diffraction loss versus distance and radio frequency respectively.

The COST231-Walfisch-Ikegami model is restricted to:

f: 800 ... 2000 MHz

h_{BS} : 4 ... 50 m

h_{M} : 1 ... 3 m

d: 0.02 ... 5 km

The estimation of path loss agrees rather well with measurements for base station antenna heights above rooftop levels. The prediction error becomes large for $h_{\text{BS}} \approx h_{\text{roof}}$ compared to situations where $h_{\text{BS}} \gg h_{\text{roof}}$. The performance of the model is poor for $h_{\text{BS}} \ll h_{\text{roof}}$. The parameters b , w and ϕ are not considered in a physically meaningful way for micro-cells. Therefore the prediction error for micro-cell may be quite large. The model does not consider multipath propagation and the reliability of path loss estimation decreases also if terrain is not flat or the land cover is inhomogeneous.

5.2.6. Walfisch and Bertoni model

A model developed by Walfisch and Bertoni [Walfisch, 88] considers the impact of rooftops and building height by using diffraction to predict average signal strength at street level. This model is a geometrical model of propagation process that takes place in urban environments, having a range dependence on $1/d^{3.8}$ for low transmitting antennas. It shows how the buildings influence the propagation and hence identifies those physical properties that are significant. It can be considered as a simpler version of COST231-Walfisch-Ikegami model. The propagation path loss, in dB is given by:

$$L = L_0 + L_{ex} \quad (5.28)$$

where L_0 represents the free space loss given by equation (5.23) and L_{ex} represents the excess path loss and is given by:

$$L_{ex} = 57.1 + A + \log f + 18 \cdot \log d - 18 \cdot \log(\Delta h_{BS}) - 18 \cdot \log\left(1 - \frac{d^2}{17 \cdot \Delta h_{BS}}\right) \quad (5.29)$$

where d is the distance between transmitter and receiver antennas, f is the frequency, Δh_{BS} is given by equation (5.27) and the last term in equation (5.34) accounts for the curvature of the earth. The influence of building geometry is contained in the term

$$A = 5 \cdot \log\left[\left(\frac{d}{2}\right)^2 + (\Delta h_M)^2\right] - 9 \cdot \log b + 20 \cdot \log\left\{\tan^{-1}\left[\frac{2 \cdot \Delta h_M}{b}\right]\right\} \quad (5.30)$$

where Δh_M is given by equation (5.26) and b represents the building separation.

5.2.7. Xia model

In [Xia, 97] is presented an analytical propagation model that explains the path loss in urban and suburban environments as a result of signal reduction due to free space wavefront spreading, multiple diffraction past rows of buildings, as well as building shadowing. In this model, the key system parameters such as frequency, base and mobile station antenna height are analytically explicit; and the building environments are specified by using the average building height and street width. Hence, the model is applicable for both cellular and PCS frequencies and is valid for various antenna heights employed for both macro – cellular and micro – cellular propagation environments. The radio propagation paths and the geometrical parameters used in this model are depicted in Figure 5.3 [Xia, 97].

For base station antennas near the average rooftop level, the path loss can be evaluated

as:

$$L = -10 \log \left(\frac{\lambda}{2\sqrt{2}\pi d} \right)^2 - 10 \log \left[\frac{\lambda}{2\pi^2 r} \left(\frac{1}{\theta} - \frac{1}{2\pi + \theta} \right)^2 \right] - 10 \log \left(\frac{b}{d} \right)^2 \quad (5.31)$$

For base station antennas above the average rooftop level, the path loss can be written as:

$$L = -10 \log \left[\left(\frac{\lambda}{4\pi d} \right)^2 \right] - 10 \log \left[(2.35^2) \left(\frac{\Delta h_{BS}}{d} \sqrt{\frac{b}{\lambda}} \right)^{1.8} \right] - 10 \log \left[\frac{\lambda}{2\pi^2 r} \left(\frac{1}{\theta} - \frac{1}{2\pi + \theta} \right)^2 \right] \quad (5.32)$$

For base station antennas below the average rooftop level, the path loss is given by:

$$L = -10 \log \left(\frac{\lambda}{2\sqrt{2}\pi d} \right)^2 - 10 \log \left[\frac{\lambda}{2\pi^2 r} \left(\frac{1}{\theta} - \frac{1}{2\pi + \theta} \right)^2 \right] - 10 \log \left\{ \left[\frac{d}{2\pi(d-b)} \right]^2 \frac{\lambda}{\sqrt{(\Delta h_{BS})^2 + b^2}} \left(\frac{1}{\phi} - \frac{1}{2\pi + \phi} \right)^2 \right\} \quad (5.33)$$

with

$$\theta = \tan^{-1} \left(\frac{\Delta h_M}{x} \right) \quad [\text{rad}] \quad (5.34)$$

$$\phi = \tan^{-1} \left(\frac{\Delta h_{BS}}{d} \right) \quad [\text{rad}] \quad (5.35)$$

$$r = \sqrt{(\Delta h_M)^2 + x^2} \quad (5.36)$$

where:

d = transmitter – to – receiver distance [m];

Δh_{BS} = the base station antenna height with respect to the average rooftop level [m];

Δh_M = the height difference between the average rooftop level and the mobile station antenna [m];

b = average separation distance between the rows of buildings [m];

w = the width of the street [m];

x = the horizontal distance between the mobile station and the diffracting edge [m]; in general, x is taken as $x = w / 2$ by assuming that the mobile travel in the middle lane of the street

Since the dependence on frequency, base and mobile station antenna heights, building height and street width are analytically explicit in the above expressions, the model is applicable for cellular as well as PCS in both macro – cell and micro – cell environments.

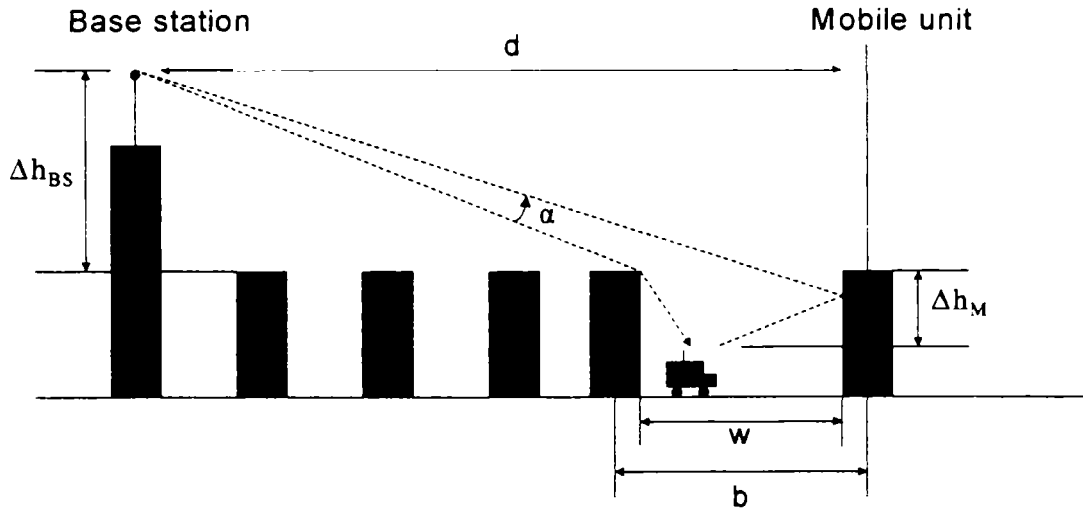


Figure 5.3. Radio propagation paths and geometrical parameters

5.2.8. The Sakagami – Kuboi model

In this model nine input parameters are used like street widths and street orientation. The new parameters used are the average height of buildings around the base station and around the mobile station. The base station antenna height has to be greater than the height of the buildings.

This model is restricted to the following range of parameters:

Frequency f:	400 ... 2200 MHz
Distance d:	0.5 ... 3 km
Location base and mobile stations: $h_{BS} - h_M$:	20 ... 100 m
Average height of buildings around mobile station:	5 ... 50 m
Height of buildings near mobile station:	5 ... 80 m
Street width w:	5 ... 50 m

$$L = 100 - 7.1 \log w + 0.023 \phi + 1.4 \log H_{R,M} + 1.1 \log h_{R,M} - \left(24.37 - 3.7 \left(\frac{h_{R,BS}}{h_{BS}} \right)^2 \right) \log h_{BS,M} + (43.42 - 3.1 \log h_{BS,M}) \log d + 20 \log f + \exp(13(\log f - 3.23)) \quad (5.37)$$

where:

L = the basic path loss [dB],

f = the frequency [MHz],

d = the distance between transmitter and receiver [km],

w = the street width [m],

$h_{BS,M} = h_{BS} - h_M$ [m],

$h_{R,BS}$ = the average height of buildings around base station [m],

$h_{R,M}$ = the average height of buildings around mobile station [m],

$H_{R,M}$ = the height of buildings nearby mobile station [m],

φ = the angle between the transmitter – receiver direction and the street direction.

5.2.9. The Log-distance path loss model

Most radio propagation models are derived using combination of analytical and empirical models. The empirical approach is based on fitting curves or analytical expressions that recreate a set of measured data. This has the advantage of implicitly taking into account all propagation factors, both known and unknown, through actual field measurements. However, the validity of an empirical model at transmission frequencies or environments other than those used to derive the model can only be established by additional measured data in the new environment at the required transmission frequency.

Both theoretical and measurement-based propagation models indicate that average received signal power decreases logarithmically with the distance between the transmitter and the receiver. The average path loss for an arbitrary transmitter-receiver separation is expressed as a function of distance by using a path loss exponent, n .

$$L[\text{dB}] = L(d_0) + 10 \cdot n \cdot \log\left(\frac{d}{d_0}\right) \quad (5.38)$$

where n is the path loss exponent that indicates the rate at which the path loss increases with distance, d_0 is the reference distance determined from measurements close to the transmitter and d is the distance between the transmitter and the receiver. The value of n depends on the specific propagation environment. For example, in free space, n is equal to 2 and when obstructions are present, n has a large value. In macro-cells systems, the reference distance commonly used is 1 km [Lee, 85]. The reference distance should always be in the far field of the antenna so that near-field effects do not alter the reference path loss. The reference path loss is computed using the free space path loss formula or through field measurements at distance d_0 .

5.2.10. Discussions

A number of prediction methods have been described in this section. The methods differ widely in approach, complexity and accuracy. Quite often, the application of two different methods to precisely the same problem will yield results that differ by a wide margin, thereby producing a degree of uncertainty and lack of confidence on the part of the user. One fact is quite clear; there is no one method that outperforms all others in all conditions as far as accuracy is concerned. Very often, accurate prediction of signal strength is a secondary consideration, especially close to the transmitter. Often the primary concern is to predict the limits of the coverage area of a given base site and to identify the inevitable “black spots” that occur; other objectives may be to predict the probability of interference between services or to plan a frequency assignment strategy for radio channels. Choosing a method applicable to the specific problem under consideration is a vital step in reaching a valid prediction.

In general, the models described are a mixture of empiricism and the application of propagation theory. The empirical approach relies on fitting curves or analytical expressions to sets of measured data and has the advantage of implicitly taken all factors (known and unknown) into account. However, a purely empirical model must always be subjected to stringent validation by testing it on data sets collected at locations and transmission frequencies other than those used to produce the model in the first place. Theoretical equations such as those for the free space or plane earth propagation loss often underpin models that provide additional empirical (or semi – empirical) factors to account for diffraction loss, earth curvature, atmospheric effects or building clutter. In deriving prediction models and in considering the applicability of a particular model to a specific problem, it is prudent to consider the input data required by the model, the availability of that data and the effect on the prediction if only partial input data is available.

5.3. Micro-cell propagation models

A micro-cell is a relatively small outdoor area such as a street with the base station antenna below the rooftops of the surrounding buildings. The coverage area is smaller compared to macro-cells and it is shaped by surrounding buildings. A micro-cell enables an efficient use of the limited frequency spectrum and it provides a cheaper infrastructure. The main assumptions are relatively short radio paths (on the order of 200 m to 1000 m), low base station antennas (on the order of 3 m to 10 m) and low transmitting powers (on the order of 10 mW to 1 W).

5.3.1. Model 1. Two-ray model.

Numerous propagation models for micro-cells are based on the ray-optic theory. In comparison with the case of macro-cells, the prediction of micro-cell coverage based on the ray-model is more accurate. One of the elementary models is the two-ray model. The two-ray model [Xia, 93] is used for modeling of the LOS radio channel and is described in Figure 5.4.

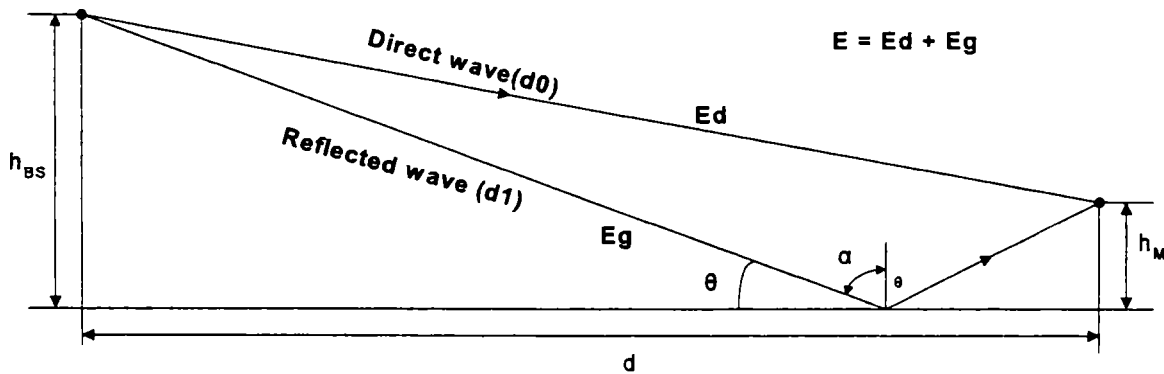


Figure 5.4. Two-ray model

The transmitting antenna of height h_{BS} and the receiving antenna of height h_M are placed at distance d from each other. The received signal P_r for isotropic antennas, obtained by summing the contribution from each ray, is expressed as:

$$P_r = P_t \left(\frac{\lambda}{4\pi} \right)^2 \left| \frac{1}{d_1} \exp(-jk d_1) + \Gamma(\theta) \frac{1}{d_2} \exp(-jk d_2) \right|^2 \quad (5.39)$$

where P_t is the transmitted power, d_1 is the direct distance from the transmitter to the receiver, d_2 is the distance through reflection on the ground and $\Gamma(\theta)$ is the reflection coefficient depending on the angle of incidence α and the polarization.

The reflection coefficient is given by:

$$\Gamma(\alpha) = \frac{\cos \alpha - a\sqrt{\epsilon_r - \sin^2 \alpha}}{\cos \alpha + a\sqrt{\epsilon_r - \sin^2 \alpha}} \quad (5.40)$$

where $\alpha = 90^\circ - \theta$ and $a = 1/\epsilon$ or 1 for vertical or horizontal polarization, respectively. ϵ_r is a relative dielectric constant of the ground.

For large distances α is small, and $\Gamma(\alpha)$ is approximately equal to -1 . For short distances, the value of $\Gamma(\alpha)$ decreases and it can even be zero for vertical polarization.

Also, there are more complex models based on the ray-optic theory. The four-ray model consists of a direct ray, ground reflected ray and two rays reflected by buildings. The six-ray model, besides the direct and the ground reflected rays, takes four rays reflected by the building walls along the street. If a model considers a larger number of rays, the prediction tends to be more accurate, but the computational time is significantly increased. The problem deserving special attention is that of corner diffraction. Two popular models considering this effect are the GTD (Geometrical Theory of Diffraction) model [Luebbers, 84] and the UTD (Uniform Theory of Diffraction) model [Kouyoumijan, 74].

5.3.2. Model 2

The models proposed in [Harley, 89] and [Lotse, 90] describe the measured signal level along the line-of-sight path. According to these models, the road guided waves are expected to exist only for short ranges. This situation can be described by two distinct path-loss slopes and a breakpoint. The breakpoint is the distance from the base station that is equal to the maximum distance that has the first Fresnel zone clear. The breakpoint can be used to define the size of a micro-cell because the signal level decreases more rapidly when the distance increases after the breakpoint.

The form of the proposed propagation models is given by:

$$S = -20 \log \left[d^a \cdot \left(1 + \frac{d}{g} \right)^b \right] + c \quad (5.41)$$

where S is the signal level in $\text{dB}\mu\text{V}/\text{m}$, d is the distance from the transmitting antenna (m), a is the basic attenuation rate for short distances, b is the additional attenuation rate coefficient for the distances greater than 100 to 200 m, g is the distance to the breakpoint and c is a scalable factor.

The expression is valid for 5-20 m antenna heights and 200m-1 km distances. This model, whose coefficients are relatively independent, has two boundary cases:

1. At distances less than the breakpoint, the form of the propagation model is:

$$S = -20 \log(d^a) + c \quad (5.42)$$

2. At distances greater than the breakpoint, the form of the propagation model is:

$$S = -20 \log(d^{a+b}) + c + \text{const} \quad (5.43)$$

In addition, the signal around the corner decreases by 20–25 dB in a short transition distance of only several tens of meters.

It was shown that for the same conditions, the results of the proposed models for a micro-cell situation are better than those of the normal linear regression and the Okumura model.

5.3.3. Model 3. Wideband PCS micro-cell model

In [Feuerstein, 94], using base station antenna heights of 3.7 m, 8.5 m and 13.3 m, and a mobile receiver with an antenna height of 1.7 m above ground, statistics for path loss, multipath and coverage area were developed from extensive measurements in line-of-sight (LOS) and obstructed (OBS) environments, at 1900 MHz. This work revealed that a 2-ray ground reflection model is good to estimate for path loss in LOS micro-cells and a simple log-distance path loss model holds well for OBS micro-cell environments.

For a flat earth ground reflection model, the distance d_f at which the first Fresnel zone just becomes obstructed by the ground (first Fresnel zone clearance) is given by:

$$\begin{aligned} d_f &= \frac{1}{\lambda} \sqrt{(\Sigma^2 - \Delta^2)^2 - 2(\Sigma^2 - \Delta^2) \left(\frac{\lambda}{2}\right)^2 + \left(\frac{\lambda}{2}\right)^4} \\ &= \frac{1}{\lambda} \sqrt{16h_t^2 h_r^2 - \lambda^2 (h_t^2 + h_r^2) + \frac{\lambda^4}{16}} \end{aligned} \quad (5.44)$$

For LOS cases, a double regression path loss model that uses a regression breakpoint at the first Fresnel zone clearance was shown to fit well to measurements. The model assumes omnidirectional vertical antennas and predicts average path loss as:

$$L(d) = \begin{cases} 10n_1 \log(d) + p_1 & \text{for } 1 < d < d_f \\ 10n_2 \log\left(\frac{d}{d_f}\right) + 10n_1 \log(d_f) + p_1 & \text{for } d > d_f \end{cases} \quad (5.45)$$

where p_1 is equal to $L(d_0)$ (the path loss in dB at the reference distance of $d_0 = 1$ m), d is in meters and n_1, n_2 are path loss exponents which are a function of transmitter height, as given in Table 5.2. It can easily be shown that at 1900 MHz, $p_1 = 38$ dB.

Table 5.2. Parameters for the wideband microcell model at 1900 MHz from [Feuerstein, 94]

Transmitter antenna height	1900 MHz LOS			1900 MHz OBS	
	n_1	N_2	σ [dB]	n	σ [dB]
Low (3.7 m)	2.18	3.29	8.76	2.58	9.31
Medium (8.5 m)	2.17	3.36	7.88	2.56	7.67
High (13.3 m)	2.07	4.16	8.77	2.69	7.94

For the OBS case, the path loss was found to fit the standard log-distance path loss law

$$L(d)[\text{dB}] = 10n \log(d) + p_1 \quad (5.46)$$

where n is the OBS path loss exponent given in Table 5.2 as a function of transmitter height. The standard deviation (in dB) of the log - normal shadowing component about the distance-dependent mean was found from measurements. The log - normal shadowing component is also listed as a function of height for both the LOS and OBS micro - cell environments. Table 5.2 indicates that the log - normal shadowing component is between 7 and 9 regardless of antenna height. It can be seen that the LOS environments provide slightly less path loss than the theoretical 2-ray ground reflected model that would predict $n_1 = 2$ and $n_2 = 4$.

5.3.4. Model 4. Lee model

When the size of the cell is small, less than 1 km, the street orientation and individual blocks of buildings make a difference in signal reception. Those street orientations and individual blocks of buildings do not make any noticeable difference in reception when the signal is well attenuated at a distance over 1 km. Over a large distance the relatively great mobile radio propagation loss of 40 dB/decade is due to the situation that two waves, direct and reflected, are more or less equal in strength. The local scatterers (buildings surrounding the mobile unit) reflect this signal causing only the multipath fading

not the path loss at the mobile unit. When the cells are small, the signal arriving at the mobile unit is blocked by the individual buildings; this weakens the signal strength and is considered as part of the path loss. Therefore, the loss is computed based on the dimensions of the building blocks. Since the ground incident angles of the waves, are, in general, small due to the low antenna heights used in small cells, the exact height of the buildings in the middle of the propagation paths is not important, as depicted in Figure 5.5 [Lee, 93].

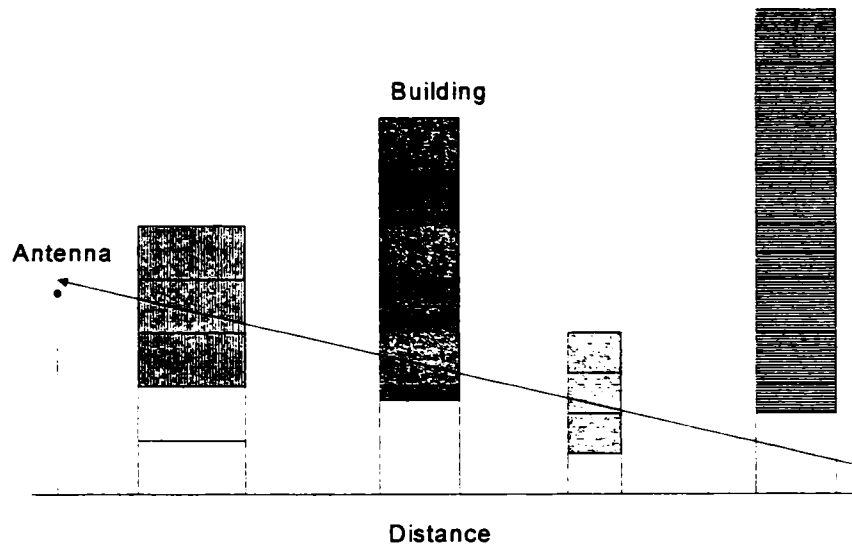


Figure 5.5. The propagation mechanism of low-antenna height at the cell site

The Lee model assumes that there is a high correlation between the signal attenuation and the total depth of building blocks along the radio path. This assumption is not entirely true because the signal received at the mobile unit comes from the multipath reflected waves and not from the waves penetrating through the buildings. However, according to the assumption, if the building blocks are larger, the signal attenuation is higher. An aerial photograph can be used to calculate the proportional length of a direct wave path being attenuated by the building blocks. The line-of-sight reception curve P_{los} is determined from the measurement data along the streets in an open line-of-sight condition. From the measured signal P_{nlos} along the streets in no-line-of-sight conditions within cells, it is formulated the additional signal attenuation α_B curve due to the portion of building blocks over the direct path by subtracting the received signal from P_{los} . The additional signal attenuation α_B can be obtained in the following way:

- Calculate the total blockage length by adding the individual building blocks.
- Measure the signal strength P_{los} for line-of-sight conditions.
- Measure the signal strength P_{nlos} for no-line-of-sight conditions.

- If the signal level at a particular point is P_{nlos} , the distance from the base station to the mobile unit is d_A and B is the blockage length between the transmitter and the receiver. Then the value of α_B for a blockage B can be expressed as:

$$\alpha_B(B) = P_{los}(d = d_A) - P_{nlos} \quad (5.47)$$

A series of measurements have been done for different antenna heights in LOS conditions along many streets and it is observed that the antenna height gain for different antenna heights is 30 dB/decade.

In conclusions, in a micro-cell prediction model the two curves P_{los} and α_B are used to predict the received signal strength. Therefore, the micro-cell prediction model is given by:

$$P_r = P_{los} - \alpha_B \quad (5.48)$$

where P_{los} is the line-of-sight path loss (measured) and α_B is the additional loss due to the length of the total building blocks B along the paths.

The original Lee model exhibits large errors in the following situations:

- When the prediction point is in the main street, but there is no direct LOS path.
- When the prediction point is in a side street near an intersection and large building blocks exist between the point of prediction and the transmitter (the case when the side street and the transmitter location are on the same side of the main street).

The accuracy of the model can be significantly improved by introducing specific corrections based on the arrangement of the streets and their types [Neskovic, 97], [Neskovic, 00a]. There are significant differences in the propagation of radio waves in different types of streets (for example, a main street under LOS conditions, a main street under NLOS conditions, a narrow side street, a wide side street and a street parallel to the main). After these corrections are added to the model, the signal level in side streets and in the main street under NLOS conditions is given by:

$$P_{nlos}[\text{dB}] = P_{los}(\text{LOS - distance}) - \alpha_{st}(\text{NLOS - distance}) \quad (5.49)$$

where P_{nlos} is the estimated signal level in the street under NLOS propagation conditions, P_{los} is the signal level on the LOS path at the intersection of the main and side street (at a LOS-distance from the transmitter) and α_B is the correction of the signal level in the side street at the NLOS-distance from the intersection.

5.3.5. COST 231 models

The propagation models developed in COST 231 are based on theoretical and empirical approaches. Ray optical methods with either simplified analytical solutions or pure ray tracing techniques have been proposed. The availability and usage of proper urban terrain databases in combination with ray tracing methods enables site-specific propagation modeling for the prediction of path loss and time spreading of the signal; the latter has a major impact on the performance of digital radio systems.

Radio transmission in urban environments is subject to strong multipath propagation. To consider these effects in a propagation model, it is necessary to gain knowledge of all dominant propagation paths. These paths depend primarily on the base station height with respect to the building heights around. A study on micro-cellular multipath propagation effects with respect to DECT system performance is given in [Kauschke, 94]. For simplification of propagation modeling several two-dimensional models have been developed under the assumption of infinitely high buildings. Hence these models only take into account wave propagation around buildings. As a result, computation-time efficient, analytical path loss models have been derived considering simple building geometries. In case of low building heights, over-rooftop propagation has to be regarded, too.

The second group of small-cell models allows a very site specific three-dimensional path loss and signal spread prediction for base station heights below as well as above rooftop level of the buildings. Hence, not only the shape but also the height of a building has to be incorporated. Of course, due to the three-dimensional ray tracing these models require a higher computation time than the simplified approaches mentioned above.

The micro-cell models are generally valid only for flat urban area. Investigations on the influence of terrain on micro-cell propagation are presented in [Bertoni, 94] and [Wiert, 95]. Further on the effects of urban-type vegetation (like line-up trees, parks, etc) on radio propagation [Causebrook, 90] are not included in these micro-cell models. But aspects are of great interest from an engineering point of view and should be regarded in further developments of these models.

5.3.6. Model based on UTD and Multiple Image Theory

The model is quasi three-dimensional UTD propagation model [Tan, 96], which functions well for micro-cellular applications. A multiple image concept and generalized Fermat's principle are used to describe the multiple reflections and diffractions. It is assumed that the building walls are much higher than the transmitter height so that the diffraction from the rooftops can be neglected. The model considers various line-of-sight propagation paths and, also, non-line-of-sight paths. The propagation paths taking a large number of corners and building walls are not necessarily coplanar. This model includes multiple reflections between wall-to-wall, wall-to-ground, ground-to-wall, the diffraction from the corners of buildings, and also subsequent reflections from such diffracted signals. The relative contributions of the diffraction and reflection components to the total received signal along a side street depend on the parameters such as the widths of the main street, side streets, parallel streets, the distance from the transmitter to the junction, the reflectivity of the surfaces, etc.

The location and the sequence of all images have to be determined for making use of multiple images concept. A "test ray" or ray-launching technique is used to the plan view of the street grid. The intersection of a ray with an object is the fundamental operation in the ray-launching technique. This image concept makes the determination of the exact point of reflection at a wall or at the ground surface. In the case of diffraction, the location of the point of diffraction at the edges has to be determined. The local ray-fixed coordinate system or edge-fixed coordinate system and appropriate reflection or diffraction coefficients are used for each reflection or diffraction. The accuracy of this model is limited mainly by the assumption of characterizing the tall building walls as "smoothed-out" flat surfaces with average relative permittivity ϵ_r and conductivity σ .

The UTD model considers a single ray at a time. Naturally, many rays will contribute to the received signal at a particular location of the receiver. The UTD approach takes the vector sum of all the reflected and diffracted rays. In general, a total of j multiple wall reflections from the main street, side streets, parallel streets and, at most, one ground reflection, with or without diffractions from the building corners at the junctions can arrive at the receiver. This is equivalent to including the multiple transmitter images. Since each reflection or diffraction causes a loss in signal strength, the value of j will depend on the values of σ and ϵ_r of the walls and ground surfaces as well as the geometry of the environment.

5.3.7. Discussions

The concept of micro-cells has been suggested as the best solution in heavily built-up areas. A micro-cell may have dimensions of only a few hundred meters up to 1 kilometer with (see Table 5.1) with base transceivers mounted at the street – lamp level. Each fixed base transceiver is associated with a micro-cell and in order to provide adequate coverage there is a partial overlap. The extent of coverage from a given base can be adjusted by the use of different antenna patterns and heights and by controlling the transmitted power [Parsons, 92].

The fact that propagation is required only over short ranges suggests that much higher frequencies can be used in micro-cellular systems. Microwave frequencies offer a much-enhanced bandwidth potential and the wavelengths are such that space diversity can be accommodated even within the dimensions of hand – portable equipment. A suggestion that has been much discussed is the use of frequencies where the radiation is partially absorbed by oxygen molecules in the atmosphere. Such resonant absorption lines are found in the frequency range from about 50 to 70 GHz and the portion of the radio spectrum from 51.4 to 66 GHz has been designated absorption band A1. Some parts of this band have already been provisionally allocated for mobile communications. The attenuation due to oxygen in the atmosphere at ground level has been measured [King, 77] as 16 dB/km and this is in addition to the normal spatial attenuation and losses due to rain. The 60 GHz band has been advocated for use in micro – cellular systems because of the oxygen absorption effects, the potentially large available bandwidths and the prospect of affordable transceivers for the public. Micro – cellular systems, however, are not restricted to any particular frequency band and investigations of short range propagation characteristics have also been carried out at other frequencies [Parsons, 92].

In addition to 60 GHz, there have also been measurements made at lower frequencies, particularly around 11 GHz [Reudink, 72], [Rustako, 89] in both urban and rural areas. Reudink's results at 11.2 GHz in New Jersey and New York City were not obtained with micro-cellular systems in mind and the base station antenna heights were typical of large-cell systems.

Several research workers, e.g. Young, Okumura have made measurements in urban areas at frequencies up to 3.7 GHz. These were not short range measurements with relatively low base station antennas, designed with micro-cellular systems in mind.

Measurements of the latter kind show somewhat different results, principally because a line-of-sight path often exists over much of the coverage area [Parsons, 92].

Whitteher [Whitteker, 88] has conducted measurements at 910 MHz in Ottawa, Canada, using street lamp level transmitter at height 8.5 m above the ground while the receiving antenna was 3.65 m. The path loss was found to be close to free space value along the street on which the transmitter was located (LOS) and 20 dB or more greater than free space value on other streets (NLOS) except where there was an open area between transmitter and receiver. The received power was found to depend in a detail way on the distribution of buildings and the open areas between them.

A comparative study at 900 MHz and 1.8 GHz has been undertaken in Melbourne, Australia [Harley, 89]. Antenna heights between 5 and 20 m were used for the fixed site, these being well below the height of the surrounding buildings. The signal strength was measured at distances up to 1 km, the measuring vehicle having an antenna mounted 1.5 m above street level. A line-of-sight path always existed. It was found that the attenuation could be modeled by two straight lines, one representing the region between the base station and a point about 150 m away, the other representing greater distances. Close to the base station the slope is inverse square, which is suggestive of a ducting mode: beyond the so-called “turning – point” the slope is greater. Other authors have also reported a turning point in the path loss curve, for example, results presented by Kaji and Akeyama [Kaji, 85] show a change at about 350 m and a similar trend was observed at 900 MHz in New Zealand [Williamson, 84]. There is a high correlation between the measurements made at 900 MHz and 1.8 GHz, although there is often a larger spread in the results at the higher frequency. It has been observed that the 1.8 GHz signal is more sensitive to shadowing when other vehicles move in front of the measurement vehicle and this could be the main reason for the increased variability.

5.4. INDOOR PROPAGATION MODELS

5.4.1. Indoor radio propagation environment

Indoor radio propagation is a very complex and difficult radio propagation environment and differs from the outdoor mobile radio channels in two aspects: the distances between transmitter and receiver is shorter (due to high attenuation caused by internal walls and furniture and often also because of the lower transmitter power) and the variability of the environments is much greater for a much smaller range of transmitter-receiver separation distances.

Indoor radio propagation is dominated by the same mechanisms as outdoor: reflection, diffraction and scattering. However, conditions are much more variable. For example, signal levels vary greatly depending on whether interior doors are open or closed inside a building. Where antennas are mounted also impacts large-scale propagation. Antennas, mounted at desk level in a partitioned office, vastly different signals than those mounted on the ceiling. Also, the smaller propagation distances make it more difficult to insure far-field radiation for all receiver locations and types of antennas.

Signals propagate along the corridors and other open areas, depending on the structure of the building. The results of measurements indicate that the signal variation inside buildings is approximately Rayleigh distributed for non-line-of-sight (NLOS) cases, whereas a Rice distribution fits in the case of line-of-sight (LOS). Therefore, in modeling indoor propagation the following environmental parameters must be considered:

1. Construction materials: reinforced concrete, brick, metal, glass, wood and plastic.
2. Types of interiors: rooms with windows, rooms without windows, hallways with doors, hallways without doors, large hallways or open area, corridors with corners and curved corridors.
3. Locations within a building: ground floor, n^{th} floor and basement or garage.
4. Location of transmitting (T) or receiving (R) antennas: transmitting and receiving antennas on the same floor; transmitting and receiving antennas within the building, but on different floors and transmitting antenna outside the building, receiving antenna inside the building.

5.4.2. Empirical narrowband models

5.4.2.1 Model 1

A common model for estimating the mean path loss (in the case where the transmitter and the receiver are on the same floor, inside a hall or a large area) is given by:

$$\bar{L}(d)[\text{dB}] = L(d_0)[\text{dB}] + 10n \log\left(\frac{d}{d_0}\right) \quad (5.50)$$

where:

$L(d)$ = the mean path loss, in dB,

d = distance from the transmitter, in meters,

d_0 = the reference distance, in meters

$L(d_0)$ = the path loss at reference distance d_0 ,

n = specifies the path loss behavior for a particularly type of building.

In micro-cellular and indoor environments, a reference distance of $d_0 = 1$ m is typically used.

This model is easy to use because only the distance between transmitter and receiver appears as an input parameter. However, the dependency of these parameters on environment category has to be taken into account.

Local large-scale path loss is often log-normally distributed about the mean power level described by equation (5.50). That is

$$L(d)[\text{dB}] = \bar{L}(d)[\text{dB}] + X_\sigma[\text{dB}] \quad (5.51)$$

where:

$L(d)$ = path loss, in dB, at distance d ,

X_σ = a normal random variable in dB having a standard deviation of σ dB (the uncertainty of the model).

The standard deviation σ of the zero-mean log-normal random variable X_σ provides a quantitative measure of the variability of the model used to predict the path loss.

This model is simple, efficient and suitable for computer implementation. During implementation, the environmental database is unnecessary. Therefore, there is no requirement for investing time and resources in surveying building layouts. Due to the model simplicity, great accuracy could not be expected. The main parameter n is very sensitive to the

propagation environment, i.e. the type of construction material, type of interior, location within building, etc.

Several researchers have attempted to modify equation (5.50) in order to obtain a better fit (smaller σ) to their measured data for different indoor environments. The values of n range from 1.2 to 6. In addition, the value of n depends on the way the statistical analysis of the measurement data is performed. In free space, $n = 2$. For $n < 2$ we have a wave guide effect [Cox, 84]. Preliminary measurements conducted by [Bergljung, 89] gives the value of $n = 2.9$.

5.4.2.2. Model 2

This model describes the non-line-of-sight (NLOS) situation. If a wall(s) exists between the transmitter and the receiver (they are on the same floor) and the only signal path is through the wall(s), with no channeling effect around corners between transmitter and receiver, then path loss is given by:

$$L = L_0 + 10n \log(d_1) + \sum L_w \quad (5.52)$$

where:

L = the path loss in dB,

L_0 = the path loss at 1 meter distance from the transmitter,

n = the exponent depending on the environment outside the wall,

d_1 = distance between transmitter and external surface of the wall, in meters,

L_w = the penetration loss due to the wall(s).

The parameter L_w depends on the type of the wall construction between the transmitter and the receiver and the angle of incidence of the transmitted wave.

In the case where more than one wall exists between the transmitter and the receiver, a detailed analysis is required to calculate the total loss ($\sum L_w$). In this case, the actual geometry of the walls with respect to the incident wave must be considered as well as the construction materials of each wall.

5.4.2.3. Model 3. Floor Attenuation Factor model

An indoor propagation model that includes the effect of building type as well as the variations caused by obstacles was described in [Seidel, 92]. This model provides flexibility

and was shown to reduce the standard deviation between measured and predicted path loss to around 4 dB, as compared to 13 dB when only a log-distance model was used in two different buildings. The attenuation factor model is given by:

$$L [\text{dB}] = L(d_0)[\text{dB}] + 10 n_{\text{SF}} \log\left(\frac{d}{d_0}\right) + \text{FAF} [\text{dB}] \quad (5.53)$$

where:

L = the path loss, in dB,

$L(d_0)$ = the path loss at the reference distance, d_0 ,

d = distance from transmitter, in meters,

d_0 = the reference distance, in meters,

n_{SF} = the exponent value for the “same floor” measurement,

FAF = floor attenuation factor, in dB

Thus, if a good estimate for n exists on the same floor, then the path loss on a different floor can be predicted by adding an appropriate value of FAF [Rappaport, 96]. Typical values for FAF that are a function of a of the number of intervening floors, are about 15 dB for one floor of separation and an additional 6-10 dB for every additional floor up to five floors. There is no significant increase in FAF for more than five floor of separation. The values of FAF have been observed to be smaller at 900 MHz than at 1.7 GHz by about 6 dB.

5.4.2.4. Model 4. The COST 231 – Motley model

The COST 231 – Motley model takes into account obstacles (walls and floors) between antennas. This model is one of the most complete empirical models and has been developed for a vertical polarization. The wall and floor attenuation have been taken into account in this model [Lähteenmäki, 90].

The validity of the model:

- Frequency range: 1700 ... 1900 MHz
- Distance range: 2 ... 100 m
- Location base: 1.5 m
- Location mobile: 1.5 m – ceiling

$$L = L_0 + 10 n \log(d) + \sum_{i=1}^I K_{fi} L_{fi} + \sum_{j=1}^J K_{wj} L_{wj} \quad (5.54)$$

where:

L_0 = the path loss at the reference point (1 meter distance), in dB,

L = the path loss, in dB,

n = power decay index,

d = distance between transmitter and receiver, in meters,

I = number of floor categories,

J = number of wall categories,

L_{fi} = the path loss factor for floor category i ,

L_{wj} = loss factor wall category j ,

K_{fi} = number of transversed floors of category i ,

K_{wj} = number of transversed walls of category j .

5.4.2.5. Model 5. Lafortune model

This empirical model is based on estimations of transmission, reflection and diffraction phenomena occurring in the transmission path [Lafortune, 90]. The model has been developed in the 900 MHz band but the principles used by the author to elaborate it could be applicable in the 1800 MHz band. Note that the model is not reciprocal with respect to transmitter and receiver positioning.

Distance range: 1 ... 50 m

Location base: 1.7 m – ceiling

$$L = L_0 + L_{0B} + G_{RM} \quad (5.55)$$

In case of obstacles

$$L_{0B} = -3.7 + 1.5n + 10.7 \log(d) + \begin{cases} 0 & \text{for } d' < 4\text{m} \\ 7.8 - 15.3 \log(d') & \text{for } d' > 4\text{m} \end{cases} \quad (5.56)$$

and

$$G_{RM} = \begin{cases} 0 & \text{in the general case} \\ -0.2 - 1.8 \log(d) & \text{in main corridor} \end{cases} \quad (5.57)$$

Different correction factors can be added to L_{0B} and/or G_{RM} to take into account specific configurations (see [Lafortune, 90]).

Parameters:

L = the basic path loss, in dB,

L_0 = the free space loss, in dB,

L_{0B} = the loss due to obstacles, in dB,

G_{RM} = the gain due to reflections, in dB,

d = distance from transmitter to receiver, in meters,

f = frequency, in GHz,

n = number of walls between antennas,

d' = distance from transmitter to the first wall, in meters.

This model is complex enough to be implemented and needs a detailed database to be used efficiently.

5.4.2.6. Model 6. The Multi-wall model

The **Multi-wall model** (MWM) gives the path loss as the free space loss added with losses introduced by the walls and floors penetrated by the direct path between the transmitter and the receiver. It has been observed that the total floor loss is a non-linear function of the number of penetrated floors. This characteristic is taken into account by introducing an empirical factor b [Törnevik, 93]. The multi-wall model can then be expressed in the form:

$$L = L_{FS} + L_c + \sum_{i=1}^I k_{wi} L_{wi} + k_f^{\left(\frac{k_f+2}{k_f+1} b\right)} L_f \quad (5.58)$$

where:

L_{FS} = the free space loss between the transmitter and the receiver,

L_c = constant loss,

k_{wi} = number of penetrated walls of type i ,

k_f = number of penetrated floors,

L_{wi} = loss of wall type i ,

L_f = loss between adjacent floors,

b = empirical parameter,

I = number of wall types.

The constant loss L_c is a term that results when wall losses are determined from measurement results by using the multiple linear regression. Normally it is closed to zero. The third term in equation (5.58) expresses the total wall loss as a sum of the walls between transmitter and receiver. For practical reasons the number of different wall types must be kept low. Otherwise, the difference between the wall types is small and their significance in the model becomes unclear. A division into two wall types according to Table 5.3 is proposed in [COST231, 99].

Table 5.3. Wall types for the Multi-wall model

Wall Type	Description
Light wall (L_{w1})	A wall that is not bearing load: e. g. plasterboard, particle board or thin (<10 cm), light concrete wall
Heavy wall (L_{w2})	A load-bearing wall or other thick (>10 cm) wall, made of e. g. concrete or brick

It is important to notice that the loss factors in equation (5.56) are not physical wall losses but model coefficients that are optimized along with the measured path loss data. Consequently, the loss factors implicitly include the effect of furniture as well as the effect of signal paths guided through corridors.

5.4.3. Deterministic models

Deterministic models are used to simulate physically the propagation of radio waves. Therefore the effect of the environment on the propagation parameters can be taken into account more accurately than in empirical models. Another advantage is that deterministic models make it possible to predict several propagation parameters. For example, the path loss, impulse response and angle of arrival can be predicted at the same time.

Several deterministic techniques for propagation modeling can be identified. For indoor applications, especially, the Finite Difference Time Domain (FDTD) technique and the geometrical optics (GO) technique have been studied. In COST 231 the main effort is on the geometrical optics that is more computer efficient than the FDTD. There are two basic approaches in the geometrical optics techniques: the ray-launching approach and the image approach. Computational complexity of ray-tracing methods is considered in [Huschka, 94].

5.4.3.1. Ray-launching model (RLM)

The ray-launching approach involves a number of rays launched at the transmit antenna in specified directions. For each ray its intersection with a wall is determined and the incident ray is divided into a wall penetrating ray and a reflected ray; each of them is traced to its next intersection and so on. A ray is determined when its amplitude falls below a specified threshold or a chosen maximum number of ray-wall interactions are succeeded. In, e.g. [Honcharenko, 92] a uniform angular separation of launching rays is maintained, where the spherical surface is subdivided by a geodesic polyhedron with resulting hexagonally shaped,

wave front portions that are further approximated by circular areas. Whether a ray reaches a receiver point or not can be accomplished by a reception sphere [Schaubach, 92].

In figure 5.6.a, a two-dimensional view of the reception sphere is shown, where the unfolded total path length d and the angular separation γ of adjacent rays launched at the transmitter determines its radius R_{rS} :

$$R_{rS} = \gamma d / \sqrt{3} \quad (5.59)$$

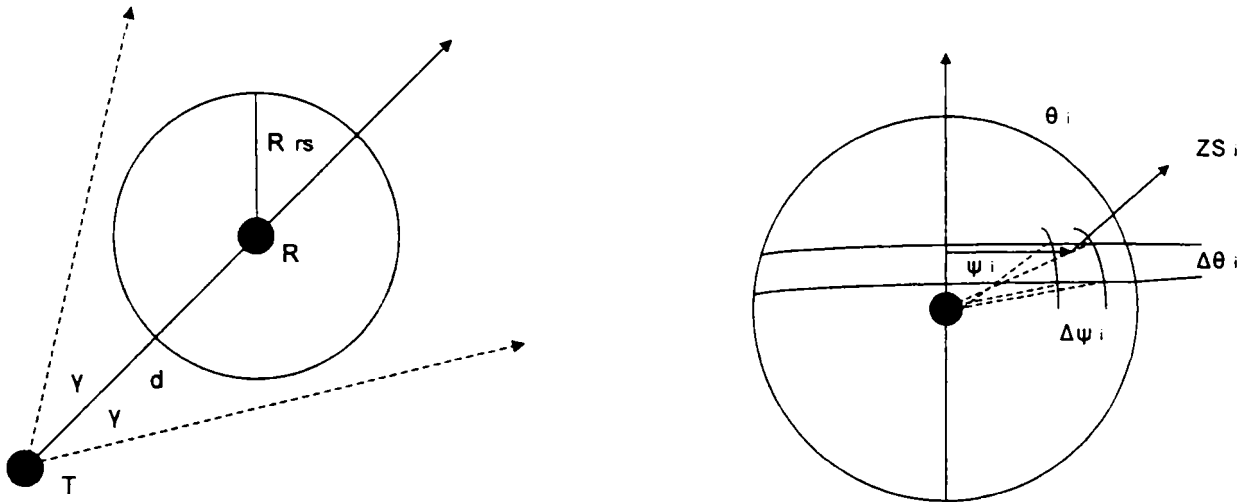


Figure 5.6. (a) - 2-D view of the reception sphere and (b) - Ray launching

However, this is an approximate solution for the 3-D propagation case. To achieve a complete solid angle discretization under maintenance of unambiguous and practical reception determination, the entire solid angle 4π is subdivided into rectangularly shaped, incremental portions of the spherical wave front [Kreuzgruber, 93], [Cichon, 95]. In [Cichon, 94a], [Cichon, 95] the propagation directions θ_i and ψ_i of the central rays of the ray tubes and the corresponding $\Delta\theta_i$ and $\Delta\psi_i$ as depicted in Figure 5.6.b) are determined by (5.60) and (5.61).

$$\Delta\psi_i(\theta_i) = \frac{\Delta\theta}{\sin\theta_i} \quad (5.60)$$

$$\theta_i = \frac{\Delta\theta}{2} + (i-1)\Delta\theta \quad i = 1, \dots, N_\theta \quad \Delta\theta = \text{const.} \quad (5.61)$$

Ray launching approaches are flexible, because different diffracted and scattered rays can also be handled along with the specular reflections. To maintain a sufficient resolution the so-called “ray splitting technique” can be used as given in [Kreuzgruber, 93] and [Cichon, 94b].

5.4.3.2. Ray Tracing method

The ray-tracing algorithm determines all relevant rays for each receiver point independent of the other points. The computation time increases in comparison to the ray launching but on the other hand a constant resolution and accuracy can be obtained. The computation of the field strength is done with the GTD/UTD for diffracted rays and with the Fresnel equations for transmission and reflection.

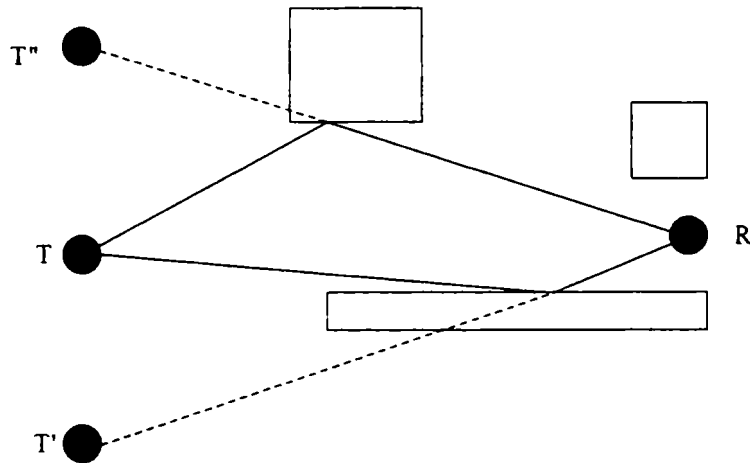


Figure 5.7. Ray tracing.

The differences between empirical and ray-optical predictions are obvious. Especially the wave guiding in corridors due to multiple reflections, the coupling of the waves into the rooms and the diffraction around corners are responsible for the high accuracy of the ray-optical models. Among the main disadvantages of the deterministic prediction models is their very long computation time. Different papers presented some ideas to accelerate the prediction and some ideas lead to acceleration factor beyond 1000, but for these algorithms a preprocessing of the database is necessary and the preprocessing is also very time consuming. A further disadvantage with this type of propagation models is the marked dependence on the accuracy of the database. Small errors in the positions and materials of the walls influence the predicted results. These effects are very important and they limit the use of ray-optical models for the computation of the indoor wave propagation.

6. Neural Networks applications for propagation prediction

6.1. Outdoor environment

The early work reported in [Stocker, 92] and [Gschwendtner, 93] showed that artificial neural networks are able to give good estimates of propagation path loss in rural environment using height profiles as inputs for the network. The proposed neural network is the multilayer perceptron and it is trained with measurement data from different environment databases. In [Stocker, 92] the neural network has the following configuration: 11 input units, 8 hidden units and 1 output unit (11-8-1). The topography is sampled at 10 points that provide the inputs for 10 units. The sampling width is varied and fed to the 11th input. The neural network output is represented by the normalized field strength for an arbitrary point along the path profile at the height of the mobile receiver. It was found a deviation below 6% for the test set.

In [Gschwendtner, 93] the neural model is based on narrowband measurements at 943 MHz taken in the city of Mannheim that is characterized by a regular building structure and almost no variation in terrain height. Therefore, only land usage has been considered and five classes of morphography were distinguished: extremely dense city, dense city, rather dense city, wood and streets. First, a neural network with 21 inputs was used: 20 for the sampled values of land usage along the path profile and 1 for the distance. The prediction leads to a systematical error of - 4.9 dB and standard deviation of 6 dB. It was shown that adding the base station antenna height as an input to the neural network the results are improved: mean error of 1.4 dB and the standard deviation of 5.2 dB. Another example presented in [Gschwendtner, 93] is a neural network trained to reproduce field strength over single idealized wedge mountains. The neural network input consists of 20 sampled values for the height along the path profile and one parameter for distance. For training purposes, four different mountains were used, representing different heights and distances. Testing with an unknown mountain with arbitrary heights and distances, the standard deviation was found 1.8 dB. In a further step, a real world scenario that is, a height profile in an open rural area replaced the idealized structure. It is not obvious that a neural network trained with single idealized mountain leads to at least reasonable results in arbitrary terrain with multiple mountains.

In [Stocker, 93] theoretical investigations into the suitability of neural network simulators for the prediction of field strength based on topographical and morphographical

data are presented. The neural network has 21 inputs and 10 hidden units and is used to approximate field strength for an arbitrary point along the path profile at the height of the mobile receiver. The topographic information along the propagation path is sampled at 20 points, which provide the inputs for 20 units to the neural network. Distance between transmitting and receiving antennas is varied and used as the 21th input.

As a first example, the neural network model is trained directly by measurements made in the German town of Mannheim, at 943 MHz, which provides a typical urban environment with varying buildings heights. Antenna heights are rather low just above rooftops. The situation is characterized by virtually no variation in terrain height. There are distinguished extremely dense city, dense city and rather dense city. The input of the neural network consists of 20 sample point values of morphographical data plus one for the distance parameter. The hidden layer has 15 neurons. The improvement obtained by the neural model is about 1.25 dB in standard deviation, comparing with Hata formula.

The second example presented is a different urban scenario in the city of Darmstadt [Germany]. For training purposes only morphographical data are considered; topographical variations are neglected. The neural network has 21 inputs and 10 hidden units. In this case, improvement obtained by the neural network is about 1.5 dB in standard deviation, comparing with Hata formula.

In [Balandier, 95] a semi-empirical model of field strength prediction combining neural networks and theoretical results of propagation loss algorithms is considered. Field strength measurements used to design the neural network based model have been performed at the 170 MHz in the city of Paris (France). The covered urban area is very dense and it is characterized by irregular street grids, inhomogeneous built-up structures and street widths. All possible environments are encountered: narrow and large streets, squares, etc.

The architectures of the multilayer perceptron used are determined by the nature of the problem to be solved. The number of input neurons is set by the dimension of the input vector used. The output layer always consists in one single neuron, which provides the field strength value. The number of hidden units has been set to two layers of neurons.

There are presented two neural network based models in order to examine the potential benefits of the combination of physical and theoretical data. The first network is trained with physical data only including distance between transmitter and receiver, the angle of incidence and the description of the main obstacle on the path profile (represented by its relative position with regard to the transmitter: distance and altitude). The second network uses the same data except the information about the main obstacle that is replaced by results

of wave propagation algorithms, which consist of multiple diffraction calculus over knife-edges by the Deygout method [Deygout, 66]. The two neural network models have been processed and compared on three files representing the three distinct base stations. It appears that the neural network based hybrid system gives more accurate prediction than the other one; the improvement obtained in standard deviation is in the range [0.3, 1.3] dB. Furthermore, the neural based hybrid system is compared to classical prediction models: a semi-empirical built on the base of Deygout method and linear regression analysis, processed from the same propagation loss parameters, field strength measurements and terrain databases. The models are tested upon five urban files. The 13% gain on the standard deviation (0.8 dB) obtained by the neural network model is quite important in spite of the good values of standard deviation for the semi-empirical model.

In [Balandier, 95] are also presented two experiments on the base station antenna heights; the three base station antenna studied were set-up above rooftops. First of all, the neural network was built up considering one single base station then the network has been tested with patterns corresponding to the other base stations. The second experiment includes antenna height parameters as input to the neural network and the model is computed with several base stations. In this case, the average standard deviation obtained by the neural based hybrid system is 4.6 dB.

In [Gschwendtner, 96a,b] a hybrid modeling approach for the prediction of terrestrial wave propagation is presented, based on the Multilayer Perceptron trained with the backpropagation algorithm. The base station antenna height is used as a correction parameter. Therefore, the neural network consists of one input for the normalized antenna height, 1 hidden layer with three neurons and one output unit. The model is based on the field strength level measurements carried out in the city of Mannheim (Germany). The output unit is represented by the difference between predicted field strength values computed by COST231 Walfisch-Ikegami model [COST231, 99] and measured field strength values. It was obtained 1 dB gain with the neural model on the standard deviation.

In [Fraile, 97] it is presented a neural network based method for path loss prediction in urban environment. Its performance is compared to that of COST 231 models [COST231, 99], Walfisch-Bertoni [Walfisch, 88] model and Saunders-Bonar [Saunders, 91] model in terms of prediction error and computation time. The multilayer perceptron used has 41 input units, 4 hidden units and 1 output. The 41 inputs contain information about the building height profile and the distance between transmitter and receiver, while the output represents the estimate of the difference between the propagation loss and the free space loss. The neural network has

been trained and tested with measurements from a campaign carried out in Munich (Germany). The transmitter height was 13 m about ground and the measurements were taken along three different routes. One of these routes has been used as the training set for the neural network, while the others have been used for testing purposes. The values obtained for the standard deviation for the test routes are 6.3 dB and 6.6 dB, respectively. It is outlined that the neural model is at least four times faster than any of the other methods used for comparison.

In [Chang, 97] is presented a propagation model based on a radial basis function neural network that is capable of predicting the field strength based on topographical and morphographical. The radial basis function based propagation loss prediction model is trained by Okumura's field measurements taken in Tokyo area and the method is compared with Hata's formula. The neural network has 4 inputs, corresponding to frequency, base station antenna height, mobile station antenna height and distance between transmitter and receiver. The number of hidden nodes was chosen 30. The neural approach provides a uniform approximation to the propagation loss over a wide range of path profile. The mean-square error achieved by the radial basis prediction model for the test patterns is 2.35 dB and for the Hata's formula is 7.5 dB.

In [Fraile, 98] it is presented a neural network based method for path loss prediction in urban macrocellular environments. It is shown that this method is valid for all ranges of antenna height and for propagation distances up to 2 km. In [Fraile, 97] was showed a method for propagation loss prediction using a multilayer perceptron and the results were restricted to one measurement campaign carried out in Munich, at which the antenna was below the mean building height. In [Fraile, 98], the validity of the model is extended by using measurements taken from antenna sites placed over rooftop height, hence achieving a method for path loss prediction for all ranges of antenna height. It is also obtained a slight improvement in computation speed.

In [Leros, 98] typical backpropagation neural network with different inputs parameters are developed and used to evaluate and assess the relative importance of a set of radio propagation parameters for field strength prediction. This study develops a number of neural network models trained on an extended data set of propagation loss measurements taken in an urban area (Athens region) in the 900 MHz band. Each propagation loss measurement is associated with a set of parameters with environmental and topographic features. The neural model with the best performance is used to indicate the most important set of parameters for field strength prediction.

Each measurement record consists of the following set of parameters:

- Measured local mean value of propagation loss [dB];
- Horizontal coordinate [m] of the measurement point in the area digital map;
- Vertical coordinate [m] of the measurement point in the area digital map;
- Horizontal trace of distance [m] of the measurement point from the transmitting antenna;
- Altitude relative to the sea level [m] of the measurement point;
- Ground surface slope relative to the horizontal level ($0^\circ - 90^\circ$);
- Ground surface aspect (or azimuth) relative to the North ($0^\circ - 360^\circ$);
- Effective height [m], a derived parameter representing the local topographic features close to the receiver.

It is outlined that the most important parameters from the initial set of parameters for modeling the propagation path loss for an area with no diffraction loss (LOS conditions) and terrain morphology forming a positive effective height are the ground surface slope relative to the horizontal level, altitude relative to the sea level of the measurement point and horizontal trace of distance of the measurement point from the transmitting antenna.

In [Nešković, 98] is proposed a model based on the feedforward networks – multi-layer perceptron. The implementation of the model relies on two databases: terrain elevation and groundcover (“clutter”) database. Extensive field strength measurements were carried out in the wide area of Belgrade, Yugoslavia, a typical European town, in 450 MHz and 900 MHz frequency bands.

The presented neural network model has three groups of inputs (totally 14 inputs). The first group consists of an input only and it is the normalized distance between transmitter and receiver. The second group of inputs (4 inputs) is based on the terrain profile analysis. These inputs are: the portion through the terrain, the modified “clearance angle” factor for transmitter site, the modified “clearance angle” factor for receiver site and the rolling factor. The third group of input parameters (9 inputs) is based on the land category analysis along the straight line drawn between the transmitter and the receiver. There is a single input for each defined land usage category. The network has one output represented by the normalized electric field level. The architecture of the neural network is 14-17-7-3-1. After the testing phase, the RMS errors between measurements and the predicted values by the artificial neural network model are 5.9 dB (450 MHz) and 6.1 dB (900 MHz).

In [Bargallo, 98] it is presented a comparison of the performance of propagation models based on multilayer perceptron neural networks, radial basis function neural networks

and conventional multiple linear regression techniques. A number of methods for selection of training examples are also compared. It is shown that, in most cases, neural networks can provide an improvement in performance over conventional empirical models. The multilayer perceptron neural network is trained with Levenberg-Marquardt algorithm and the radial basis function neural network is trained with the orthogonal least squares algorithm. The following input parameters to the neural models were used:

1. Distance d between transmitter and receiver distance [m],
2. Diffraction Loss, D_1 [dB]: a combination of the Bullington and Epstein-Peterson methods is used to merge multiple knife edges found along the T – R terrain profile. The loss is then calculated using standard Fresnel diffraction theory.
3. Effective antenna height, [m]: is calculated using the Slope Algorithm. For the measured data set used in this study, the slope algorithm was found to provide the highest correlation between path loss and effective antenna height.
4. Land usage type at receiver locations: the terrain and land usage databases from which this information was extracted.

Inputs 1–3 are each fed into a different neural network input node after normalization to the continuous range [-1, 1]. Receiver location clutter codes, ranging from 1 to 10 are assigned 10 inputs in the neural network. The neural prediction models have a single output that represents predicted loss.

The results were generated based on a set of five test sites in the western Missouri/Kansas City area (denoted sites A to E). The area is characterized by rolling hills, in some cases with pronounced slopes. Sites A and C-E are located in suburban areas with 2-3 floors houses being most typical. Site B is located in a light urban section of town with a few medium size buildings evenly distributed across the area.

When the models are designed based on data measured at site A, the multilayer perceptron and the radial basis function prediction models achieve similar performances. However, the multilayer perceptron model required 8 hidden layer neurons, while the radial basis function model required 35 radial basis neurons. When the models are designed based on data measured at site B, the multilayer perceptron and radial basis function models have similar performance with improvement over the multiple linear regression model of about 1 dB in test root mean squared error. In the next example, the models are trained with the aggregated data sets from sites C, D and E and are tested with the full data set from site A. In this case, the use of neural prediction models allow for a reduction in test root mean square error as large as 1.7 dB with respect to the multiple linear regression model.

In [Yang, 99] are presented algorithms for extracting the values of relevant parameters from field measurements and 3-dimensional geographical data to be used in neural network modeling of wave propagation loss in micro-cells. The algorithms extracts the feature values from 3-dimensional elevation map and vector maps based on the theory of Computational Geometry. The input values of neural networks are quantities representing wave propagation environment between a transmitting antenna and a receiving antenna and equivalent to the variables of propagation models. The process of feature extraction produces the neural network input vectors. The output values become the predicted loss.

The neural network is composed of 10 input layer neurons, 2 hidden layers, 10 neurons in the first hidden layer and 5 in the second and 1 single output neuron (10-10-5-1). Training sample data set consists of samples generated from the field measurements of five 1,8 GHz PCS cell sites operating in Seoul.

The trained neural network model was used to produce the predicted values for 2 cell sites not included in the training and the mean squared errors were 88.2 dB and 70 dB. The COST 231 model, used for comparison, produced the prediction with the mean square error of 450 dB and 819.7 dB, respectively.

6.2. Indoor environment

In [Zhou, 95] a three layers feedforward is applied to model the indoor environment. The clusters analysis is employed to describe the complicated indoor propagation environments. The clusters are used as input patterns; the corresponding radio propagation distributions of indoor environments are used as output patterns of the network.

The model presented in [Wolfle, 97a] is based on Huygen's principle for the propagation of electromagnetic waves. The prediction model is described as follows: in a first step a grid of prediction points is generated from the database. The distance between the prediction points should be equidistant. Distances used for their prediction are in the range of 0.4 m to 0.6 m. Larger distances lead to less accuracy, smaller distances to a longer computation time. The field strength for each point is predicted by the field strength of its neighboring points using a NN.

In the second step, based on the known field strength at a central point, the field strength at the four neighboring points is predicted with a NN. After this initial prediction, the former neighbor pixel will become the new center pixel. All new neighboring points

except the old center pixel are predicted in this phase of the prediction. A feedforward perceptron was trained with the Resilient Backpropagation algorithm. 8 input parameters are used for the neural prediction model:

1. Field strength at the center point,
2. Distance between transmitter and receiver,
3. Visibility. LOS (Line-of-sight), OLOS (obstructed line of sight – same room but no line-of-sight) and NLOS (non line of sight – different rooms),
4. Orientation. Neighbor point nearer to the transmitter as central point or vice versa,
5. Shape of the room. Corridor, small room, hall,
6. Transmission: transmission loss of a wall between neighbor and center point,
7. Immunity. Improbability of time-variant effects,
8. Wave guiding. Number and distance of walls, parallel to the line between center and neighboring point.

The network contains 1 simple output representing the field strength in one of the neighboring points. Each pixel is predicted from its four neighboring pixels. From four different values of the field strength thus obtained, the maximum value is chosen. The four predictions from the neighboring points are not computed at the same time and if one or two predictions for a point have been made, this point can be a center point for new predictions. If the third prediction for this point is higher than the first or second prediction, the field strength for this point will be set to this higher field strength and all predictions, based on the old field strength for that point must be computed again. To reduce the number of iterations, the difference of the predicted field strength for the same pixel must be more than 1 dB to start a new iteration.

The prediction model is visualized by a tree structure. The root of the tree is the transmitter and for each of the four prediction directions a branch links with a further pixel. Prediction back to the center point is not possible, so only trees branches must be considered for all pixels except the transmitter.

The model presented in [Wolfe, 97a] is fast in computing the field strength and it is possible to include the wave guiding of corridors and also the propagation around corners. The dependence on the accuracy of the database is very small, because no coordinates of any reflection or diffraction point are necessary. Only the orientation and the materials of the walls are included in the prediction. This prediction model can also be calibrated by measurements because the neural network is trained with measurements done in different environments.

In [Wolfe, 97b], a neural network based model for the prediction of electric field strength inside buildings is presented and two algorithms for the selection of training patterns are presented and compared to each other.

The described prediction model is an extension of the empirical models, because it is based on the direct ray between transmitter and receiver. It is possible to predict the field strength for a single point or an area. In contrast to the model described in [Wolfe, 97a], each point in the area is predicted as a single point and the prediction is independent from the neighboring points. This leads to a very small computation time and very fast algorithms. The output of the NN represents the field strength at the prediction point and the parameters of the prediction model can be subdivided into four groups:

1. General parameters
 - a) Free space attenuation (distance, frequency)
 - b) Visibility (LOS, NLOS, OLOS)
2. Influence of the walls between transmitter and receiver
 - a) Transmission loss of the direct ray
 - b) Wave guiding effect of the walls
3. Local arrangement of the walls at the transmitter site
 - a) Local reflectors
 - b) Local shielding effects
4. Influence of the location at the receiver site
 - a) Local reflectors
 - b) Local shielding effects
 - c) Shape of the room of the receiver
 - d) Size of the room at the receiver

Different structures for the neural network have been examined and the Multilayer Perceptron leads to the best results. The neural network is trained with the Backpropagation Algorithm. When the prediction is done on measurements used to train the network, a mean error of 0.2 dB and a standard deviation of 3.5 dB are obtained. For measurements and the environment that were not used for training, the mean error of the prediction is 7.7 dB and the standard deviation is 8 dB.

In [Wolfe, 97c] dominant paths for the wave propagation are computed and a prediction of the field strength is obtained with neural networks, trained with measurements in different buildings. This model has been developed for buildings with very long corridors, as often found in office buildings. A feedforward perceptron was trained with the standard

Backpropagation algorithm. The network is subdivided into three parts, each part representing one of the dominant paths, and leads to the most accurate prediction.

Input parameters of the neural network are transmission loss, path length, accumulated angle of the path [Wolfe, 97c], [Wolfe, 97d] and the wave guiding effects along the dominant paths.

An improvement of the prediction is possible, if different dominant paths are determined for each prediction point. For the determination of different paths, the criterion for the selection of incoming direction of first order must be altered. So it is possible, to assign different weights to the parameters transmission loss, length and accumulated angle. Best results were obtained with simultaneous consideration of three different paths: shortest path, path with lowest transmission loss and path with minimal angle.

In [Wolfe, 97d] is presented a model based on the determination of the dominant paths between the transmitter and the receiver. The field strength is predicted with artificial neural networks, trained with measurements. The best results were obtained with the following input parameters:

1. General parameters
 - a) Free space attenuation along the path (includes distance and frequency)
 - b) Visibility (LOS, OLOS, NLOS)
 - c) Accumulated transition angle
2. Influence of walls along the paths
 - a) Transmission loss along the path
 - b) Wave guiding along the path (the wave guiding effect due to walls oriented more or less parallel to the path)
3. Local arrangements of the walls at the transmitter site
 - a) Local reflectors (local directivity due to the arrangement of walls)
 - b) Local shielding effects (local shielding influenced by walls)
4. Influence of the receiver site
 - a) Local reflectors
 - b) Local shielding effects
 - c) Shape of the room containing the receiver
 - d) Size of the room containing the receiver

These parameters are described in [Wolfe, 97b].

The training patterns of a multilayered feedforward perceptron, trained with the backpropagation algorithm, are obtained from measurements at the University of Stuttgart. The difference between the measurements and the prediction for a transmitter location (not used during the training process of the network) is small and the mean error of the magnitude is smaller than 3.3 dB. The field strength prediction within a building of the Technical University of Viena gives a mean error smaller than 8 dB.

In [Wolfe, 98] is presented an algorithm for the determination of the dominant paths for indoor wave propagation. Based on these dominant paths, three different prediction models are presented and compared with one another and with measurements. Two of the models are based on neural networks and the third model is an empirical model.

For the computation of the field strength with neural networks, the parameters of the minimum-loss dominant path are determined. Good results have been obtained with the following parameters:

1. Free space attenuation along the path,
2. Transmission loss,
3. Interaction loss,
4. Wave guiding along the path. Each dominant path represents different rays. All of them are guided by reflections at the walls or by diffractions at the corners of wedges in the same direction. To include all these rays in the prediction, they are described by a parameter wave guiding. The wave guiding of the wall depends on their material (reflection loss), their orientation (reflection angle) and the distance between the walls and the path. These three parameters are combined to give the parameter wave guiding as described in [Wolfe, 97c].
5. Local reflectors and shielding effects at the transmitter site.
6. Local reflectors and shielding effects at the receiver site.

For each dominant path, all mentioned parameters are gained from a vector-oriented database and are normalized for the input of the neural network. The prediction of the field strength is based only on these parameters.

The backpropagation algorithm was used to train the neural network, and the measurements were taken at the University of Stuttgart.

In the second neural network model presented in [Wolfe, 98] besides the minimum-loss dominant path (MLDP) are considered also the alternative dominant paths (ADP). The structure of the neural network is similar but four additional parameters are taken into account for the description of the alternative paths. These additional inputs are represented by the

normalized difference of the parameters (path length, transmission loss, interaction loss and wave guiding) between the alternative and the minimum-loss dominant path.

In contrast to the model described in [Wolfe, 97d], the relative and not the absolute parameters of the alternative paths are used. The mean errors of the prediction made by the two neural models are -0.4 dB and 0.2 dB, with standard deviations of 3.3 dB and 1.6 dB.

In [Wolfe, 99] two different approaches to the computation of the field strength are presented: empirical computation and neural networks. The parameters of the dominant path, that are path length, accumulate penetration loss of the walls passed, loss due to changes in the direction of propagation and wave guiding gain) represent the input values of the neural network and the measured field strength is the output value. Both approaches require many measurement values from different buildings for the calibration of the models.

The model presented in [Neskovic, 00b] has the form of a Multilayer Perceptron with three hidden layers (12 inputs and 1 output). The implementation of the model requires a database of the floor plan in which all particular locations are classified into several environmental categories, for example wall, corridor, classroom, window, etc. The way proposed by the authors to do that, is to make a color picture over the scanned floor plan. There are several inputs based on the number of previously defined environmental categories. One of the inputs represents the normalized distance from the transmitter. The remaining inputs are based on the analysis of the straight line drawn between the transmitter and the receiver with respect to the environmental categories, e. g. how many doors, what percentage of the line passes through the classroom, etc. The proposed model has a single output that is the normalized electric field level. The average results of the prediction obtained in two buildings are: mean errors of 1.52 dB and 2.3 dB, standard deviations of 6.5 dB and 5.24 dB.

7. Proposed neural network models

In the next sections several proposed neural network models for the prediction of propagation path loss in different environments (urban, suburban and indoor) are investigated. In the presented studies the multilayer feedforward networks, commonly referred to as Multilayer Perceptron (MLP) and Radial Basis Function (RBF) Neural Networks (NN), were used. Specifically, the following NN are proposed:

- MLP-NN propagation prediction models for outdoor and indoor environment;
- MLP-NN hybrid models for outdoor environment;
- RBF-NN propagation prediction models for outdoor and indoor environment;
- RBF-NN hybrid models for outdoor environment.

In section 7.1 is given a short description of the measurements used to design the neural models presented in sections 7.2 and 7.3. In section 7.2.1 is also presented a proposed neural network model for the implementation of Hata's formula and the knife-edge diffraction.

7.1. The measurements

Field strength measurements used to design and to test the neural network models presented in the next sections, were performed at 1890 MHz frequency band in the city of Kavala (Greece), in Oia village on Santorini Island (Greece) and at the Hellenic Telecommunication Organization premises in Athens (Greece). A detailed description of the measurement procedure can be found in [Kanas, 99] and [Papadakis, 98].

In case of the measurements collected in the urban and suburban environment, respectively, the fast fluctuations effects were eliminated by averaging the measured received power over a distance of 6 m that corresponds to approximately 40λ sliding window. After converting the values from received power to path loss versus distance, the measured path loss is compared to the predicted path loss by the proposed neural network models and the empirical models based on the absolute mean error (μ), standard deviation (σ) and root mean square error (RMS). The absolute error between the measured and predicted path loss is computed with:

$$E_i = \left| PL_i^{\text{measured}} - PL_i^{\text{predicted}} \right| \quad (7.1)$$

where i represents the number of the measured sample. The absolute mean error is computed by:

$$\mu = \frac{1}{N} \sum_{i=1}^N E_i \quad (7.2)$$

where N is the total number of measured samples. The standard deviation is determined from the absolute error (7.1) and the mean absolute error (7.2):

$$\sigma = \sqrt{\frac{1}{N-1} \left(\sum_{i=1}^N E_i^2 - N \cdot \mu^2 \right)} \quad (7.3)$$

The RMS error is given by:

$$\text{RMS} = \sqrt{\mu^2 + \sigma^2} \quad (7.4)$$

The neural network models proposed for propagation prediction in outdoor environments and the examined empirical models require parameters that describe the propagation environment such as the street width, the roof top height and the building block spacing. Average values were used since these variables change continuously along one route. For the determination of these geometric parameters a map with building database was used [Kanas, 99].

The measurements conducted in the 1890 MHz frequency band, at the Hellenic Telecommunication Organization premises are following different scenarios. Each floor of the building consists of a circular sector of 60 m in circumference located at the center of each floor and 3 branches, denoted A, B and C, departing from the circular sector, where at each branch there are one main long corridor and two short back corridors with offices flanked on both sides of corridors, as shown in Figure 7.1 [Papadakis, 98].

Offices are in consecutive order and are separated by soft partitions. Measurements were done along the corridors and inside the offices, in all three branches. In every position of the receiver inside the offices about 10000 samples of the received power were recorded while the receiving antenna was rotating. The transmitting antenna was located always in the same sector of the eleventh floor in two different sites (position: 1 or 2 in Figure 7.1). The base station antenna heights used were 2.2m, 2.6m and 2.7m. The measurements were performed using two different types of transmitting antenna: OMNI and directional. The receiving antenna was always an OMNI antenna.

In case of indoor environment, the performance of the neural network models is evaluated by making a comparison between predicted and measured values based on the absolute mean error, standard deviation and root mean squared error, as described by equations 7.2-7.4.

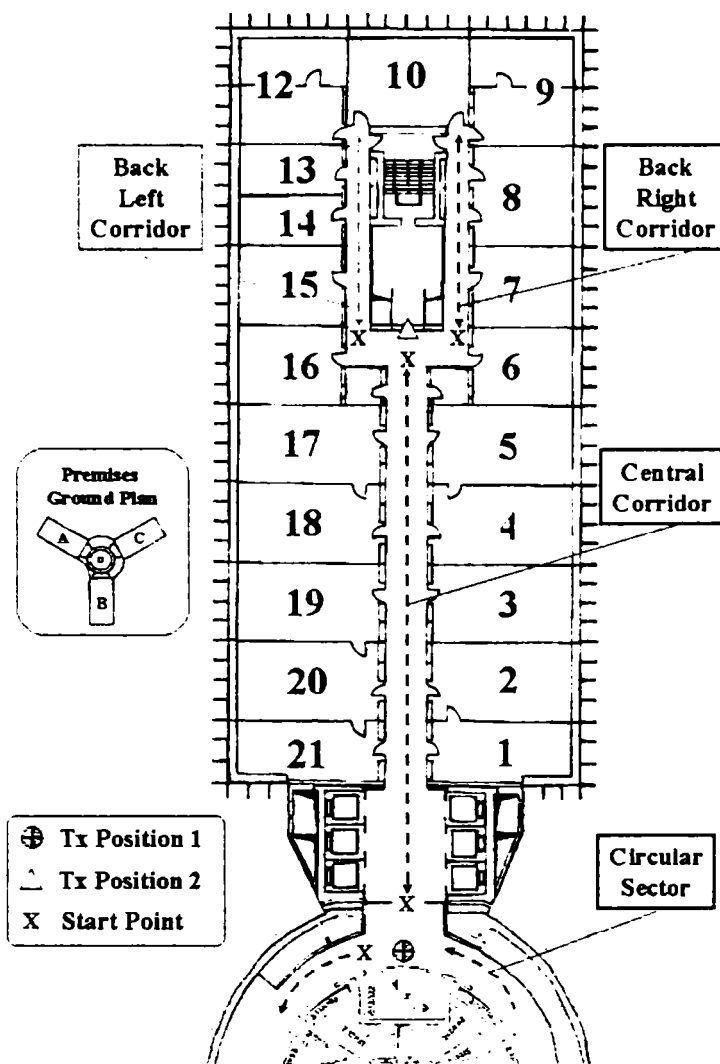


Figure 7.1. The building topology and the transmitter positions

7.2. Proposed MLP-NN models for the prediction of propagation path loss

In this section the performance of MLP neural networks for the prediction of propagation path loss in different environments are presented. A detailed description of this type of neural network is given in chapter 2.

The learning phase of the MLP neural network proceeds by adaptively adjusting the free parameters of the system based on the mean squared error between predicted and measured path loss for a set of appropriately selected training examples. When the MSE between the MLP neural network output and the desired output is minimized, the learning process is terminated and the neural network can be used in the testing phase with test vectors. At this stage, the neural network is described by the optimal weight configuration, which ensures the output error minimization.

The goal of the prediction is not only to produce small errors for the set of training examples but also to be able to perform well with examples not used in the training process. This generalization property is very important in this practical prediction situation where the intention is to use the propagation prediction model to determine the coverage area of potential transmitter locations for which no or limited measured data are available.

The Universal Approximation Theorem states that a single hidden layer is sufficient for a MLP to compute a uniform approximation to a given training set represented by the set of inputs and a desired output. However, the theorem does not say that a single hidden layer is optimum in the sense of learning time, ease of implementation or (more importantly) generalization [Haykin, 99].

It is very difficult to know which training algorithm will be faster and a good generalization will be obtained for a given problem. It depends on many factors, including the complexity of the problem, the number of data points in the training set, the number of weights and biases in the network, the error goal and on the task performed by the neural network (function approximation in the presented work).

The selection of the set of training examples is very important in order to achieve good generalization properties [Wolfe, 97] [Haykin, 94]. The set of all available data is separated in two disjoint sets that are training set and test set. The test set is not involved in the learning phase of the networks and it is used to evaluate the performance of the neural model. One of the problems that may occur during neural network training is overfitting. The error on the training set is driven to a very small value but when the new data is presented to the network, a large error is obtained and the network has not learned to generalize to new situations. It is difficult to know beforehand how large a network should be for a specific application. In this work, in order to improve generalization and due to the fact that the dimension of the training data is much smaller than the dimension of the test data, the early stopping method is used. However, it was investigated the performance of some algorithms (on the same training set) without the use of early stopping method.

Neural network training can be more efficient if certain preprocessing steps are performed on the network inputs and desired outputs. Two approaches are used in this work:

- The inputs and the desired outputs are scaled so that they fall in the range $[-1, 1]$.
- The inputs and the desired outputs are scaled so that they will have zero mean and unity standard deviation.

In section 7.2.2 – 7.2.5 a number of comparisons are made for MLP neural networks with different architectures and different training algorithms. At a first step, the performance of the

MLP-NN trained with Levenberg-Marquardt (LM) algorithm, with different numbers of neurons in one and two hidden layer was investigated. These simulations were done with the use of early stopping method. Following these simulations, the MLP-NN with the optimum configuration is established and investigations are done on the performance of two other training algorithms: the Resilient Backpropagation (RP) and the Powell-Beale (PB) version of the conjugate gradient algorithm.

7.2.1. MLP implementation of Hata's formula and knife-edge-diffraction model

This section presents a neural network implementation for path loss prediction models [Popescu, 00a]. The models implemented are Okumura-Hata formula for path loss prediction in urban area, COST231-Hata model [COST231, 99] for path loss prediction in suburban areas and knife-edge diffraction model.

A MLP with one hidden layer, trained with the Levenberg-Marquardt algorithm was used for these simulations. As a first step, the Hata's formula for urban area is simulated. The neural network has four inputs corresponding to the four parameters:

- frequency (f),
- base station antenna height (h_{BS}),
- mobile station antenna height (h_M),
- distance between transmitter and receiver (d)

and ten units in one single hidden layer (4-10-1). The output of the neural network represents the normalized path loss. Free space field strength was not separated in this experiment. As a quality criterion the mean square error is defined as follows:

$$E = \frac{\frac{1}{N} \sum_{i=1}^N (L_{p,i} - L_{p,net,i})^2}{(L_{max} - L_{min})^2} \quad (7.5)$$

that is to be minimized during the training process. $L_{p,i}$ represents the path loss computed according to Hata formula and $L_{p,net,i}$ represents the path loss derived from the network output. For comparison with other methods the mean error (μ) and the standard deviation (σ), as described by equations 7.2 and 7.3, are used.

The number of the training points in the 4-D parameter space is 2162. Testing the neural network with a test pattern that has not been presented to the network during the training phase ($f = 922\text{MHz}$, $h_{BS} = 220\text{ m}$ and $h_M = 3\text{m}$) it was obtained a mean error of 0.34 dB and a standard deviation of 0.1 dB.

In order to simulate the COST231 - Hata model it was used the same configuration for the neural network. The number of training pattern used for training was 722. For a test example the obtained values for mean error and standard deviation were 0.14 dB and 0.44 dB, respectively.

The configuration of the neural network used to simulate the diffraction losses due to wedge obstruction is the following: 6 inputs units representing

- distance between transmitter and diffraction point,
- distance between diffraction point and receiver,
- height of the obstruction wedge,
- height of the transmitter antenna,
- receiver antenna height,
- wavelength.

By taking into account different sites and heights for the knife-edge obstruction, different frequencies and antenna heights, the training set consists in 2720 patterns. The simulated neural network has 12 neurons in a single hidden layer.

The tests were done using a test set containing a height of the knife-edge diffraction that was not presented to the network in the training phase. The number of the tested patterns is 1991, due to the sampling rate used for the distance between the diffraction point and receiver.

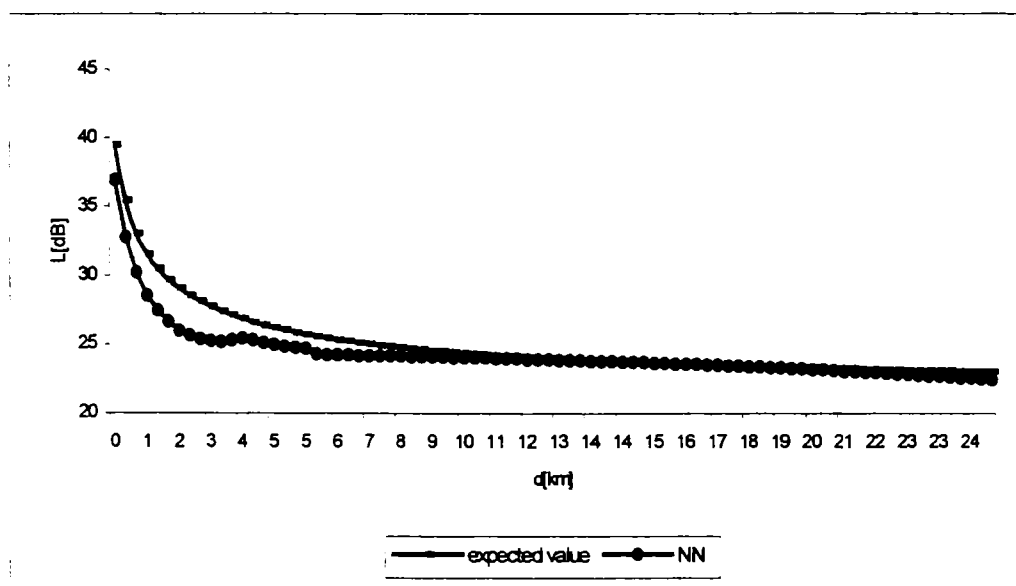


Figure 7.2. Comparison between predicted values achieved by neural network and knife-edge diffraction model versus the distance between the diffraction point and receiver

Figure 7.2 shows the comparison between the path loss value computed according to the knife-edge diffraction model and the predicted value derived from the neural network

output. when the height of the obstruction point is 90 m. the frequency is 1200 MHz and the base station antenna height is 20 m. That is a situation that was not used for training. The mean error obtained in this case is 0.69 dB and the standard deviation is 0.91 dB, respectively.

7.2.2. Proposed MLP-NN models for propagation prediction in urban environment

In this section are presented MLP-NN based models for the prediction of propagation path loss in urban environments (city of Kavala). The neural networks are designed separately for line-of-sight (LOS) and non-line-of-sight (NLOS) cases. The applications of neural network are considered as function approximation problems consisting of a nonlinear mapping from a set of input variables containing information about potential receiver onto a single output variable representing the predicted path loss.

The performance of the proposed MLP-NN models is compared to that of the COST231-Walfisch-Ikegami model (CWI) [COST 231, 99], the Walfisch-Bertoni model (WB) [Walfisch, 88] and the single slope model (SSM) [Rappaport, 96], based on the absolute mean error, standard deviation and the root mean square error between predicted and measured values.

A. LOS case

For the LOS case, the neural network model is trained with physical data, as follows:

- distance between transmitter and receiver,
- width of the streets,
- height of the buildings,
- building separation,
- position of the transmit antenna with respect to the rooftop.

The model has a single output which represents the normalized propagation path loss. Each training example consists of target values of measured path loss (normalized to the range [0,1] and the corresponding input parameter values for the measurement location.

A set of 1013 examples is used for training (out of which, 63 examples are used for validation) while 2026 patterns were used to test the neural models. Different architectures of the MLP were studied using the Levenberg-Marquardt (LM) algorithm. In the case of two hidden layers, only the case with an equal number of neurons in both layers was investigated.

In figure 7.3 is shown the performance of the MLP-NN trained with LM algorithm, for different number of neurons in one and two hidden layers.

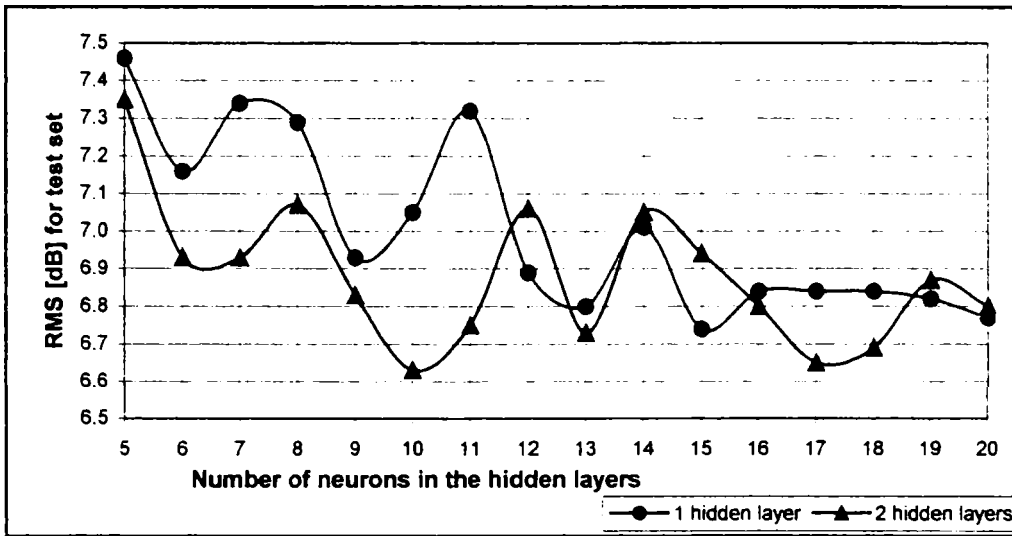


Figure 7.3. Performance of MLP-NN on the test set, with different number of neurons in 1 and 2 hidden layers, 5 inputs, LOS case, urban environment

As it can be seen from Figure 7.3 the smallest value for the RMS was obtained in case of a MLP with 10 neurons in 2 hidden layers (6.63 dB). Further investigations were done on the 5-10-10-1 MLP-NN's configuration with RP and PB algorithms. Due to the fact that the RMS errors for the test set obtained by using the early stopping method were big (9.45 dB in case of RP and 8.12 dB for PB), the simulations were done without the use of early stopping method, for different number of epochs. The results are presented in Figure 7.4.

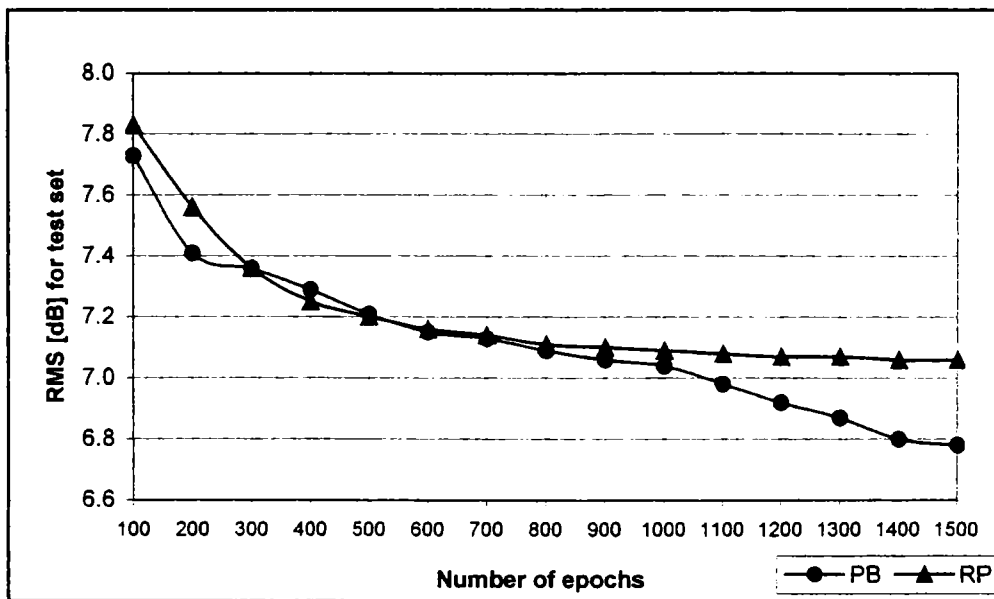


Figure 7.4. Performance of MLP-NN on the test set, with RP and PB training algorithms, 5-10-10-1, LOS case, urban environment

The results obtained in the case of the neural model designed with two hidden layers with 13 neurons each are presented in [Popescu, 00b]. As it can be seen from Figures 7.3-7.4, the best performance is achieved with a neural model with two hidden layers with 10 neurons each, trained with LM algorithm [Popescu, 01a] and the results obtained in this case are presented below. A test set of 2026 examples is used to compare the proposed MLP-NN model to the Walfisch-Bertoni model (WB), the single slope model (SSM) and the COST 231–Walfisch-Ikegami model (CWI). Table 7.1 represents the performance achieved by each of the above-mentioned models for the entire test data.

Table 7.1. Comparison between the proposed MLP-NN model and the other empirical models, LOS case

	MLP-NN	SSM	CWI	WB
μ [dB]	5.08	5.43	7.04	9.09
σ [dB]	4.23	4.80	4.06	4.51
RMS [dB]	6.63	7.24	8.23	10.24

Figure 7.5 represents the measured and predicted propagation path loss by the proposed MLP-NN model, the single slope model and the Walfisch-Bertoni model, for a specific route, in the LOS case. For this route, the results are presented in Table 7.2. As it is shown in Table 7.2, making a comparison with the single slope model, the gain obtained by the neural model lays around 3.69 dB for the mean error, 1.77 dB for the standard deviation and 4.06 dB for the RMS error.

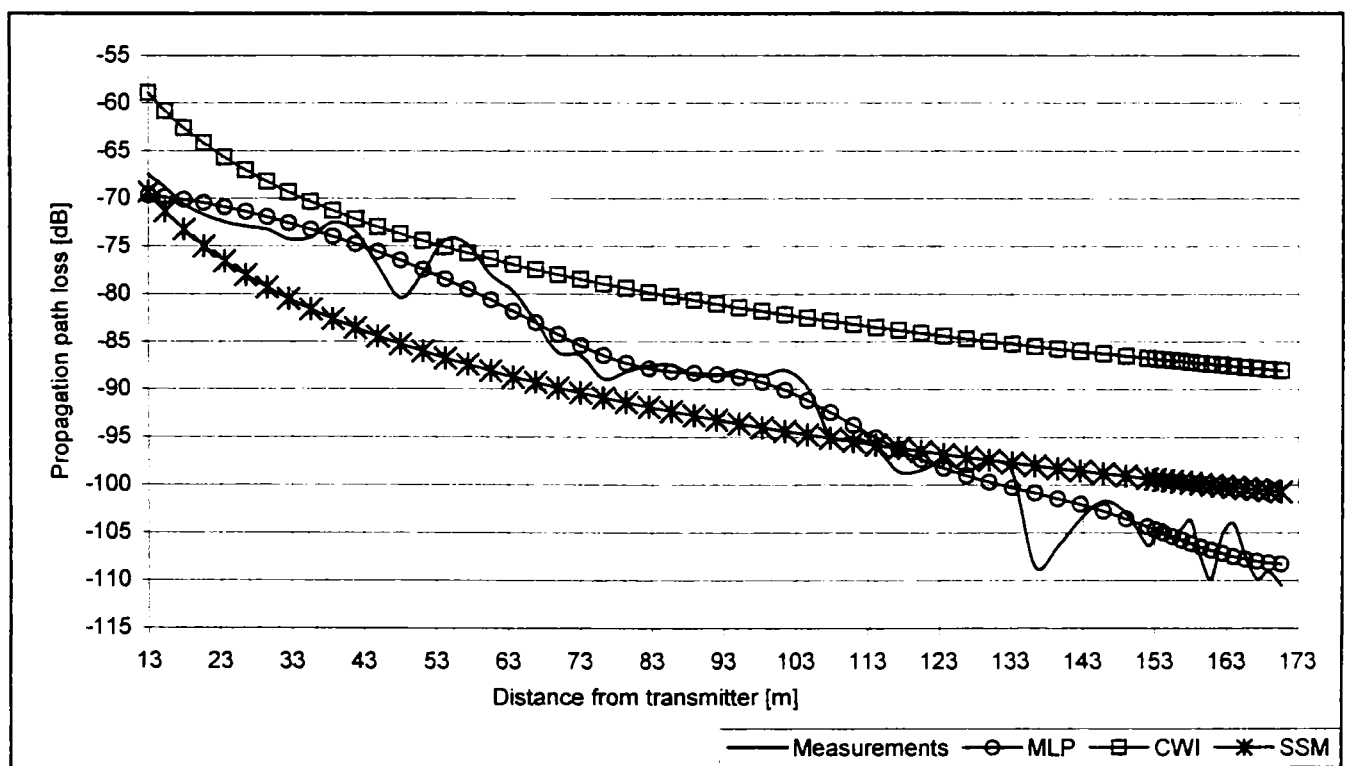


Figure 7.5. Measured and predicted path loss for LOS case in urban environment.

Table 7.2. Comparison between the prediction models in LOS case, for a particular route

	MLP-NN	SSM	CWI
μ [dB]	1.65	5.34	11.20
σ [dB]	1.35	3.12	6.51
RMS [dB]	2.13	6.19	12.95

Between the empirical algorithms, in LOS case, the single slope model achieves the best performance. However, this model is based on the distance between transmitter and receiver, the frequency and the propagation factor. It was found that in the LOS paths the power decay factor ranges from minimum 1.56 up to 3.05. The better performance achieved by the neural network model is due to the various inputs parameters used to predict the propagation path loss and due to the generalization properties of the network

B. NLOS case

The study is based on a MLP-NN model trained with physical data that includes:

- distance between transmitter and receiver,
- width of the streets,
- height of the buildings,
- building separation,
- street orientation,
- position of the transmit antenna with respect to the rooftop.

The MLP-NN model has a single output which represents the normalized propagation path loss. A set of 394 examples is used for training, a set of 26 examples is used for validation and for testing purpose a set of 1680 examples is used.

In Figure 7.6 the performance of the MLP-NN over the entire test set is shown for different numbers of neurons in 1 and 2 hidden layers, using the LM training algorithm.

Due to the fact that the value obtained for RMS obtained for the test is above 6 dB, in addition to the above-mentioned data further investigations were done by including in simulations some parameters used in the propagation path loss computation by COST231-Walfisch-Ikegami algorithm:

- free space loss L_0 ,
- roof-to-street diffraction, L_{rts} ,
- multiple screen diffraction loss, L_{msd} ,
- loss factors due to the street orientation, L_{ori} ,

- loss factor due to the difference between base station antenna height and roof top height, L_{bsh} .

The formulas used for the computation of these parameters are presented in section 5.2.5.

By using these parameters as additional inputs to the MLP the potential benefits of the combination of physical parameters and a theoretical model for this specific environment is studied.

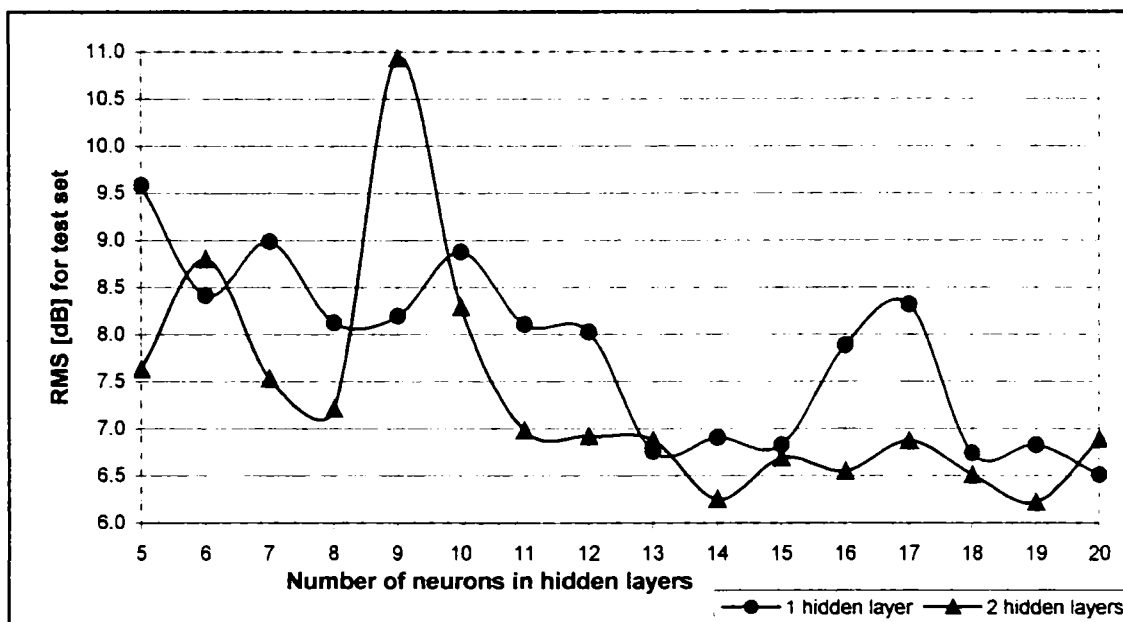


Figure 7.6. Performance of MLP-NN on the test set, with different number of neurons in 1 and 2 hidden layers, 6 inputs, NLOS case, urban environment

The performance of the MLP-NN over the test set, having 11 inputs, with one and two hidden layers, trained with LM algorithm, is presented in Figure 7.7.

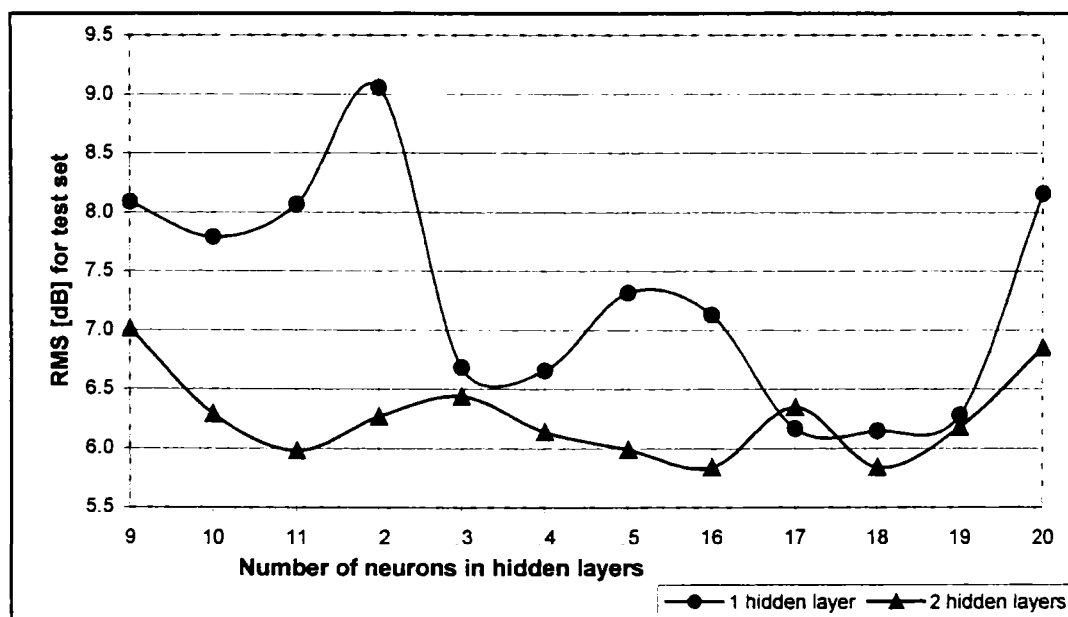


Figure 7.7. Performance of MLP-NN on the test set, LM algorithm, 11 inputs, different number of neurons in 1 and 2 hidden layers, NLOS case, urban environment

As it can be seen from the above figure, the “best” prediction is made by the MLP-NN with 11 inputs and two hidden layers of 18 [Popescu, 01a] or 16 neurons each, trained with LM algorithm, with an RMS error of 5.84 dB for the entire test set. Further investigations for RP and PB algorithms are done on the configuration of MLP with 16 neurons in 2 hidden layers, for 6 inputs and 11 inputs. The RMS errors on the entire test set obtained by using the early stopping method were unsatisfactory and the simulations were done for different number of epochs. The performance of the MLP on the entire test set having the following configurations: 6-16-16-1 and 11-16-1-6-1 are presented in Figure 7.8 and Figure 7.9, respectively.

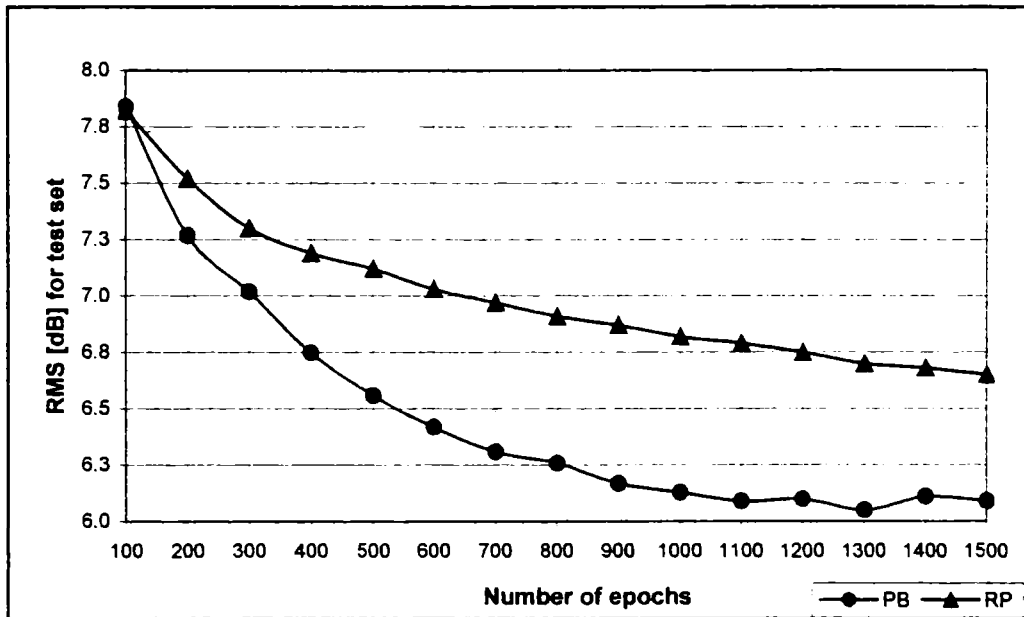


Figure 7.8. Performance of MLP-NN on the test set, with RP and PB training algorithms, 6-16-16-1, NLOS case, urban environment

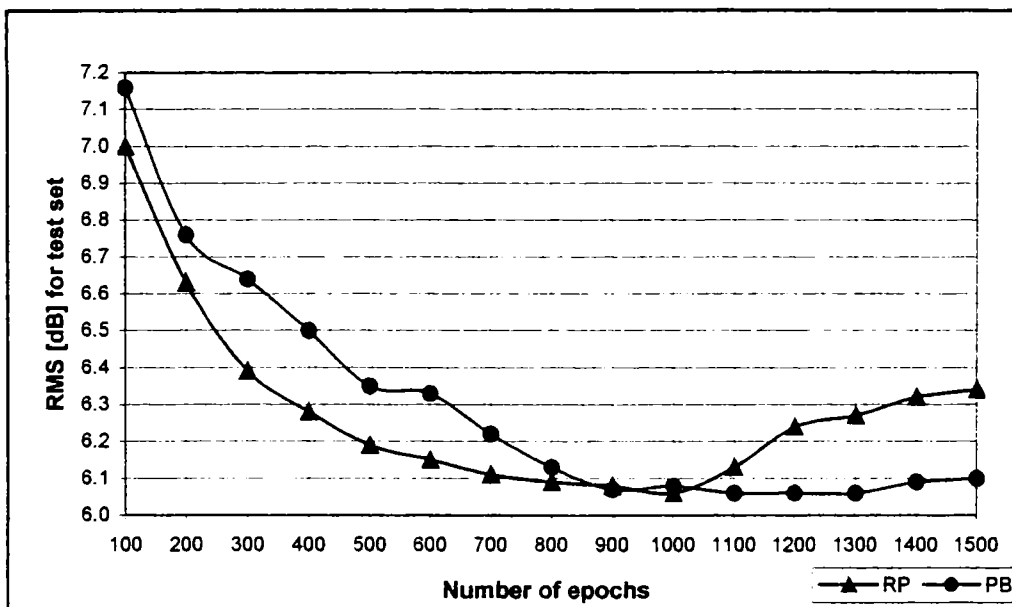


Figure 7.9. Performance of MLP-NN on the test set, RP and PB training algorithms, 11-16-16-1, NLOS case, urban environment

Table 7.3 represents the performances achieved by two MLP models, having 2 hidden layers with 16 neurons each, MLP-NN1 (with 6 inputs), MLP-NN2 (with 11 inputs) and the performance obtained by the single slope model (SSM), the Walfisch-Bertoni model (WB) and the COST 231-Walfisch-Ikegami model (CWI) for the entire test data.

Table 7.3. Comparison between the MLP-NN approach and the other empirical models in NLOS case

	MLP-NN2	MLP-NN1	SSM	WB	CWI
μ [dB]	4.01	4.61	6.35	6.08	6.96
σ [dB]	4.25	4.65	4.37	4.14	4.62
RMS [dB]	5.84	6.55	7.75	7.40	8.38

Figure 7.10 represents the measured and predicted propagation path loss by the MLP-NN models and the single slope model (SSM) for one route characterized by a base station antenna located below rooftop (NLOS case).

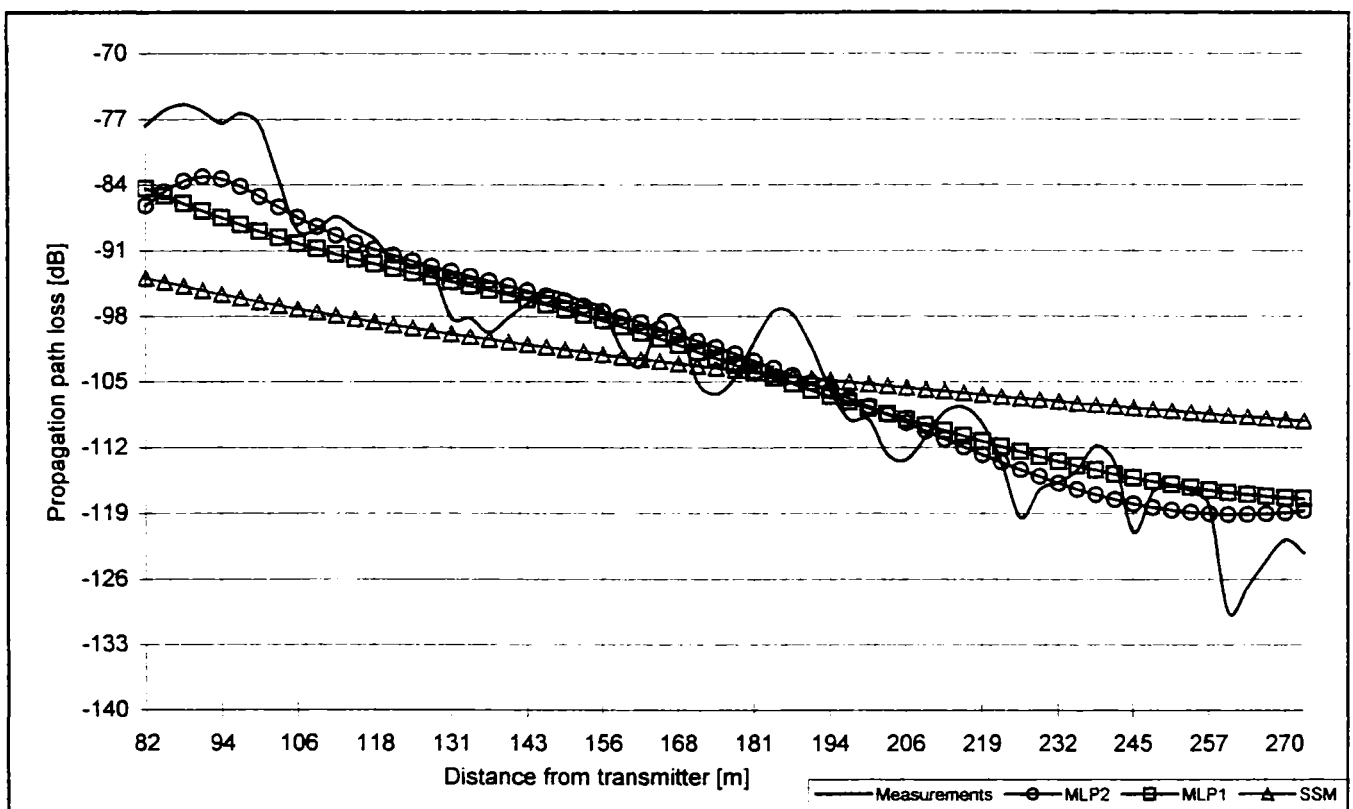


Figure 7.10. Measured and predicted path loss for NLOS case with $h_{BS} < h_{roof}$ urban case

Figure 7.11 represents the measured and predicted propagation path loss by the Multilayer Perceptron models and single slope model for one route characterized by a base station antenna placed above rooftop (NLOS case).

The performances of the models, in the case of a particular route are shown in table 7.4.

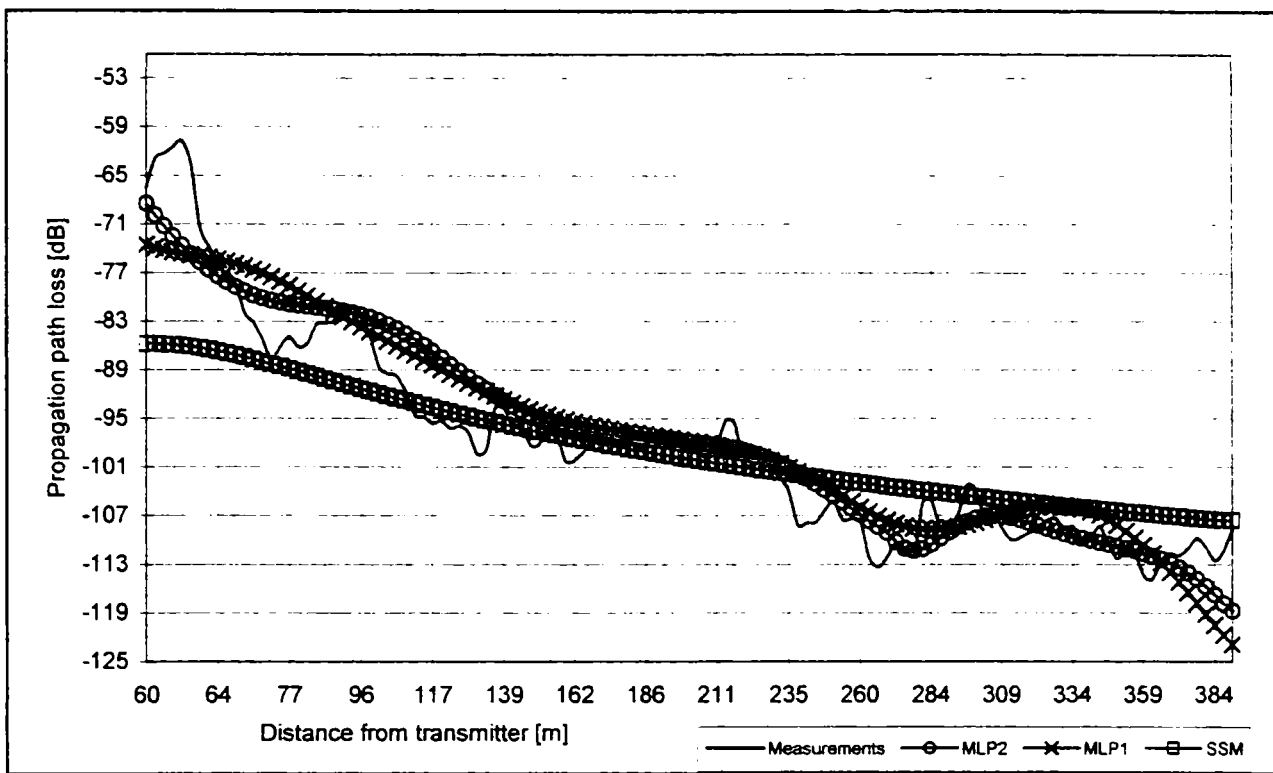


Figure 7.11. Measured and predicted path loss for NLOS condition with $h_{BS} > h_{roof}$ urban case

Table 7.4. MLP-NN models and other empirical models in NLOS case, for $h_{BS} > h_{roof}$ and $h_{BS} < h_{roof}$

	μ [dB]		σ [dB]		RMS [dB]	
	$h_{BS} < h_{roof}$	$h_{BS} > h_{roof}$	$h_{BS} < h_{roof}$	$h_{BS} > h_{roof}$	$h_{BS} < h_{roof}$	$h_{BS} > h_{roof}$
MLP-NN2	3.34	3.08	2.60	2.88	4.24	4.22
MLP-NN1	3.90	3.81	3.26	3.11	5.09	4.92
SSM	8.05	4.69	5.67	5.05	9.85	6.89

In the neural network model MLP-NN2, the training patterns combine information about the physical propagation conditions (transmitter and receiver locations) and about the diffraction losses that characterize the path between transmitter and receiver. The simple neural model MLP-NN1 gives good results but the results obtained by the system that combines a nonlinear neural network approach, propagation loss algorithms and physical data gives more accurate prediction.

7.2.3. Proposed MLP-NN models for propagation prediction in suburban environment

In this section are presented neural network based models for the prediction of propagation path loss in suburban environment for in NLOS case [Popescu, 01b]. The performance of the neural model is compared to the measured path loss values from the measurements conducted in Oia village on Santorini Island, Greece, based on the absolute

mean error, standard deviation and the root mean squared error between predicted and measured values. In addition, the measured values were compared with the results obtained by Walfisch-Bertoni model (WB) [Walfisch, 88], COST231-Walfisch-Ikegami model (CWI) [COST 231, 99] and Single Slope model (SSM) [Rappaport, 96].

At the first step, the study is based on a MLP-NN trained with physical data that includes:

- distance between transmitter and receiver,
- width of the streets,
- height of the buildings,
- building separation,
- street orientation.

A set of 333 measurement points (out of which 21 examples is used for validation purpose) is used for training and a set of 1332 measurement points is used for testing purpose. The neural model has a single output which represents the normalized propagation path loss.

In Figure 7.12 the performance of the MLP-NN on the test set is represented for different numbers of neurons in one and two hidden layer, using the LM training algorithm and the early stopping method.

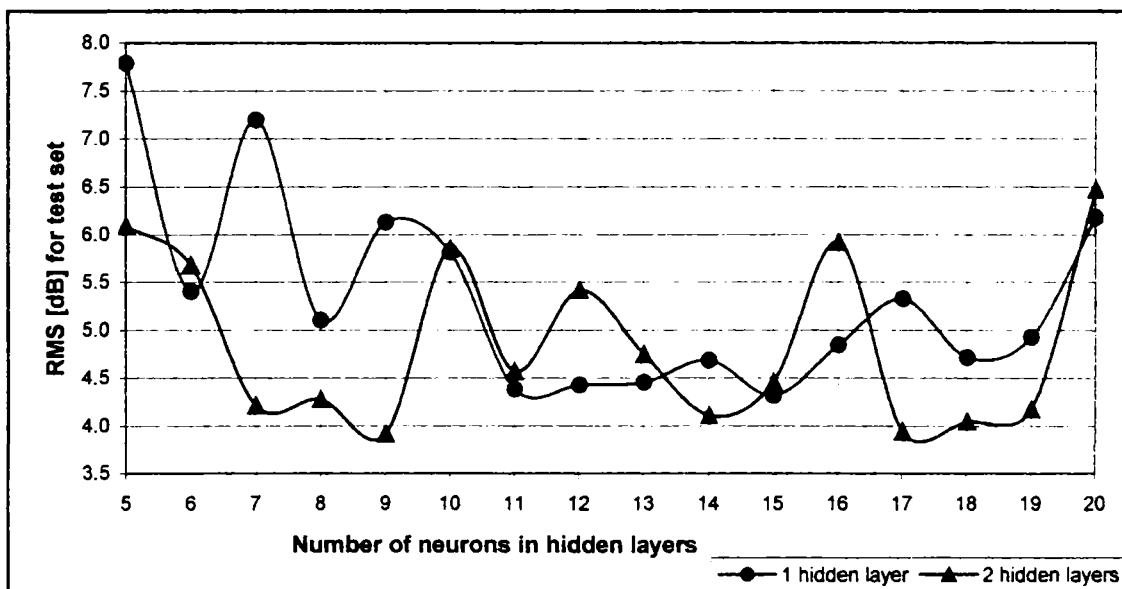


Figure 7.12. Performance of MLP-NN on the test set, LM algorithm, different number of neurons in 1 and 2 hidden layers, 5 inputs, suburban environment

In order to investigate the effect of the combination between physical and theoretical parameters used to train the MLP-NN, two more inputs were added:

- multi-screen diffraction loss, L_{msd} ,
- roof-to-screen diffraction loss, L_{rts} ,

computed by COST 231-Walfisch-Ikegami algorithm, as presented in section 5.2.5.

The performance of the MLP on the entire test set, having 7 inputs with one and two hidden layers is presented in Figure 7.13.

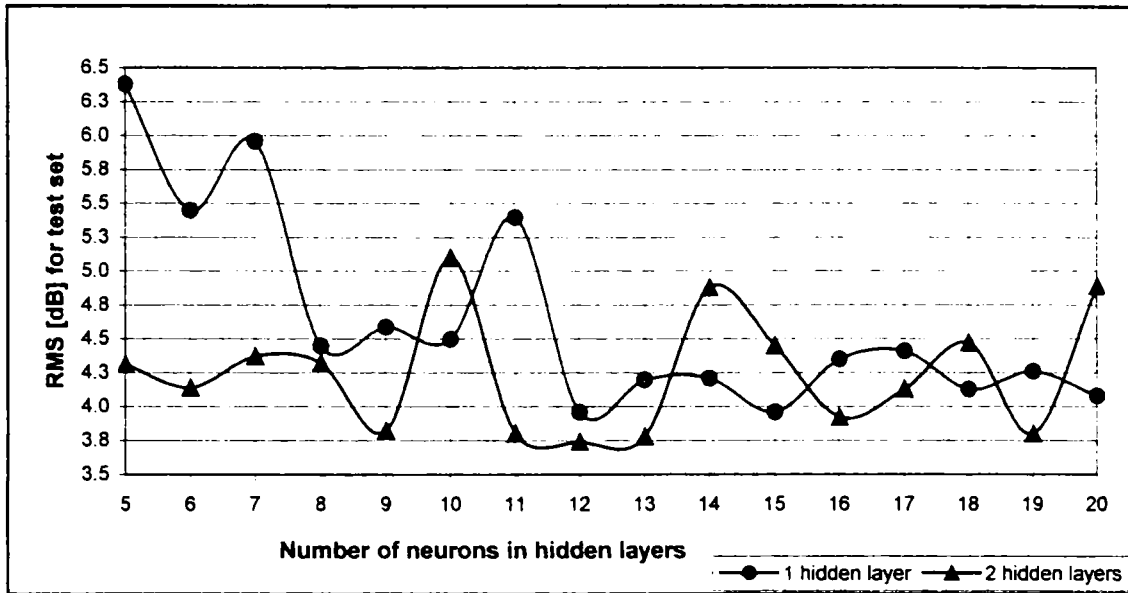


Figure 7.13. Performance of MLP for LM training algorithm, different number of neurons in 1 and 2 hidden layers, 7 inputs, suburban environment

As it can be seen from the above figure, a good prediction is made by the MLP-NN with 7 inputs and 2 hidden layer having 12 neurons each, with an RMS error of 3.74 dB for the entire test set. Using this configuration of the MLP-NN, further investigations for RP and PB algorithms are done for different number of epochs. The performance of the MLP-NN (having the following configurations: 5-12-12-1 and 7-12-12-1) over the entire test set, for RP and PB training algorithms, are represented in Figure 7.14 and Figure 7.15, respectively.

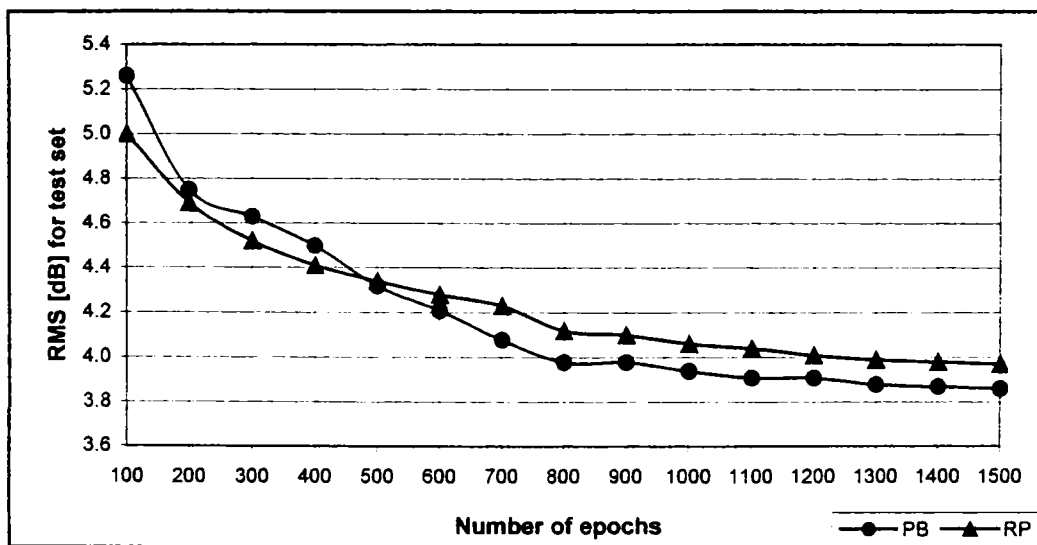


Figure 7.14. Performance of MLP-NN, 5-12-12-1, RP and PB algorithms, suburban environment

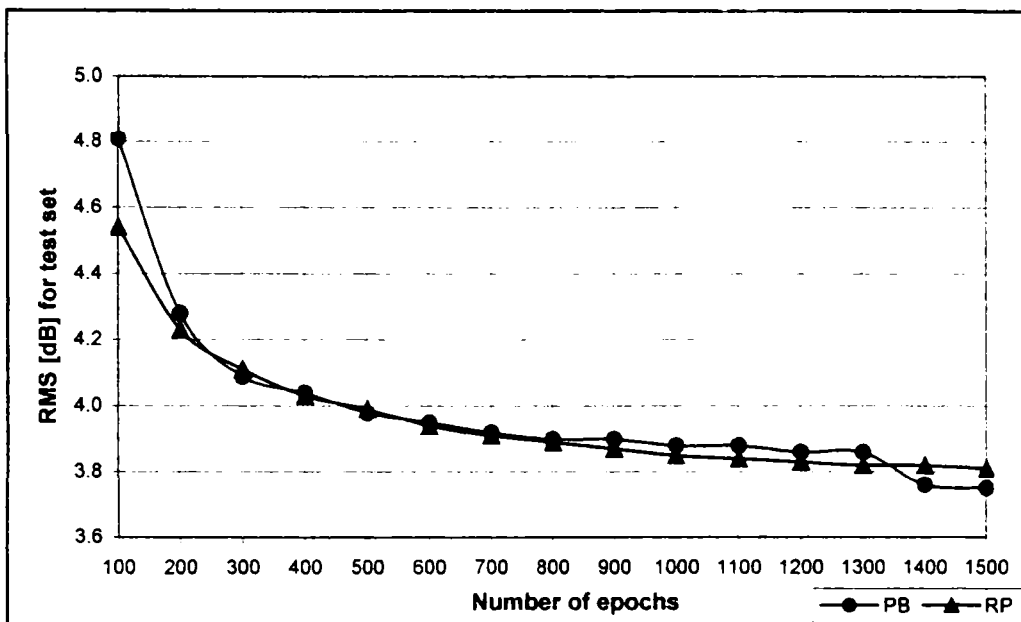


Figure 7.15. Performance of MLP-NN, 7-12-12-1, RP and PB algorithms, suburban environment

The proposed MLP neural network model is the one with 7 inputs, two hidden layers with 12 neurons each, trained with LM algorithm by using the early method stopping. The results obtained over the entire test set by the proposed MLP-NN model have been compared with the results obtained by using the Walfisch-Bertoni model (WB), the COST231-Walfisch-Ikegami model (CWI) and the Single Slope model (SSM) and are presented in Table 7.5.

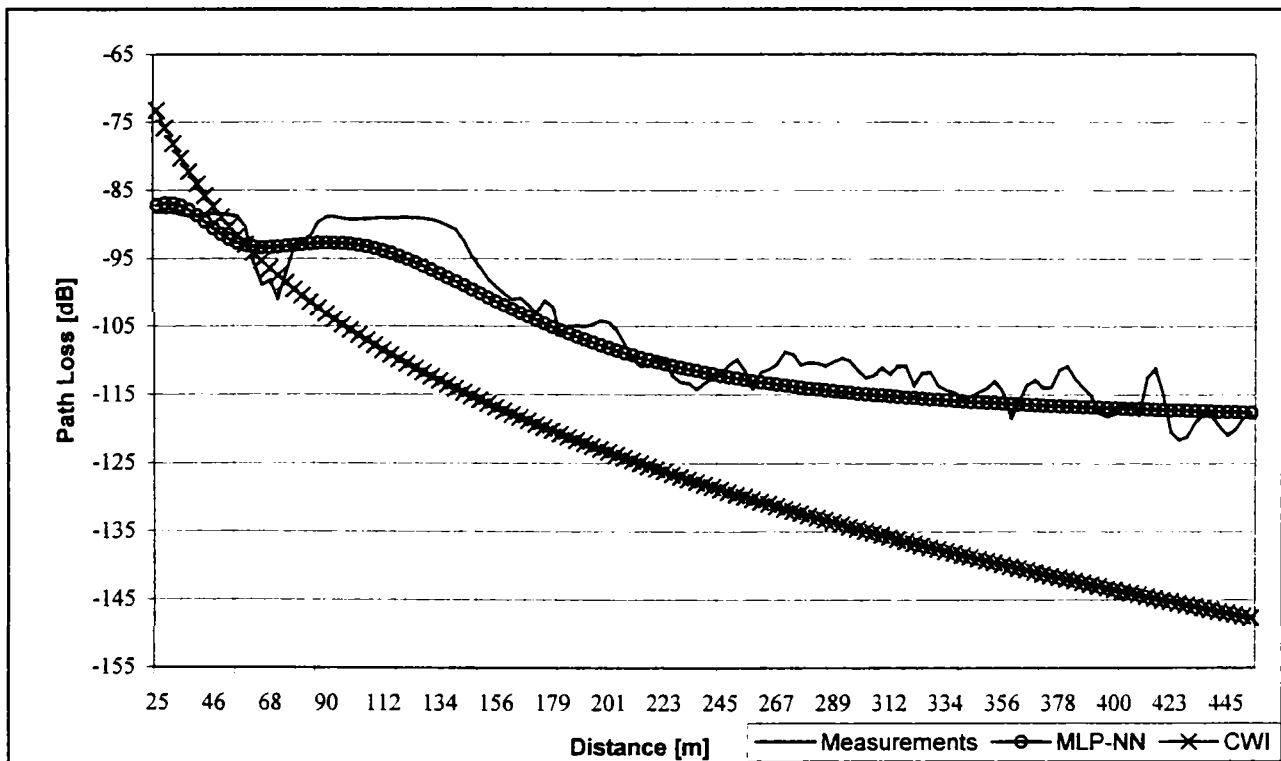


Figure 7.16. Comparison between the measured and the predicted propagation path loss by the neural model and the CWI model for one particular route.

Table 7.5. Comparison between the proposed MLP-NN model and the other empirical models, suburban environment

	MLP-NN	SSM	WB	CWI
μ [dB]	3.01	5.61	5.29	5.71
σ [dB]	2.23	3.81	3.92	3.70
RMS [dB]	3.74	6.82	7.06	6.84

Figure 7.16 represents the measured and predicted propagation path loss by the proposed MLP model and CWI model in case of one particular route.

7.2.4. Hybrid model based on MLP-NN

This section presents a hybrid model for the prediction of propagation path loss based on the combination between a theoretical model and a neural network [Popescu, 01c]. The neural network, based on the principles of feedforward multilayer neural networks, is used to compensate for the errors obtained by applying COST231-Walfisch-Ikegami (CWI) model [COST231, 99] in the urban (city of Kavala) and the suburban area (Oia village), respectively.

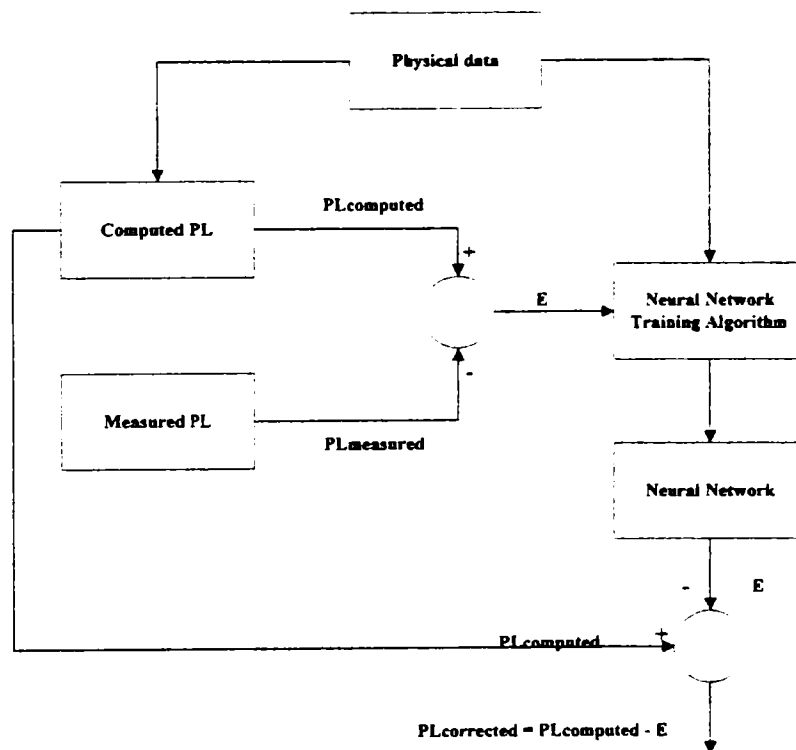


Figure 7.17. The schematic diagram of the training process

The purpose is to build a correction model: the neural network is used to compensate for the errors obtained by applying CWI model. In Figure 7.17 is represented the training phase of the neural network and in Figure 7.18 is represented the schematic diagram of the

prediction phase. The inputs of the neural network consist of the available physical parameters and the output is represented by the difference error between the path loss computed by COST-Walfisch-Ikegami model and measurements:

$$E = PL_{\text{computed}} - PL_{\text{measured}} \tag{7.6}$$

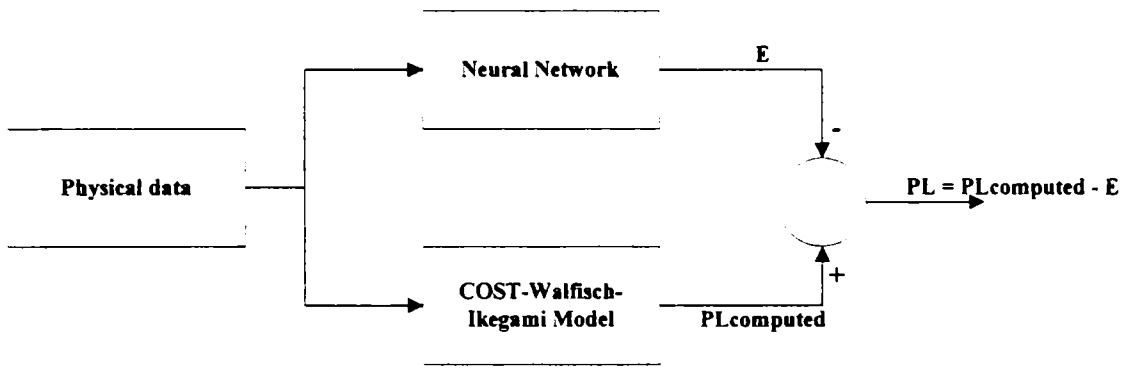


Figure 7.18. The schematic diagram of the prediction

A. Urban environment

The MLP-NN error correction model is built for the NLOS case and the size of the training and test sets are the same as the ones used to build the prediction model presented in section 7.2.2.B.

The first MLP-NN model investigated has as inputs the following physical parameters:

- distance between transmitter and receiver,
- width of streets,
- height of the buildings,
- building separation,
- street orientation.

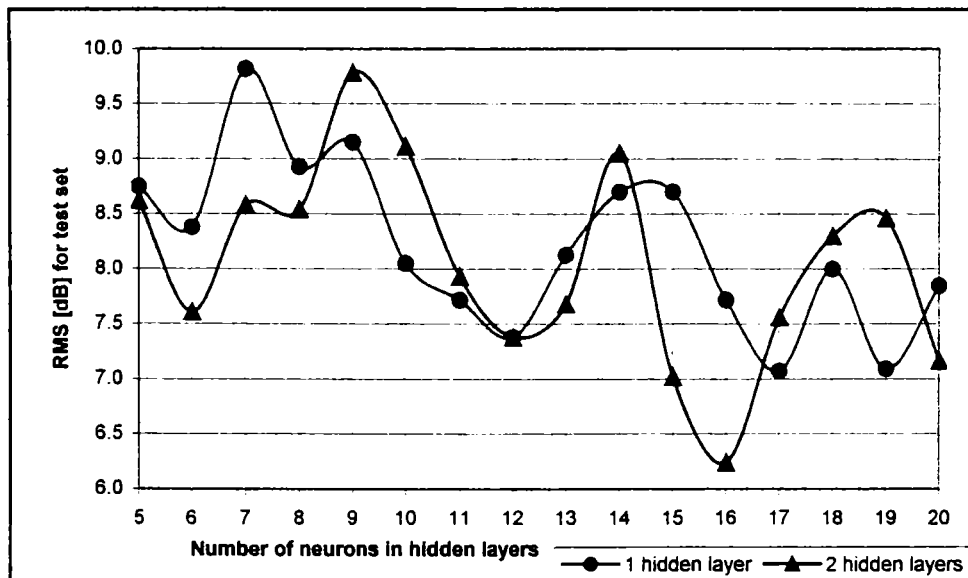


Figure 7.19. Performance of hybrid MLP-NN, LM algorithm, 1 and 2 hidden layers, 5 inputs, urban environment

In Figure 7.19 the performance of the MLP-NN trained with LM algorithm is presented, for different number of neurons in one and two hidden layers.

In addition to the above-mentioned parameters, the position of the transmit antenna with respect to the rooftop is used as input to the MLP-NN. The performance of the MLP-NN over the entire test set, obtained for 6 inputs, LM algorithm, is presented in Figure 7.20.

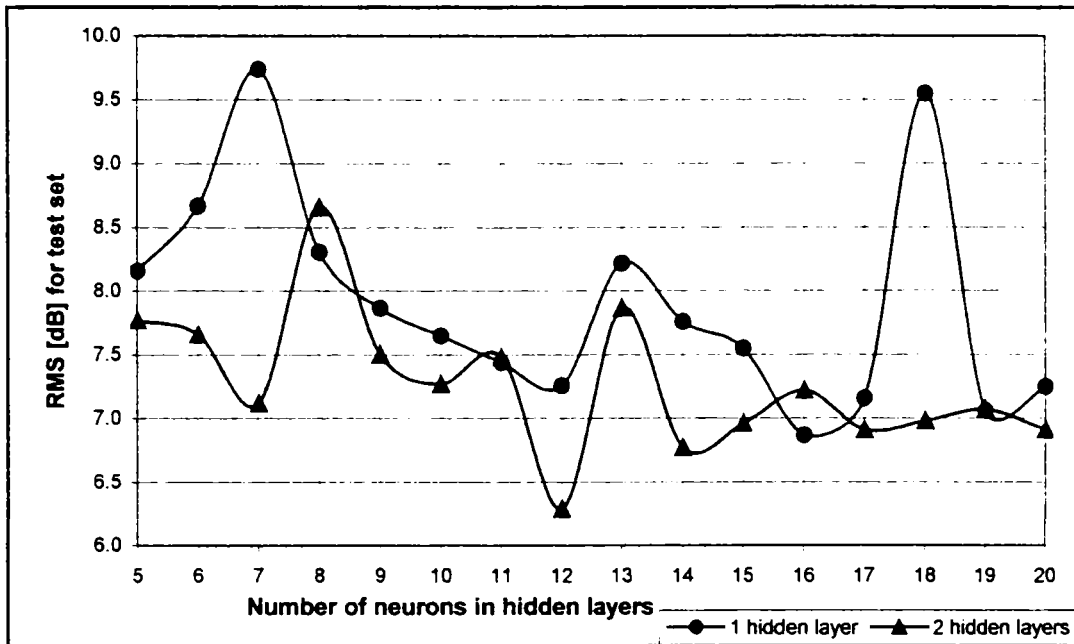


Figure 7.20. Performance of hybrid MLP-NN, LM algorithm, 1 and 2 hidden layers, 6 inputs, urban environment

As it can be seen from Figure 7.20, the results obtained in the case of MLP-NN with 6 inputs, 2 hidden layers with 16 neurons each, are RMS = 6.24 dB.

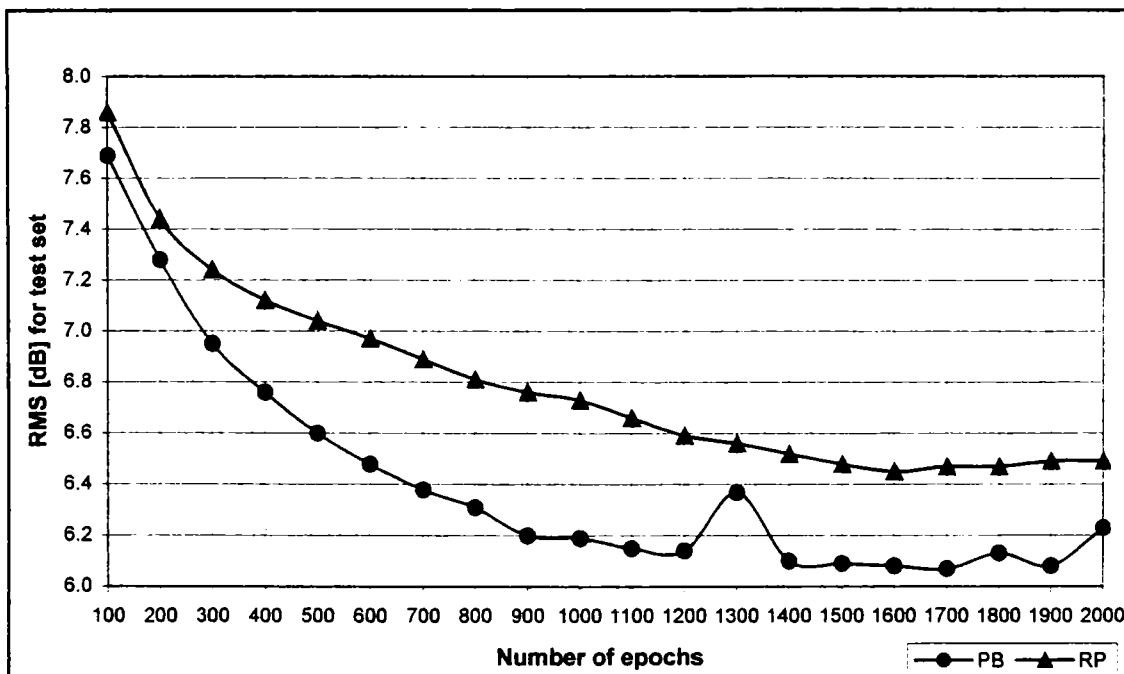


Figure 7.21. Performance of hybrid MLP-NN, 5-16-16-1, PB and RP training algorithms, different number of epochs

Further investigations were done on the configurations 5-16-16-1 and 6-16-16-1 for RP and PB training algorithms and for different number of epochs and the results are presented in Figure 7.21 and Figure 7.22, respectively.

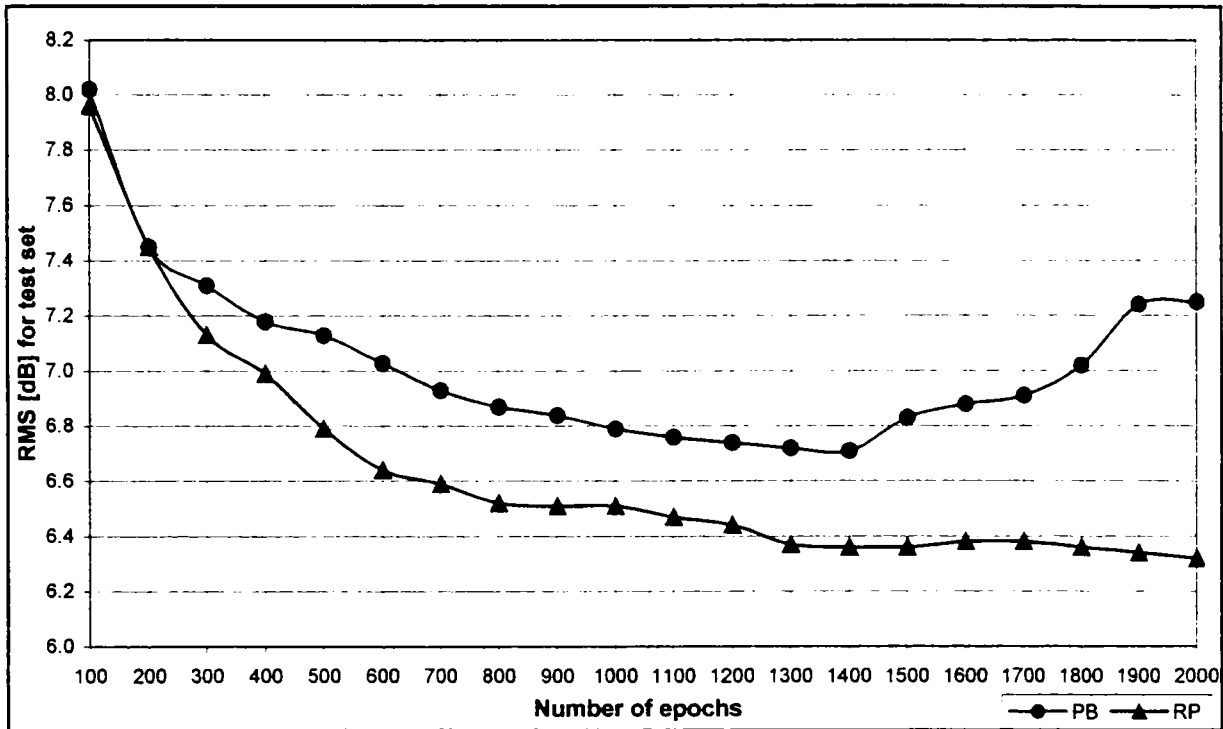


Figure 7.22. Performance of hybrid MLP-NN, 6-16-16-1, PB and RP training algorithms, urban environment.

The best performance was achieved by the MLP-NN having 2 hidden layers with 16 neurons each and 5 inputs and trained with the PB algorithm (1700 training epochs). The street orientation parameter gives no improvement in the prediction done by the neural network. The performance of the hybrid MLP-NN model, as described above, is compared with the measurement values and with the COST231-Walfisch-Ikegami (CWI) model (see Table 7.6).

Table 7.6. Comparison between the proposed hybrid MLP-NN and the CWI model in urban environment

	μ [dB]	σ [dB]	RMS [dB]
MLP-NN&CWI	4.14	4.43	6.07
CWI	6.97	4.61	8.35

In Figure 7.23 is represented the comparison between the prediction made by the error correction model, CWI model and measurements for a particular route in urban environment.

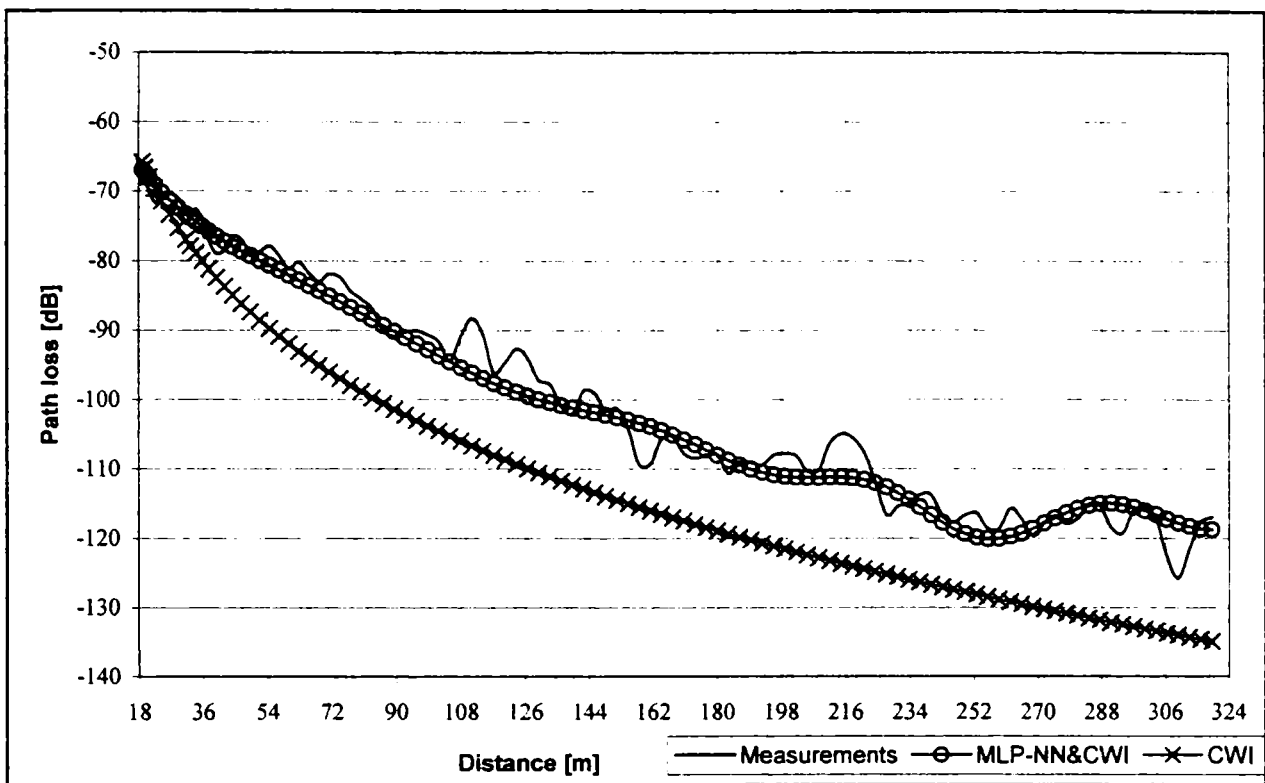


Figure 7.23. The comparison between the prediction made by the proposed error correction model, CWI model and measurements in case of a particular route, urban environment

B. Suburban environment

In order to model the suburban environment, the size of the training and test sets are the same as the ones used to build the prediction model presented in section 7.2.3.

As a first step, the study is based on a MLP-NN trained with the following physical parameters:

- distance between transmitter and receiver,
- width of the streets,
- height of the buildings,
- building separation.

In Figure 7.24 the performance of the MLP-NN on the test set, trained with LM algorithm is presented, for different number of neurons in one and two hidden layers.

In addition to the above-mentioned parameters, the street orientation is used as an input to the MLP-NN, The performance of the MLP-NN over the entire test set, obtained in case of 5 inputs, LM algorithm, is presented in Figure 7.25.

As it can be seen from Figure 7.25, a good prediction is made by the MLP-NN with 5 input parameters and 2 hidden layers having 17 neurons each (RMS = 4.21). Following the results of these simulations, further investigations were done on the configurations 4-17-17-1 and 5-17-17-1 for RP and PB training algorithms and for different number of epochs. The results are presented in Figure 7.26 and Figure 7.27, respectively.

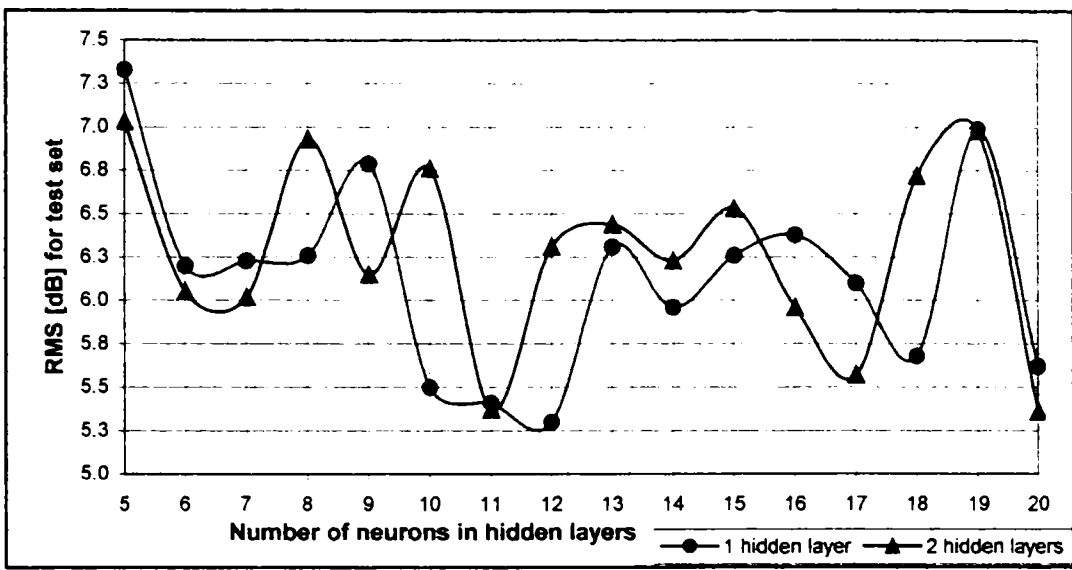


Figure 7.24. Performance of hybrid MLP-NN, LM algorithm, 4 inputs, 1 and 2 hidden layers, suburban environment

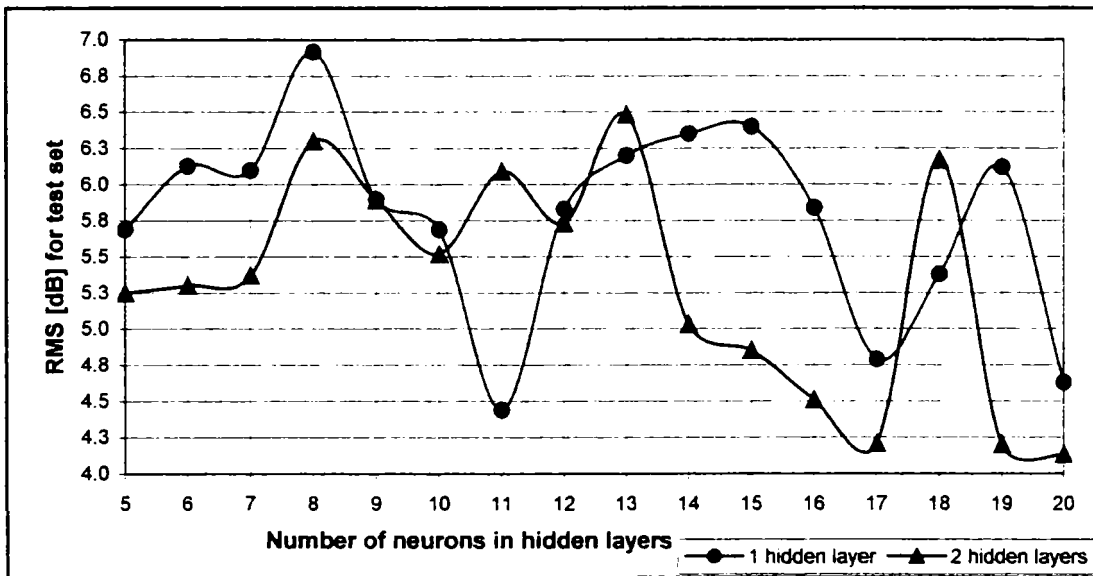


Figure 7.25. Performance of hybrid MLP-NN, LM algorithm, 5 inputs, 1 and 2 hidden layers, suburban environment

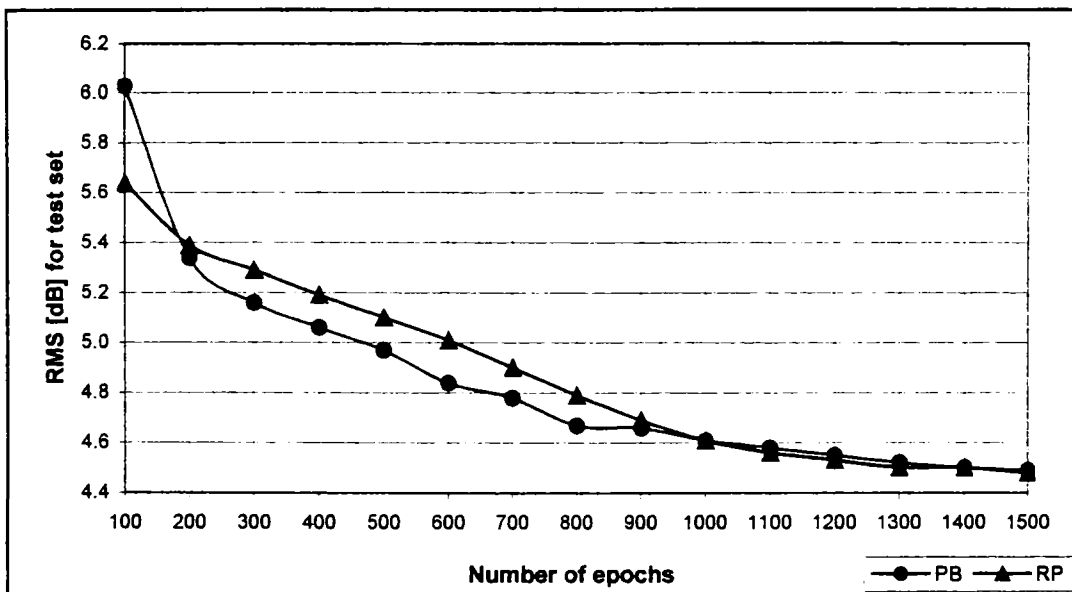


Figure 7.26. Performance of hybrid MLP-NN, RP and PB algorithms, 4-17-17-1, suburban environment

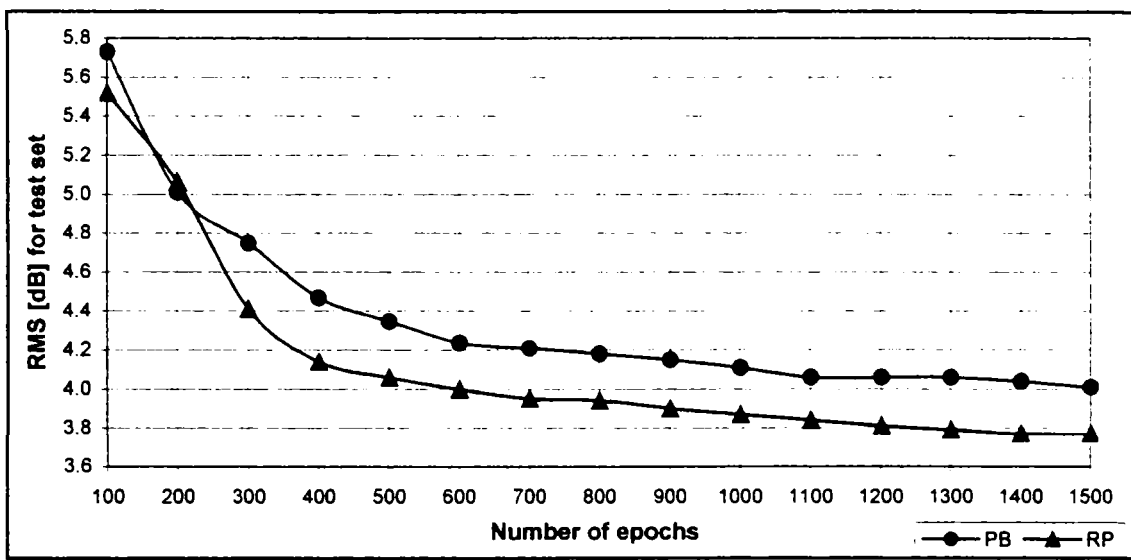


Figure 7.27. Performance of hybrid MLP-NN, 5-17-17-1, RP and PB algorithms, suburban environment

The best performance was achieved by the MLP-NN having the configuration of 5-17-17-1 and trained with the RP algorithm. The performance of the hybrid MLP model, as described above, is compared to the measured values and with the COST231-Walfisch-Ikegami (CWI) model (see Table 7.7).

In Figure 7.28 is represented the comparison between the prediction made by the error correction model, CWI model and measurements in case of a particular route, in suburban environment.

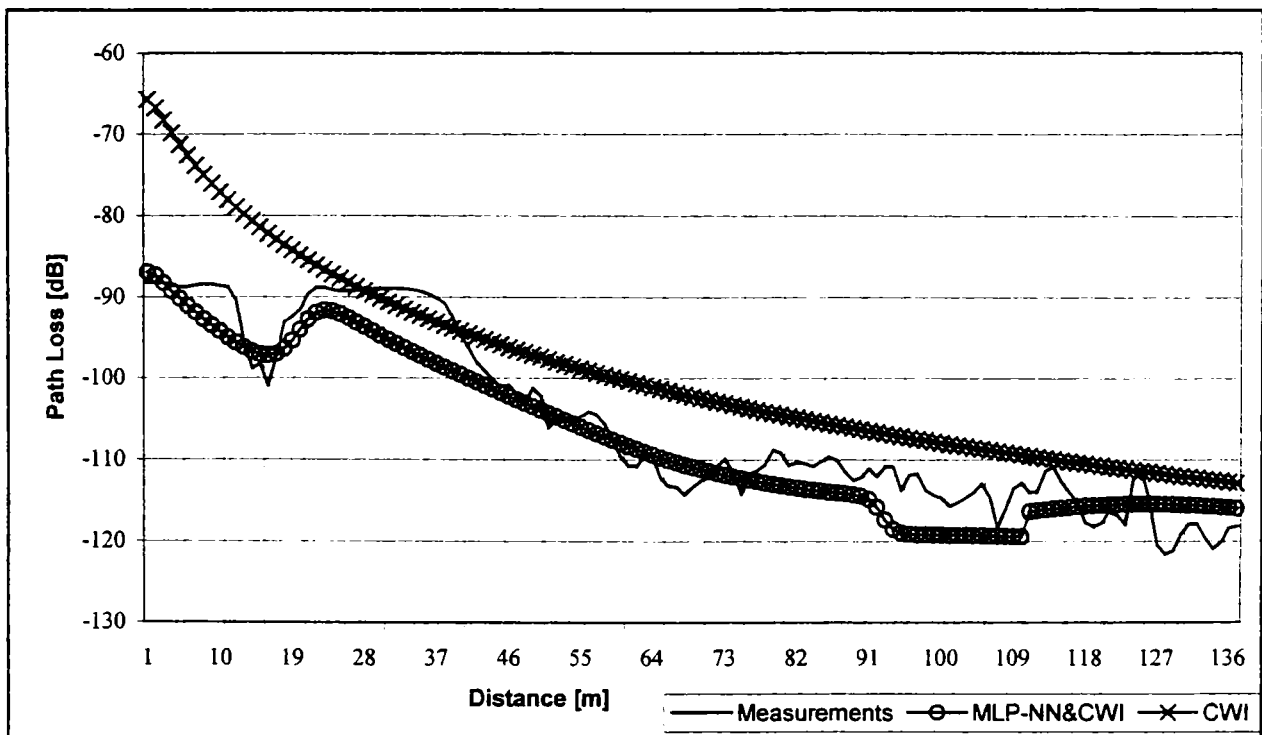


Figure 7.28. The comparison between the prediction made by the proposed hybrid model, CWI model and measurements for a particular route, suburban case.

Table 7.7. Comparison between the proposed hybrid MLP-NN model and the CWI model, suburban environment

	μ [dB]	σ [dB]	RMS [dB]
MLP-NN&CWI	2.86	2.46	3.77
CWI	5.71	3.70	6.84

7.2.5. Proposed MLP-NN models for propagation prediction in indoor environment

The presented study includes the single floor scenario and the procedure used to select the measurement data is described below.

In order to train the neural network we have used the measurements collected from two branches of the building, denoted sector B, where the transmitter was always located, and sector A. The fast fading was eliminated, in the case of longitudinal measurements (along the corridors), by averaging the measured received power using a 2λ windowing technique [Honcharenko, 92]. In the case of static measurements the average values of the recorded samples in every position of the receiver inside the offices were computed. In this way we have obtained two values for the received power in each office (with closed doors respectively with open doors) for each combination of the position, height and gain of the transmitter antenna. Following the filtering process of the measured data we obtained more than 1400 measurement locations corresponding to the non-line-of-sight (NLOS) case.

The inputs of the neural network are the following:

- Three inputs for the transmitter site: position, gain and height of the antenna;
- Two inputs for the distance: one is the distance between transmitter and the starting point of the measurements and the second input is represented by the distance covered by the mobile unit;
- Two inputs describing the receiver site: the sector and the location within the sector.
- Three inputs for the smallest number of walls and windows penetrated by the ray between transmitter and receiver and their accumulated losses.

The input parameters that describe the transmitter and receiver location are quantized so the effect of each parameter is more obvious for the neural network [Haykin, 94]. For example, the parameters like size of the corridors where the receiver is located are quantized as follows: 1 for the large corridor and 0.3 for the medium corridor. All parameters are normalized to the range $[-1, +1]$.

The output layer consists of one neuron that provides the normalized received power. A data set of 289 patterns is used for training (out of which, 18 patterns are used for validation). A set of 1155 training patterns was used to test the model.

In Figure 7.29 the performance of the MLP on the test patterns is presented for different number of neurons in 1 and 2 hidden layers, using the LM training algorithm.

Investigations were done on the configuration of the MLP with 2 hidden layers having 14 neurons each trained with RP and PB algorithms. The RMS errors obtained for the entire test set by using the early stopping method were unsatisfactory and the simulations were done for different number of epochs. In case of RP and PB training algorithms, the performance of the MLP on the entire test set is shown in Figure 7.30.

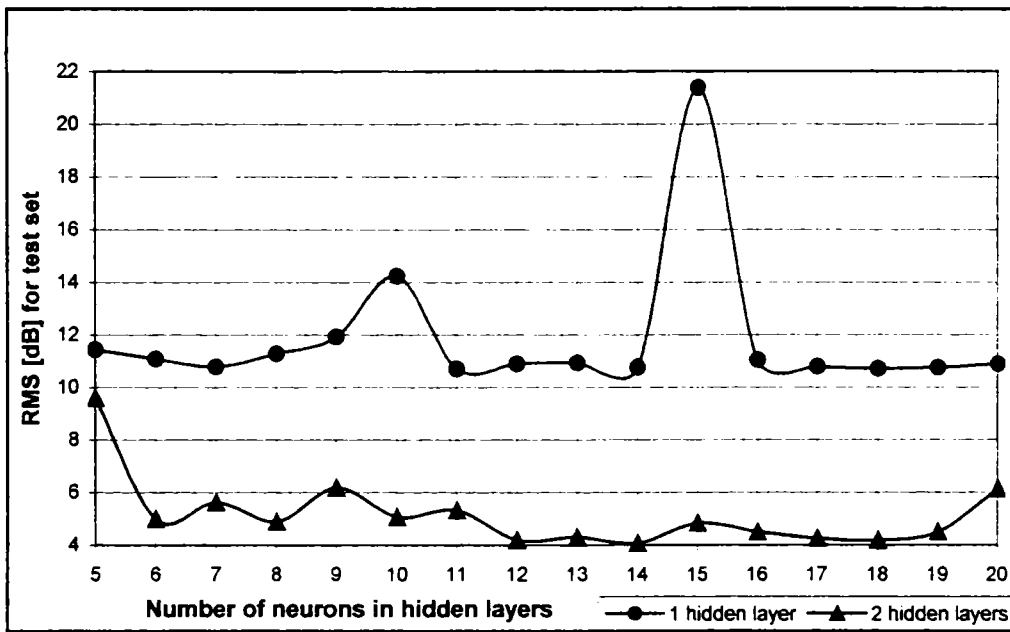


Figure 7.29. Performance of MLP-NN, 10 inputs, LM training algorithm, 1 and 2 hidden layers, indoor environment

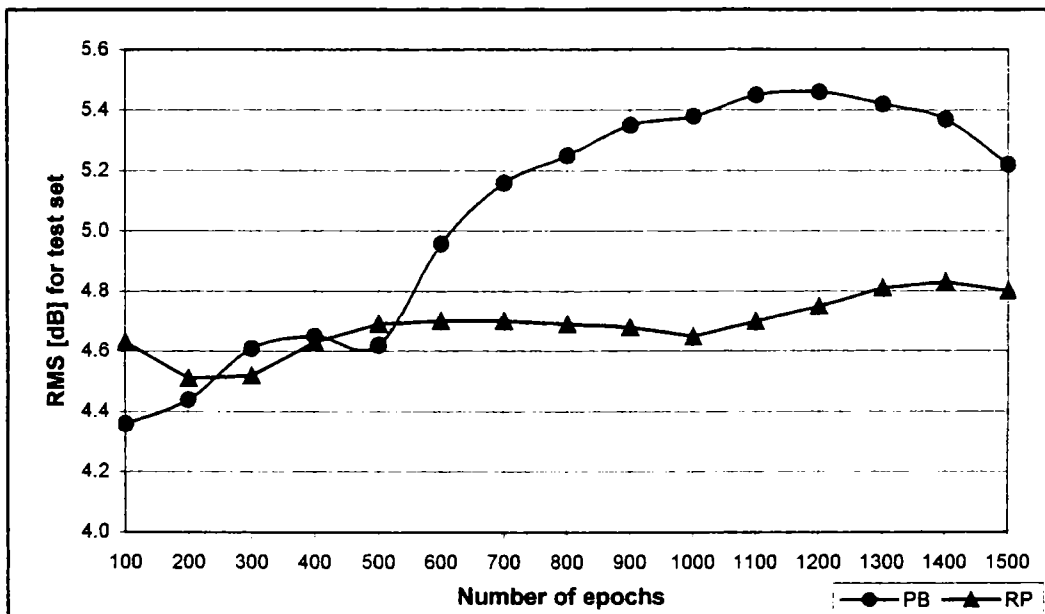


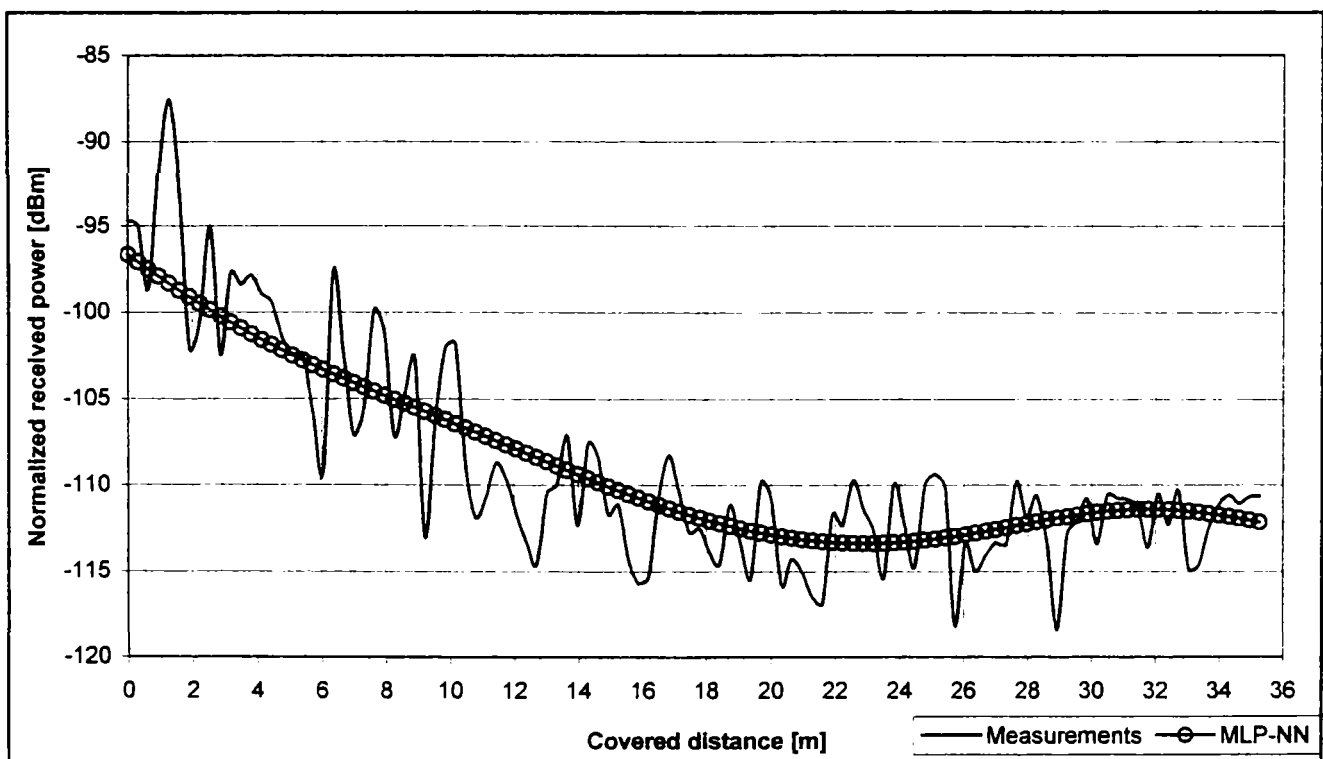
Figure 7.30. Performance of MLP-NN, RP and PB algorithms, 10-14-14-1, indoor environment

As it can be seen from Figures 7.29-7.30 the best performance is achieved with the MLP with 2 hidden layers with 14 neurons each and trained with the LM algorithm. In Table 7.8 are represented the absolute mean error, the standard deviation and the root mean squared error obtained for the training set and for the test set.

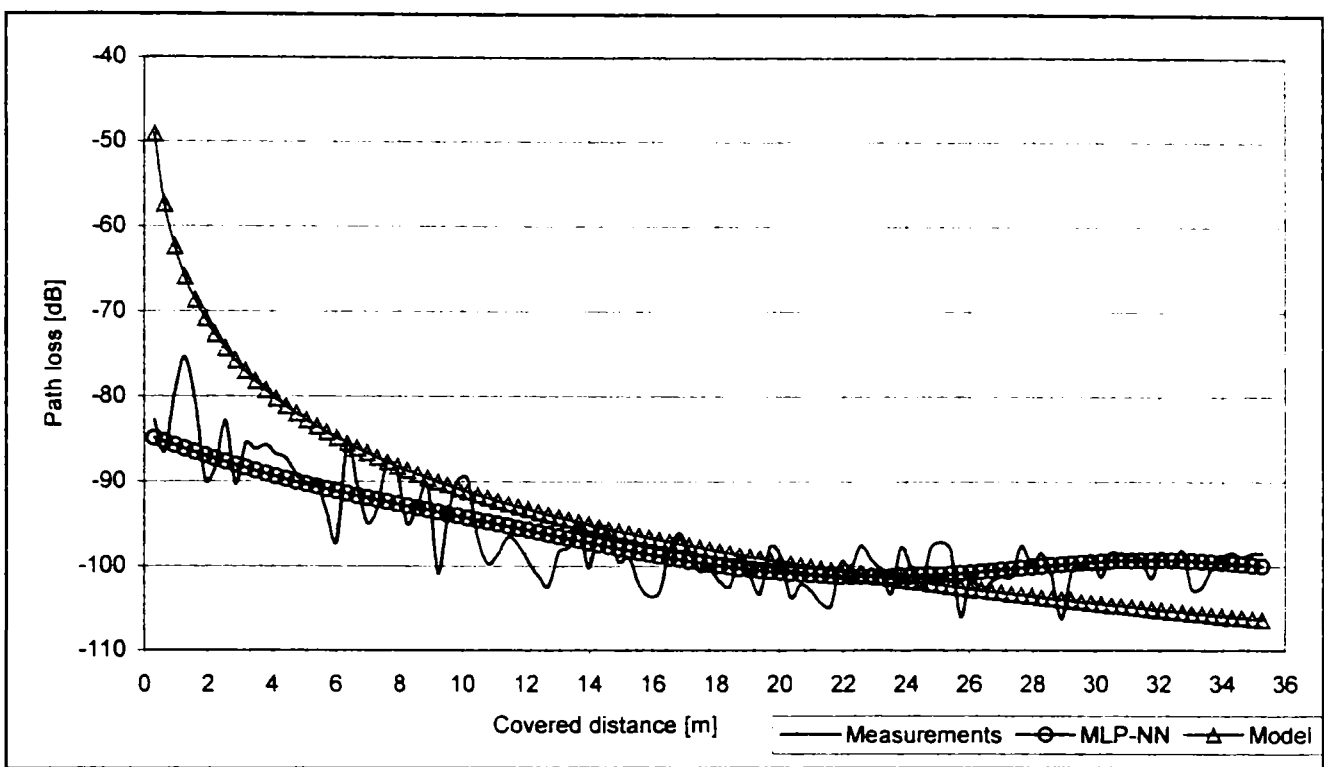
Table 7.8. Results of the prediction, MLP-NN, indoor environment

	Training patterns	Test patterns	Particular route
μ [dB]	2.77	3.05	2.33
σ [dB]	2.31	3.15	1.79
RMS [dB]	3.61	4.38	2.94

In Figure 7.31.a is represented a comparison between predicted and measured values of the normalized received power when the transmitter is in position 1 and the receiver is located along the main corridor in sector A. For this particular route, the values for the absolute mean error, standard deviation and RMS are presented in Table 7.8. The percentage of the predicted values with a mean error below 5 dB is 81.55 % in the case of the entire test set respectively 90.17 % in the case of the particular route. In Figure 7.31.b the comparison between the predicted path loss by the MLP-NN, measured values and predicted path loss by the empirical model (described in section 5.4.2.2) is shown. In case of the theoretical model the following results were obtained: $\mu = 5.29$ dB, $\sigma = 5.04$ dB and RMS = 7.31 dB.



a)



b)

Figure 7.31. Comparison between predictions and measurements with the transmitter in position 2 and the receiver located in sector A, along the main corridor: a) normalized received power, b) path loss

7.3. Proposed RBF-NN models for the prediction of propagation path loss

In this section the performance of the generalized Radial Basis Function Neural Networks (RBF-NN) applications for the prediction of propagation path loss in urban and suburban environment. A detailed description of this type of neural network is given in chapter 2. A number of RBF-NN models, trained on extended sets of propagation path loss measurements taken in different environments, are studied. The performance of all RBF-NN models with different input parameters is evaluated by comparing their prediction error statistics of absolute mean error (μ), standard deviation (σ), root mean square error (RMS) and the correlation (R) between predicted values and the measurement data.

7.3.1. Proposed RBF-NN models for propagation prediction in urban environment

In this section are studied RBF-NN models for the prediction of propagation path loss in urban environment (city of Kavala, Greece). The RBF-NN models are designed separately for line-of-sight (LOS) and for Non-line-of-sight (NLOS) cases. The performance of the

proposed RBF-NN models is compared to that of the COST231-Walfisch-Ikegami model (CWI) [COST 231, 99], the Walfisch-Bertoni model (WB) [Walfisch, 88] and the Single Slope model (SSM) [Rappaport, 96], based on the absolute mean error, standard deviation and the root mean squared error between predicted and measured values.

A. LOS case

For the LOS case, the proposed RBF-NN model is trained with physical data that includes the distance between transmitter and receiver, the width of the streets, the height of the buildings, the building separation and the difference between the base station antenna height and the rooftop height. The model has a single output which represents the normalized propagation path loss. The dimension of the training set is 1013 and the rest of 2026 examples were used to test the model.

Table 7.9 represents the performance achieved by the proposed RBF-NN model and the other empirical models over the entire test data.

Table 7.9. Proposed RBF-NN model and other empirical models, LOS case, urban environment

	RBF-NN	SSM	CWI	WB
μ [dB]	5.04	5.43	7.04	9.09
σ [dB]	4.54	4.80	4.06	4.51
RMS [dB]	6.78	7.24	8.23	10.24

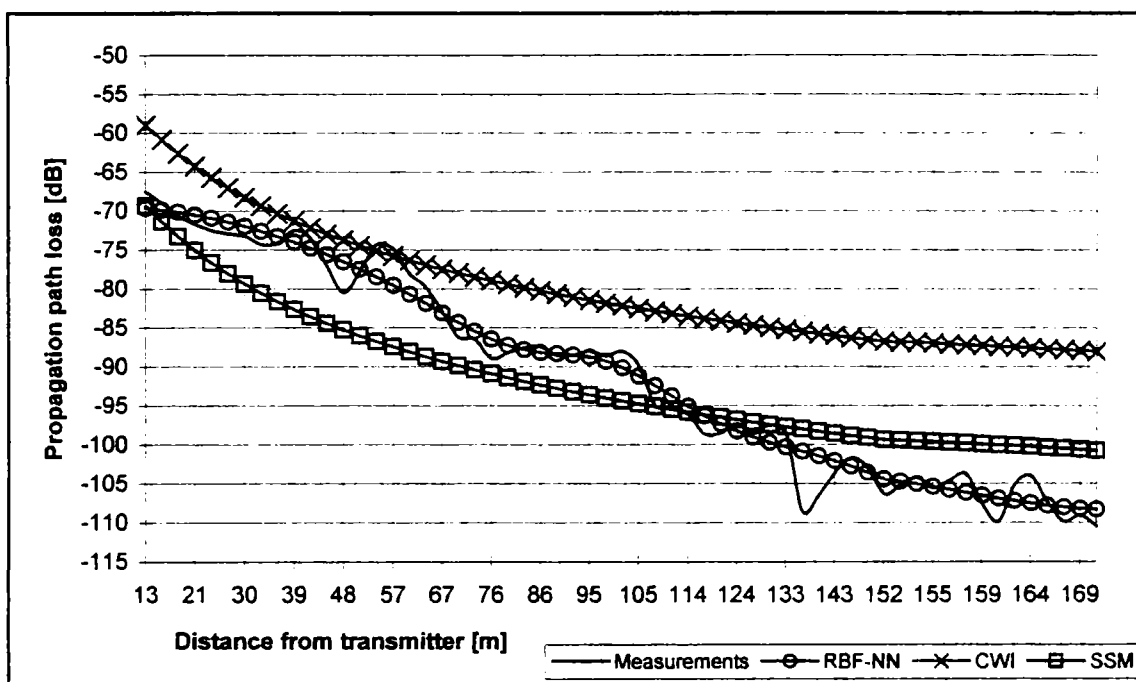


Figure 7.32. Measured and predicted path loss for LOS case, urban environment

Figure 7.32 represents the measured and predicted propagation path loss by the proposed RBF-NN model, the SSM model and the CWI model for a specific route, in LOS case.

B. NLOS case

Three RBF-NN models with different inputs were studied:

- The first one, named RBF1, is trained with four parameters: the distance between transmitter and receiver, the width of the street, the building separation and the building height.
- The second model, named RBF2, in addition to the above-mentioned data we also include the difference between the base station antenna height and the building height.
- In the third neural model, named RBF3, we have included the street orientation in addition to the parameters used for the training of RBF2.

A set of 420 examples was used for training purpose while the rest of 1680 was used for test purpose. Table 7.10 presents the results obtained by the three different RBF-NN models, for the test set. R represents the correlation between predicted values and the measurement data.

Table 7.10. RBF-NN models for NLOS case, urban environment

	μ [dB]	σ [dB]	RMS [dB]	R
RBF 1	5.50	5.27	7.62	0.911
RBF 2	5.31	5.07	7.34	0.917
RBF 3	3.67	3.88	5.35	0.957

Table 7.11 represents the comparison of the performance achieved by the RBF 3 model and the Single Slope model (SSM), the Walfisch-Bertoni model (WB) and the modified COST231-Walfisch-Ikegami model (CWI) for the entire test patterns.

Table 7.11. Proposed RBF3 model and other empirical models, NLOS case, urban environment

	RBF 3	SSM	WB	CWI
μ [dB]	3.67	6.35	6.08	6.96
σ [dB]	3.88	4.37	4.14	4.62
RMS [dB]	5.35	7.75	7.40	8.38

Figure 7.33 represents the measured and predicted propagation path loss by the proposed RBF3 model, CWI model and measurements for one route characterized by a base station antenna located below rooftop.

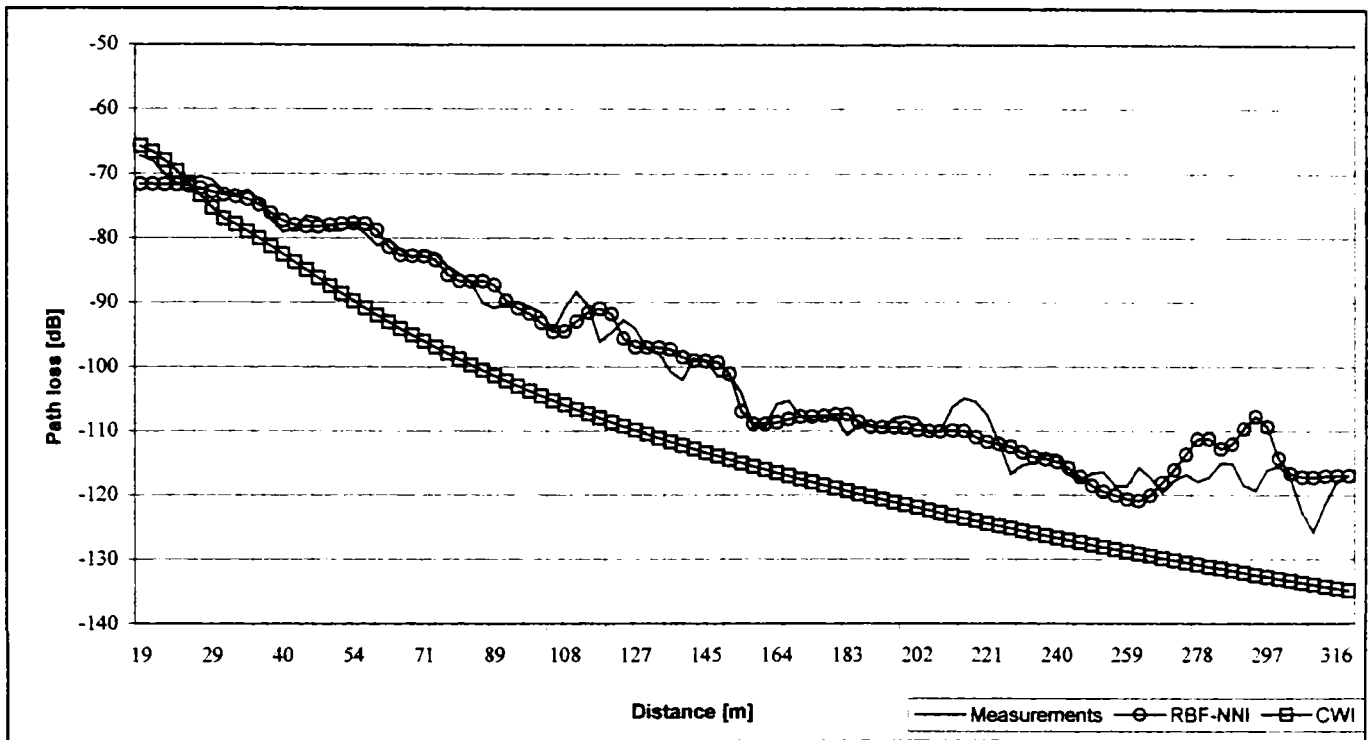


Figure 7.33. Measured and predicted path loss, in case of a particular route, NLOS case, urban environment

7.3.2 Proposed RBF-NN models for propagation prediction in suburban environment

In the case of suburban environment, four RBF-NN models were studied:

- The first model, RBF 4, has four inputs: the distance between transmitter and receiver, the width of the street, the building separation and the building height.
- The second model, RBF 5, in addition to the above-mentioned data we also included the street orientation.
- The third model, RBF 6, has six inputs: the distance between transmitter and receiver, the width of the street, the building separation, the building height, the roof-to-screen diffraction and multi screen diffraction losses computed with the modified COST231-Walfisch-Ikegami formula.
- The fourth model, RBF 7, uses as inputs all the seven parameters mentioned above.

A set of 333 examples (20% from the data measurement set) was used for training purpose while the rest of 1332 was used for test purpose. Table 7.12 presents the results obtained by the four different RBF-NN models, for the entire test set.

Table 7.12. Generalized RBF-NN models, suburban environment

	μ [dB]	σ [dB]	RMS [dB]	R
RBF4	3.68	3.71	5.23	0.958
RBF5	2.76	2.65	3.83	0.963
RBF6	2.68	2.60	3.73	0.964
RBF7	2.65	2.55	3.68	0.966

Table 7.13 represents the comparison of the performance achieved by the proposed RBF7 model and the Single Slope model (SSM), the Walfisch-Bertoni model (WB) and the modified COST231-Walfisch-Ikegami model (CWI) for the entire test patterns.

Table 7.13. Proposed RBF 7 model and the other empirical models, suburban case

	RBF7	SSM	WB	CWI
μ [dB]	2.65	5.61	5.29	5.71
σ [dB]	2.55	3.81	3.92	3.70
RMS [dB]	3.68	6.82	7.06	6.84

Figure 7.34 represents the measured and the predicted path loss by the proposed RBF7 model and CWI model in case of one particular route, in suburban environment

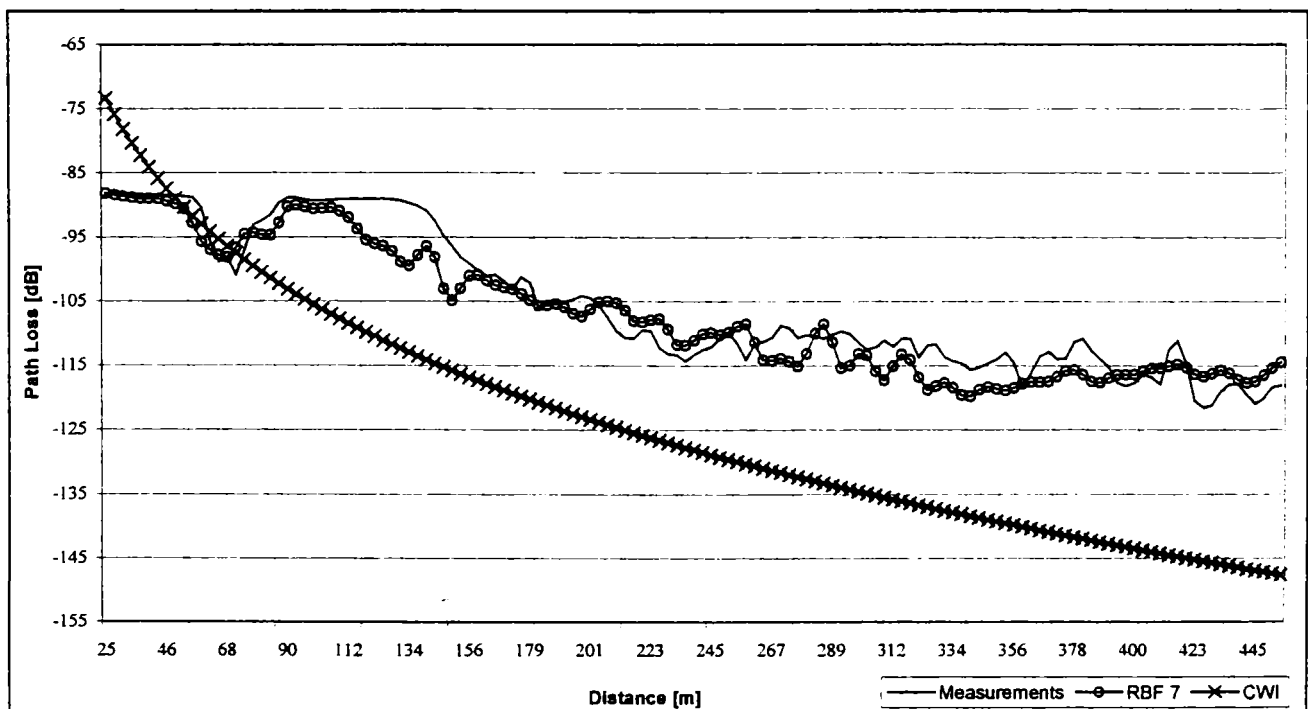


Fig. 7.34. Measured and predicted propagation path loss by the proposed RBF 7 model and the CWI model, suburban environment

7.3.3. Hybrid models based on RBF-NN

The purpose is to build an error correction model: the neural network is used to compensate for the errors obtained by applying the COST-Walfisch-Ikegami (CWI) model, as described in section 7.2.4.

For training and test purpose we have used the same number of patterns as in the prediction models, for both cases, urban and suburban environment.

A. Urban environment

In the case of urban environment, the following RBF-NN models were investigated:

- RBF8 that has four inputs: distance between transmitter and receiver, the width of the street, the height of the buildings and the building separation.
- RBF9 that has 5 inputs: the above-mentioned parameters used to design RBF 8 plus the difference between the base station height and the building height.
- RBF10 with 5 inputs: the parameters used to design RBF 8 plus the street orientation.
- RBF11 with 6 inputs represented by all parameters used to design the above-mentioned models.

The comparison between the results obtained by the Generalized RBF-NN models for the test patterns are presented in Table 7.14.

Table 7.14. Hybrid RBF-NN models for NLOS case, urban environment

	μ [dB]	σ [dB]	RMS [dB]	R
RBF8	6.12	6.13	8.66	0.774
RBF9	5.68	5.59	7.97	0.812
RBF10	3.79	4.07	5.57	0.913
RBF11	3.65	3.85	5.30	0.922

Table 7.15 represents the comparison between the proposed RBF11 hybrid model and CWI model for the entire test set.

Table 7.15. Proposed RBF11 model and the CWI model

	μ [dB]	σ [dB]	RMS
RBF11+CWI	3.65	3.85	5.30
CWI	6.97	4.62	8.38

In Figure 7.35 is represented the comparison between the prediction made by the proposed error correction model RBF11, the CWI model and measurements, for a particular route in urban environment.

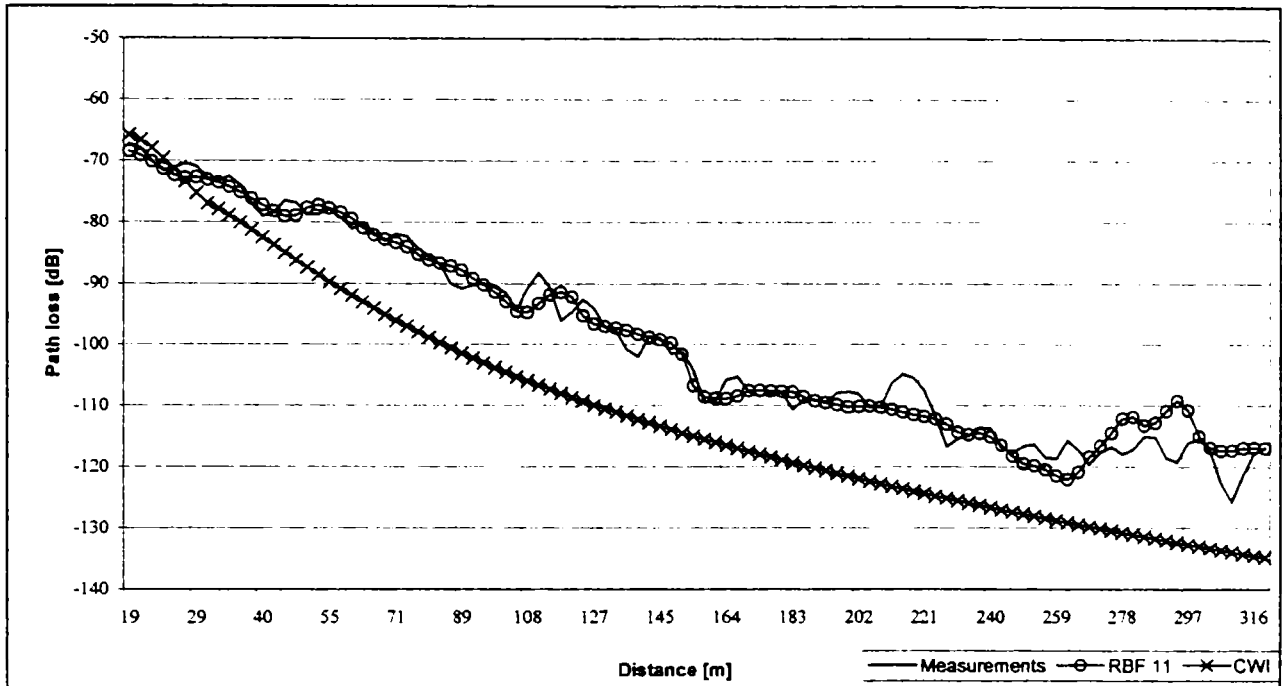


Figure 7.35. Prediction made by the proposed RBF11 hybrid model, CWI model and measurements, urban environment

B. Suburban environment

In the case of suburban environment, the following two hybrid RBF-NN models were studied:

- RBF12 that has four inputs: distance between transmitter and receiver, the width of the street, the height of the building and the building separation.
- RBF13 that has 5 inputs: the above-mentioned parameters plus the street orientation.

The comparison between the results obtained by the Generalized RBF-NN models for the entire set of test patterns are presented in Table 7.16.

Table 7.16. Hybrid RBF-NN models, suburban case

	μ [dB]	σ [dB]	RMS [dB]	R
RBF 12	3.22	2.99	4.40	0.832
RBF 13	2.71	2.54	3.71	0.883

Table 7.17. Comparison between the proposed RBF13 hybrid model and the CWI model

	μ [dB]	σ [dB]	RMS
RBF 13	2.71	2.54	3.71
CWI	5.71	3.70	6.84

Table 7.17 represents the comparison between the proposed hybrid model RBF 13 and CWI model for the entire test set.

Figure 7.36 represents the comparison between the prediction made by the proposed error correction model RBF13, the CWI model and measurements for a specific route in suburban environment.

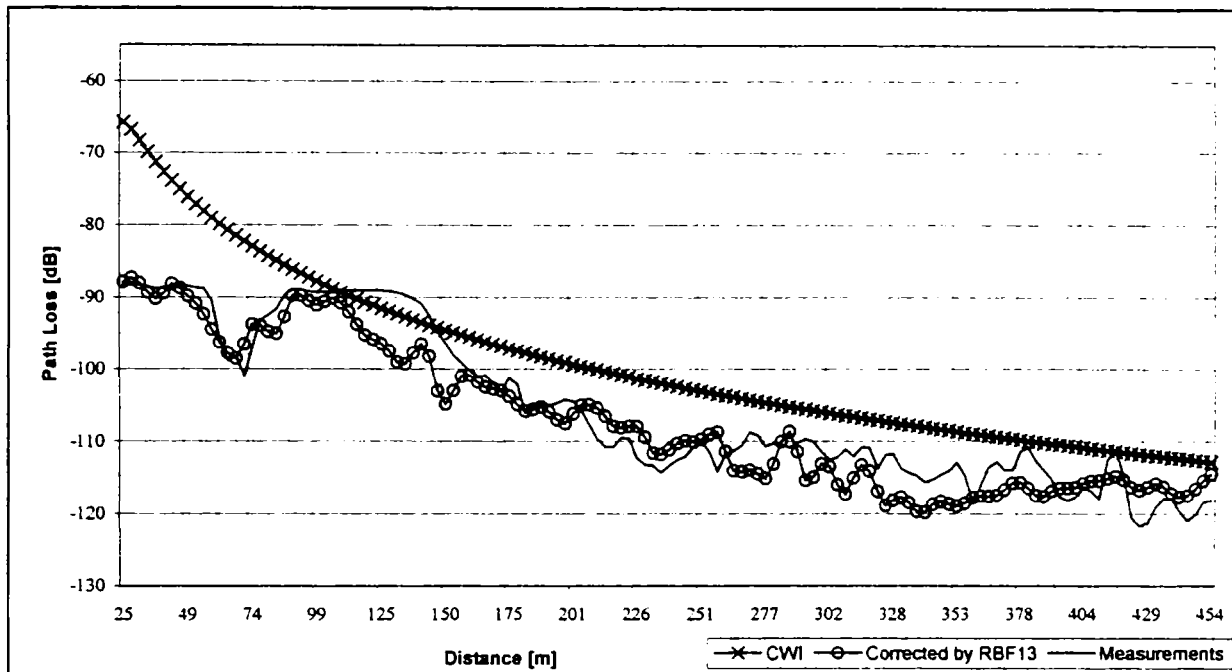


Fig. 7.36. Prediction made by the proposed hybrid model RBF13, CWI model and measurements, suburban case

7.3.4. Proposed RBF-NN model for propagation prediction in indoor environment

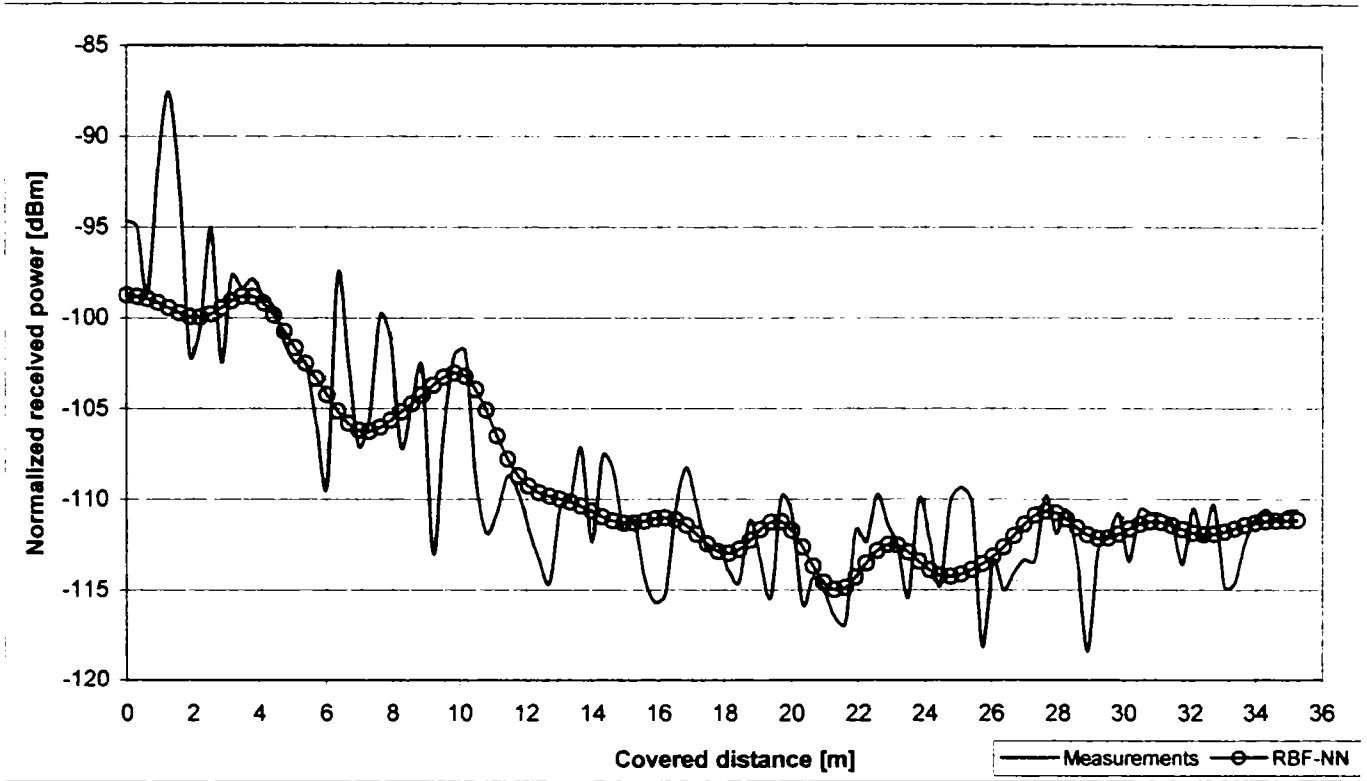
In the case of indoor environment, the input vector used to design the proposed RBF-NN model is the same as described in section 7.2.5. A data set of 289 patterns is used for training and a set of 1155 training patterns is used for test purpose.

In Table 7.18 are represented the mean error, the standard deviation and the root mean square error obtained for the training set and for the test set by the proposed RBF-NN model.

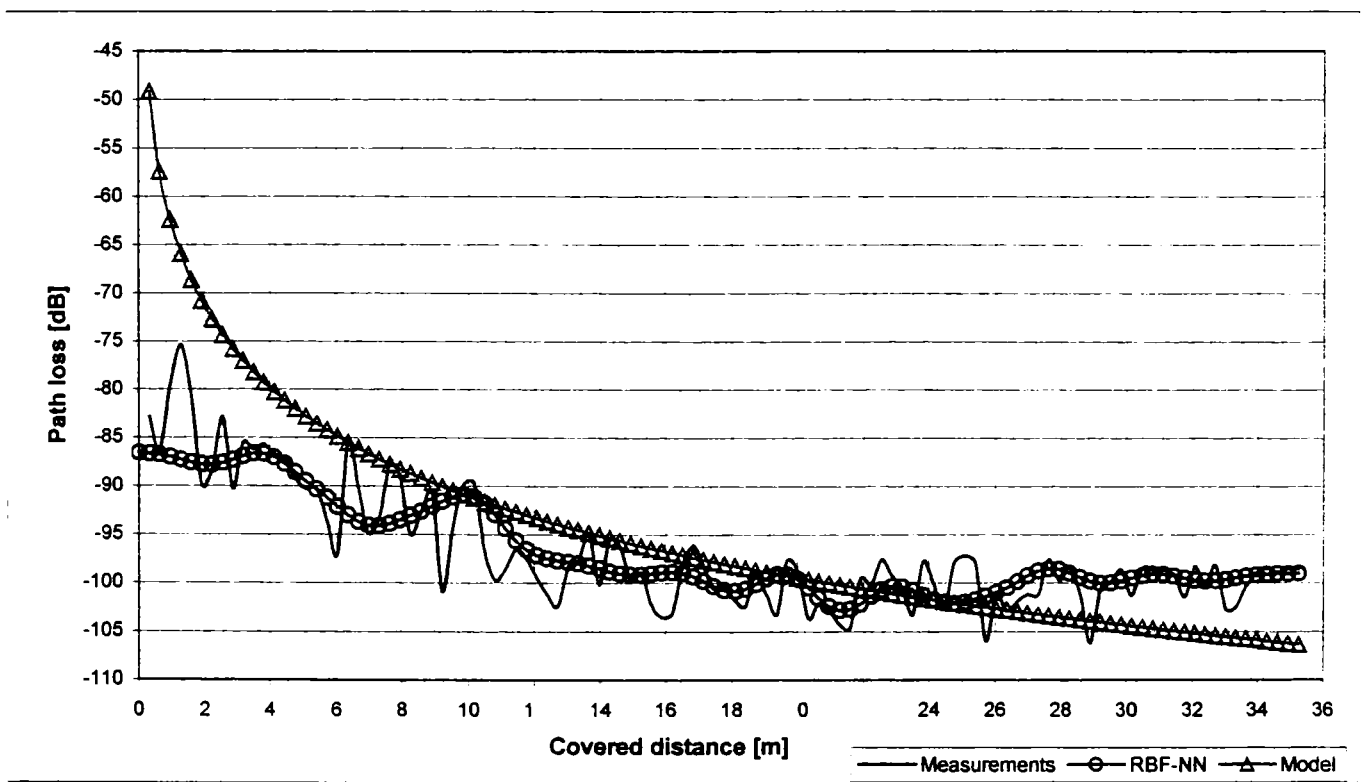
Table 7.18. Results of prediction, indoor environment

	Training patterns	Test patterns
μ [dB]	1.49	3.09
σ [dB]	1.71	2.88
RMS [dB]	2.27	4.23

In Figure 7.37.a is represented a comparison between predicted and measured values of the normalized received power for an NLOS route. In Figure 7.37.b the comparison between the predicted path loss by the RBF-NN model, by the empirical model (described in section 4.4.2.2) and measured values is shown.



a)



b)

Figure 7.37. Comparison between predictions and measurements with the transmitter in position 2 and the receiver located in sector A, along the main corridor: a) normalized received power; b) path loss

7.4. Discussions

Propagation measurement results for different environment (urban, suburban, indoor) were used to design neural network based models. In order to examine the validity of the NN models, the predicted path loss by them is compared to the measured values and to the path loss obtained by applying empirical models.

The empirical propagation models that have been examined in case of outdoor environment are the COST231-Walfisch-Ikegami (CWI) and Walfisch-Bertoni (WB) that are complex and require a more detailed description of the environment. A simple model, Single Slope model (SSM), is also examined. The SSM is determined by fitting the data in the sense of the minimum square error, while the other empirical models need more real and accurate environment parameters to validate their accuracy and capability.

Between the empirical algorithms, the single slope model achieves the best performance. However, this model is based on the distance between transmitter and receiver, the frequency and the propagation factor. The linear regression model used to predict path loss for both LOS and NLOS cases is of the form given by equation (5.38) considering the reference distance as 1 m. In LOS case, urban environment, the path loss exponent was found to range from 1.56 to 3.05. The low value implies waveguiding effect while the maximum value occurs due to the obstructed first Fresnel zone [Kanas, 99]. The total RMS error for all LOS test set is 7.24 dB. In NLOS case, urban environment, the path loss ranges from 2.1 to 5.2 that show highly obstructed routes [Kanas, 99]. The total average path loss exponent for NLOS case, urban environment is 3.65 dB. The RMS error for the all NLOS data test, urban environment, was found to be 7.75 dB.

In CWI model, LOS case, urban environment, a simple propagation loss formula is applied, as described in section 5.2.5 by equation (5.16), and an RMS error of 8.23 dB was obtained. In NLOS condition the CWI model is designed to provide better results for base station antenna heights above the rooftop. The RMS error obtained over the entire test data is 8.38 dB.

In LOS condition, the obtained RMS error with the WB model is 10.24 dB. In NLOS case the RMS error decreases to 7.40 dB and this confirms that the model is more appropriate to predict path loss in NLOS case.

In case of suburban environment, the only NLOS condition routes were examined. The average path loss exponent in this case was found to be 3.1. The overall RMS error is

6.82 dB. Applying CWI formula in suburban environment, it was obtained 8.38 dB RMS error and by applying WB formula in suburban environment, an RMS error of 7.40 dB was obtained.

The performance of the CWI model is poor for $h_{BS} < h_{roof}$. The model does not consider multipath propagation and the reliability of path loss estimation decreases also if the terrain is not flat or the land-cover is not homogeneous [Kanas, 99].

In general, the CWI model describes adequately the site-specific environment in both LOS and NLOS cases. If the examined environment had flat terrain and the land cover was more homogeneous, the model would provide much better results. Furthermore a significant parameter is that the street layout and the city plan are non-uniform (characteristic in most Greek cities) and this reduces the model reliability according to the required parameters [Kanas, 99].

The WB model is in general designed to provide better results for wide space buildings and for base station antennas placed above the rooftop.

Taking into account these observations, an investigation of the NN applicability to the prediction of propagation path loss was done.

The intention of this work was to develop micro-cell and indoor propagation prediction models with satisfactory accuracy, which relies on a database not difficult to create.

In Table 7.19, the performances, in terms of mean error (μ), standard deviation (σ) and root mean square error (RMS) (in dB), achieved by all the proposed NN models in urban, suburban and indoor environment are indicated.

Table 7.19. Proposed NN models performance in different environments

NN models	Urban						Suburban			Indoor		
	LOS			NLOS			μ	σ	RMS	μ	σ	RMS
μ	σ	RMS	μ	σ	RMS	μ						
MLP-NN	5.08	4.23	6.63	4.61	4.65	6.55	3.01	2.23	3.74	3.05	3.15	4.38
Hybrid MLP-NN	-	-	-	4.14	4.43	6.07	2.86	2.46	3.77	-	-	-
RBF-NN	5.04	4.54	6.78	3.67	3.88	5.35	2.65	2.55	3.68	3.09	2.88	4.23
Hybrid RBF-NN	-	-	-	3.65	3.85	5.30	2.71	2.54	3.71	-	-	-

In general, for both urban and suburban environment, in the prediction made by the NN models is noticed a significant improvement as comparing to the empirical models. The only exception occurs in LOS case, urban environment, the prediction is very close to the one

made by the SSM, the difference in the RMS sense being 0.46 dB in case of MLP-NN and 0.61 dB in case of RBF-NN (both values in NN favor).

In NLOS case, for both urban and suburban environment, an investigation was done on the potential benefits of the combination of physical parameters and a theoretical model, by applying MLP-NN. The improvement obtained in case of urban environment lies around 0.71 dB and in case of suburban environment is 1.11 dB. However, this combination between physical and theoretical parameters had no effect on the prediction done by RBF-NN models.

Besides the simple NN prediction models, an error correction model was studied when the NN is used to compensate for the error introduced by applying CWI model. In case of urban environment, the hybrid MLP-NN has shown an improvement at the order of 0.48 dB comparing to the simple prediction whilst in case of suburban environment the hybrid MLP-NN has the same behavior as the simple prediction model (Table 7.19).

Referring to the simple prediction models, for both urban and suburban environment, as depicted in Table 7.19, a slight improvement was obtained by the RBF-NN models over the MLP-NN models, 1.2 dB in case of urban (NLOS) and 0.06 dB in case of suburban.

In case of hybrid NN models, the hybrid RBF-NN model for urban environment gives an improvement of 0.77 dB comparing to the hybrid MLP-NN whilst there is no significant difference between the prediction done by hybrid MLP-NN and hybrid RBF-NN in suburban environment. These lead to the conclusion that both NN investigated, either as simple prediction models or incorporated in an error control system, give similar results.

RBF-NN models with different inputs, representing physical parameters, were developed and used to evaluate their importance for the prediction of propagation path loss. For both types of environments (urban and suburban) under investigation, by using the street orientation parameter as an input to the RBF-NN, together with the other physical parameters, it was noticed a significant improvement in the prediction made by the both, RBF-NN and hybrid RBF-NN models.

The best results obtained by the MLP-NN trained by LM algorithm were obtained in urban, suburban environments, for the simple prediction models and in indoor case. The proposed hybrid MLP-NN models give better prediction when the NN is trained with PB algorithm (urban environment) and RP algorithm (suburban environment).

In general, the RP algorithm is about 41-43 % time faster than the PB algorithm. The LM algorithm is slower and it was tested only with early stopping method. The time used for training, for the proposed models, ranges between 100 s and 200 s.

Referring to the previous work reported in literature, in case of urban and suburban environment, a summary of the results are presented in the following table (Table 7.20):

Table 7.20. Background studies, outdoor environment

	Environment	Frequency [MHz]	Results [dB]	Observations
[Gschwendtner, 93]	Mannheim (Germany)	943	$\mu = 1.4$ $\sigma = 5.2$	MLP Urban
[Stocker, 93]	Mannheim (Germany)	943	$\mu = 0.20$ $\sigma = 6.20$	MLP Urban
	Darmstadt (Germany)	943	$\mu = 4.9$ $\sigma = 6$	MLP Urban
[Balandier, 95]	Paris [France]	170	$\sigma = 6.73$ (average)	MLP, Urban Simple model
			$\sigma = 4.6$ (average)	Hybrid MLP model Urban
[Gschwendtner, 96]	Mannheim	943	$\mu = 0$ $\sigma = 5.7$	Error correction MLP Micro-cell, urban
[Fraile, 97]	Munich	940	$\sigma = 6.45$ (average)	MLP, Urban
[Chang, 97]	Tokyo		$\sigma = 2.35$	RBF, Urban -cell
[Fraile, 98]	Munich	940	$\mu = 1.779$ $\sigma = 7.345$	MLP, Macro-cell Urban
[Leros, 98]	Athens (LOS)	900	$\mu = 1.079$ $\sigma = 6.888$ RMS = 6.97	MLP Urban
[Nešković, 98]	Belgrade	450 & 900	RMS=5.9 dB (450 MHz) RMS= 6.1 dB (900 MHz).	MLP Urban
[Bargallo, 98]	Missouri/Kansas City 4sites: suburban 1 site: urban		$\mu = 0.37$ $\sigma = 7.6$	MLP
			$\mu = 1.13$ $\sigma = 7.8$	RBF
[Yang, 99]	Seoul	1800	$\sigma_{\text{area1}} = 9.39$	Micro-cell
			$\sigma_{\text{area2}} = 8.36$	

The applications of the NN presented in this work were designed based on measurements performed in micro-cell at 1.89 GHz. Due to that it occurs a difficult in making a comparisons with the other models, due to the different frequencies and different environments where the measurements were collected. For example, the German town of Mannheim is characterized by a regular building structure and almost no variation in terrain

height; in the city of Paris the area is very dense, characterized by irregular street grids, inhomogeneous built-up structures and street widths.

In this work the measurements used to design the proposed NN based models were collected in two different environments, characterized by a non-flat terrain, an inhomogeneous land cover, street layout and city (village) plan are non-uniform. The centre of the city of Kavala is heavily built, mainly with five storey buildings, apartments, shopping centers and shops. Due to the fact that the city is placed between the city and a hill, the elevation varies from one meter to one hundred meters above sea level. Most of the roads in the city centre are narrow and there are only a few four-lane roads. The village of Oia is built on a high peninsula on the island of Santorini. It is a unique Cycladic environment, characteristic in the Greek islands. It has narrow walking paths, except of three main streets. The buildings are mostly stony and two-storeyed and many of them are embedded into the rock, being carved out of the stones with non-canonical shape.

However, making a comparison between the results obtained by the proposed NN models in this work for urban environment with the results presented in [Yang, 99], the improvement obtained in the average standard deviation obtained ranges from 3.92 dB to 4.51 dB.

The hybrid model presented in [Balandier, 95], that includes antenna height parameter as an input to the NN, gives a standard deviation of 4.6 dB. The hybrid models proposed in this work (for urban environment) lead to a standard deviation of 4.43 dB (in case of MLP-NN) and 3.85 dB (in case of RBF-NN).

As far as the indoor environment is concerned, the previous studies are briefly described in section 6.2.

The standard deviation indicated for the prediction done with measurements that were used for training is smaller than 3.3 dB and the standard deviation for measurements not used for training is smaller than 8 dB [Wofle, 97b].

In [Neskovic, 00b] the mean errors of the field strength prediction obtained by the neural network model in two old-style university buildings range between -2 dB to 2.4 dB for one building; -5.3 dB to 1.7 dB for the second building. The standard deviations reported range between 4.7 dB to 7.4 dB for the first building and 3.4 dB to 6.3 dB for the second one.

The performances of the NN model in indoor environment presented in this work are shown in Table 7.19. In contrast to well-known empirical models high accuracy can be obtained, because the NN is trained with measurements inside buildings and thus include realistic propagation effects and also consider parameters, which are difficult to include in

analytic equations. It is difficult to make a comparison between the NN based models proposed in this work and the models proposed by the other authors due to differences in building structure.

7.5. Conclusions

New NN models are presented, through which some important disadvantages of both statistical and deterministic propagation models can be overcome. The proposed NN models are:

1. MLP-NN prediction model in urban environment,
2. MLP-NN prediction model in suburban environment,
3. Hybrid MLP-NN models in urban and suburban environment,
4. MLP-NN model for indoor environment,
5. RBF-NN prediction model in urban environment,
6. RBF-NN prediction model in suburban environment,
7. Hybrid RBF-NN models in urban and suburban environment,
8. RBF-NN model for indoor environment.

Within the proposed models, environmental characteristics are considered more subtly than in standard statistical models, what usually provides greater accuracy of the model. On the other side, the NN models are not computationally extensive as the deterministic models. The implementation of the proposed NN models requires a database that is easy to obtain.

In comparison with other field strength prediction models, the proposed NN models showed very good accuracy. The main advantage of the proposed NN models lays in the fact that the models should be easily adjusted to some specific environments and complex propagation condition. In more specific local cases, the accuracy can be improved by some additional NN training. Results are always connected with some uncertainty but accuracy may be sufficient for prediction purpose.

The results obtained by a pure MLP-NN system used for prediction are very interesting. But the results obtained by a MLP-NN system that combines a nonlinear NN approach, results of classical propagation loss algorithms and physical data open new ways of investigations. The algorithms carry a considerable expert knowledge on terrestrial wave propagation; the use of NN as field strength prediction model allows to efficiently integrate this knowledge as well as topographic and land cover information.

The hybrid modeling approach for the prediction of propagation path loss is studied and it is shown that NN can be used in highly adaptive models. By introducing of additional parameters during the training process even an extension of empirical models is feasible. In contrast to well-known regression algorithms, NNs offer many advantages owing their ability to represent highly nonlinear dependencies of many parameters simultaneously, including information that cannot be treated analytically. In addition, the application of all available information at the same time is a way of getting the most even from poorly defined databases. This flexible and computationally effective approach can be used for calibration and as an extension of conventional prediction models.

The advantage of the NN approaches is that a particular propagation model can be constructed to take account of various types of environments based on measurement data taken in the desired environment. This approach enhances the flexibility of the NN based prediction model to adapt to the terrain database of the environment. Simulation results have shown that the NN approach provides more accurate prediction of field strength loss than that of the empirical models studied in this work. This verifies the effectiveness of the best approximation capability of the NN.

7.6. Further work

- It is proposed NN to be also investigated for the prediction of the delay spread and in general for the prediction of several parameters related to fast fading (time delay, Doppler effect, etc.).
- The collection of measurements at different frequencies and in several types of environments will permit the development of a more general model.
- NN applications in Telecommunications such as:
 - Automatic Modulation Recognition
 - Direction of Arrivals Estimation
 - Call Admission Control

Appendix 1

Method of Steepest Descent

Consider a cost function $E(\mathbf{w})$ that is a continuously differentiable function of some unknown weight (parameter) vector \mathbf{w} . The function $E(\mathbf{w})$ maps the elements of \mathbf{w} into real numbers.

In this method, the successive adjustments applied to the weight vector \mathbf{w} are in the direction of steepest descent, that is, in a direction opposite to the gradient vector $\nabla E(\mathbf{w})$. For convenience of presentation

$$\mathbf{g} = \nabla E(\mathbf{w}) \quad (\text{A.1})$$

Accordingly, the steepest descent algorithm is formally described by

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \mu \mathbf{g}(n) \quad (\text{A.2})$$

where μ is a positive constant called the learning rate parameter and $\mathbf{g}(n)$ is the gradient vector evaluated at the point $\mathbf{w}(n)$. In going from iteration n to $n+1$ the algorithm applies the correction

$$\Delta \mathbf{w}(n) = \mathbf{w}(n+1) - \mathbf{w}(n) = -\mu \mathbf{g}(n) \quad (\text{A.3})$$

To show that the formulation of the steepest descent algorithm satisfies the condition for iterative descent $E(\mathbf{w}(n+1)) < E(\mathbf{w}(n))$, it is used a first-order Taylor series expansion around $\mathbf{w}(n)$ to approximate $E(\mathbf{w}(n+1))$ as

$$E(\mathbf{w}(n+1)) \cong E(\mathbf{w}(n)) + \mathbf{g}^T(n) \Delta \mathbf{w}(n)$$

the use of which is justified for small μ and yields to

$$E(\mathbf{w}(n+1)) \cong E(\mathbf{w}(n)) - \mu \mathbf{g}^T(n) \mathbf{g}(n) = E(\mathbf{w}(n)) - \mu \|\mathbf{g}(n)\|^2 \quad (\text{A.4})$$

which shows that, for a positive learning rate parameter μ , the cost function is decreased as the algorithm progresses from one iteration to the next. The reasoning presented here is approximate in that this end result is only true for small enough learning rates.

The method of steepest descent converges to the optimal solution \mathbf{w}^* slowly. Moreover, the learning rate parameter μ has a profound influence on its convergence behavior:

- When μ is small, the transient response of the algorithm is overdamped, in that the trajectory traced by $\mathbf{w}(n)$ follows a smooth path in the W -plane (weights plane).
- When μ is large, the transient response of the algorithm is underdamped, in that the trajectory of $\mathbf{w}(n)$ follows a zigzagging (oscillatory) path.
- When μ exceeds a certain critical value, the algorithm becomes unstable (i.e., it diverges).

Appendix 2

Newton's method

The basic idea of Newton's method is to minimize the quadratic approximation of the cost function $E(\mathbf{w})$ around the current point $\mathbf{w}(n)$; this minimization is performed at each iteration of the algorithm. Specifically, using a second-order Taylor series expansion of the cost function around the point $\mathbf{w}(n)$, we may write

$$\Delta E(\mathbf{w}(n)) = E(\mathbf{w}(n+1)) - E(\mathbf{w}(n)) \cong \mathbf{g}^T(n)\Delta\mathbf{w}(n) + \frac{1}{2}\Delta\mathbf{w}^T(n)\mathbf{H}(n)\Delta\mathbf{w}(n) \quad (\text{A.5})$$

As before, $\mathbf{g}(n)$ is the m -by-1 gradient vector of the cost function $E(\mathbf{w})$ evaluated at the point $\mathbf{w}(n)$. The matrix $\mathbf{H}(n)$ is the m -by- m Hessian matrix of $E(\mathbf{w})$, also evaluated at $\mathbf{w}(n)$. The Hessian of $E(\mathbf{w})$ is defined by

$$\mathbf{H} = \nabla^2 E(\mathbf{w}) = \begin{bmatrix} \frac{\partial^2 E}{\partial w_1^2} & \frac{\partial^2 E}{\partial w_1 \partial w_2} & \cdots & \frac{\partial^2 E}{\partial w_1 \partial w_m} \\ \frac{\partial^2 E}{\partial w_2 \partial w_1} & \frac{\partial^2 E}{\partial w_2^2} & \cdots & \frac{\partial^2 E}{\partial w_2 \partial w_m} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial^2 E}{\partial w_m \partial w_1} & \frac{\partial^2 E}{\partial w_m \partial w_2} & \cdots & \frac{\partial^2 E}{\partial w_m^2} \end{bmatrix} \quad (\text{A.6})$$

Equation (A.6) requires the cost function $E(\mathbf{w})$ to be twice continuously differentiable with respect to the elements of \mathbf{w} . Differentiating equation (A.5) with respect to $\Delta\mathbf{w}$, the change $\Delta E(\mathbf{w})$ is minimized when

$$\mathbf{g}(n) + \mathbf{H}(n)\Delta\mathbf{w}(n) = 0 \quad (\text{A.7})$$

Solving this equation for $\Delta\mathbf{w}(n)$ yields

$$\Delta\mathbf{w}(n) = -\mathbf{H}^{-1}(n)\mathbf{g}(n) \quad (\text{A.8})$$

That is,

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \Delta\mathbf{w}(n) = \mathbf{w}(n) - \mathbf{H}^{-1}(n)\mathbf{g}(n) \quad (\text{A.9})$$

where $\mathbf{H}^{-1}(n)$ is the inverse of the Hessian of $E(\mathbf{w})$.

Generally speaking, Newton's method converges quickly asymptotically and does not exhibit the zigzagging behavior that sometimes characterizes the method of steepest descent. However, for Newton's method to work, the Hessian $\mathbf{H}(n)$ has to be a positive definite matrix for all n . Unfortunately, in general there is no guarantee that $\mathbf{H}(n)$ is positive definite at every iteration of the algorithm. If the Hessian $\mathbf{H}(n)$ is not positive definite, modification of Newton's method is necessary [Haykin, 99].

Appendix 3

Gauss-Newton method

The Gauss-Newton method is applicable to a cost function that is expressed as the sum of error squares. Let

$$E(\mathbf{w}) = \frac{1}{2} \sum_{i=1}^N e^2(i) \quad (\text{A.10})$$

where the scaling factor 1/2 is included to simplify matters in subsequent analysis. All the error terms in this formula are calculated on the basis of a weight vector \mathbf{w} that is fixed over the entire observation interval $1 \leq i \leq n$.

The error signal $e(i)$ is a function of the adjustable weight vector \mathbf{w} . Given an operating point $\mathbf{w}(n)$, the dependence of $e(i)$ on \mathbf{w} is linearized by writing

$$e'(i, \mathbf{w}) = e(i) + \left[\frac{\partial e(i)}{\partial \mathbf{w}} \right]_{\mathbf{w}=\mathbf{w}(n)}^T (\mathbf{w} - \mathbf{w}(n)) \quad i = 1, 2, \dots, n \quad (\text{A.11})$$

Equivalently, by using matrix notation we may write

$$\mathbf{e}'(n, \mathbf{w}) = \mathbf{e}(n) + \mathbf{J}(n)(\mathbf{w} - \mathbf{w}(n)) \quad (\text{A.12})$$

where $\mathbf{e}(n)$ is the error vector

$$\mathbf{e}(n) = [e(1), e(2), \dots, e(n)]^T \quad (\text{A.13})$$

and $\mathbf{J}(n)$ is the n -by- m Jacobian matrix of $\mathbf{e}(n)$:

$$\mathbf{J}(n) = \left[\begin{array}{cccc} \frac{\partial e(1)}{\partial w_1} & \frac{\partial e(1)}{\partial w_2} & \dots & \frac{\partial e(1)}{\partial w_m} \\ \frac{\partial e(2)}{\partial w_1} & \frac{\partial e(2)}{\partial w_2} & \dots & \frac{\partial e(2)}{\partial w_m} \\ \vdots & \vdots & & \\ \frac{\partial e(n)}{\partial w_1} & \frac{\partial e(n)}{\partial w_2} & \dots & \frac{\partial e(n)}{\partial w_m} \end{array} \right]_{\mathbf{w}=\mathbf{w}(n)} \quad (\text{A.14})$$

The Jacobian $\mathbf{J}(n)$ is the transpose of the m -by- n gradient matrix $\nabla \mathbf{e}(n)$, where

$$\nabla \mathbf{e}(n) = [\nabla e(1), \nabla e(2), \dots, \nabla e(n)] \quad (\text{A.15})$$

The updated weight vector $\mathbf{w}(n+1)$ is then defined by

$$\mathbf{w}(n+1) = \arg \min_{\mathbf{w}} \left\{ \frac{1}{2} \|\mathbf{e}'(n, \mathbf{w})\|^2 \right\} \quad (\text{A.16})$$

Using equation (A.12) to evaluate the squared Euclidean norm of $\mathbf{e}'(n, \mathbf{w})$, we get

$$\begin{aligned} \frac{1}{2} \|\mathbf{e}'(n, \mathbf{w})\|^2 &= \frac{1}{2} \|\mathbf{e}(n)\|^2 + \mathbf{e}^T(n) \mathbf{J}(n) (\mathbf{w} - \mathbf{w}(n)) \\ &\quad + \frac{1}{2} (\mathbf{w} - \mathbf{w}(n))^T \mathbf{J}^T(n) \mathbf{J}(n) (\mathbf{w} - \mathbf{w}(n)) \end{aligned} \quad (\text{A.17})$$

Hence, differentiating this expression with respect to \mathbf{w} and setting the result equal to zero, we obtain

$$\mathbf{J}^T(\mathbf{n})\mathbf{e}(\mathbf{n}) + \mathbf{J}^T(\mathbf{n})\mathbf{J}(\mathbf{n})(\mathbf{w} - \mathbf{w}(\mathbf{n})) = 0 \quad (\text{A.18})$$

Solving this equation for \mathbf{w} , we may thus write in light of equation (A.16):

$$\mathbf{w}(\mathbf{n} + 1) = \mathbf{w}(\mathbf{n}) - (\mathbf{J}^T(\mathbf{n})\mathbf{J}(\mathbf{n}))^{-1} \mathbf{J}^T(\mathbf{n})\mathbf{e}(\mathbf{n}) = 0 \quad (\text{A.19})$$

which describes the pure form of the Gauss-Newton method.

Unlike Newton's method that requires knowledge of the Hessian matrix of the cost function $E(\mathbf{n})$, the Gauss-Newton method only requires the Jacobian matrix of the error vector $\mathbf{e}(\mathbf{n})$. However, for the Gauss-Newton iteration to be computable, the matrix product $\mathbf{J}^T(\mathbf{n})\mathbf{J}(\mathbf{n})$ must be nonsingular.

With regard to the latter point, we recognize that $\mathbf{J}^T(\mathbf{n})\mathbf{J}(\mathbf{n})$ is always nonnegative definite. To ensure that it is nonsingular, the Jacobian $\mathbf{J}(\mathbf{n})$ must have row rank n ; that is, the n rows of $\mathbf{J}(\mathbf{n})$ in equation (A.14) must be linearly independent. To guard against the possibility that $\mathbf{J}(\mathbf{n})$ is rank deficient, the customary practice is to add the diagonal matrix $\delta\mathbf{I}$ to the matrix $\mathbf{J}^T(\mathbf{n})\mathbf{J}(\mathbf{n})$. The parameter δ is a small positive constant chosen to ensure that $\mathbf{J}^T(\mathbf{n})\mathbf{J}(\mathbf{n}) + \delta\mathbf{I}$ is positive definite for all n . On this basis, the Gauss-Newton method is implemented in the slightly modified form:

$$\mathbf{w}(\mathbf{n} + 1) = \mathbf{w}(\mathbf{n}) - (\mathbf{J}^T(\mathbf{n})\mathbf{J}(\mathbf{n}) + \delta\mathbf{I})^{-1} \mathbf{J}^T(\mathbf{n})\mathbf{e}(\mathbf{n}) \quad (\text{A.20})$$

The effect of this modification is progressively reduced as the number of iterations, n , is increased.

Note also that the recursive equation (A.20) is the solution of the modified cost function:

$$E(\mathbf{w}) = \frac{1}{2} \left\{ \delta \|\mathbf{w} - \mathbf{w}(0)\|^2 + \sum_{i=1}^n e^2(i) \right\} \quad (\text{A.21})$$

where $\mathbf{w}(0)$ is the initial value of the weight vector $\mathbf{w}(i)$.

We are now equipped with the optimization tools we need to address the specific issues involved in linear adaptive filtering.

Appendix 4

Least-mean-square algorithm

The least-mean-square (LMS) algorithm is based on the use of instantaneous values for the cost function, namely

$$E(\mathbf{w}) = \frac{1}{2} e^2(n) \quad (\text{A.22})$$

where $e(n)$ is the error signal measured at time n . Differentiating $E(\mathbf{w})$ with respect to the weight vector \mathbf{w} yields

$$\frac{\partial E(\mathbf{w})}{\partial \mathbf{w}} = e(n) \frac{\partial e(n)}{\partial \mathbf{w}} \quad (\text{A.23})$$

As with linear least-squares filter, the LMS algorithm operates with a linear neuron so we may express the error signal as

$$e(n) = d(n) - \mathbf{x}^T(n) \mathbf{w}(n) \quad (\text{A.24})$$

Hence,

$$\frac{\partial e(n)}{\partial \mathbf{w}} = -\mathbf{x}(n) \quad (\text{A.25})$$

and

$$\frac{\partial E(\mathbf{w})}{\partial \mathbf{w}(n)} = -\mathbf{x}(n) e(n) \quad (\text{A.26})$$

Using this latter result as an estimate for the gradient vector, we may write

$$\hat{\mathbf{g}} = -\mathbf{x}(n) e(n) \quad (\text{A.27})$$

Finally, using equation (A.27) for the gradient vector for the method of steepest descent, the LMS algorithm may be formulated as follows:

$$\hat{\mathbf{w}}(n+1) = \hat{\mathbf{w}}(n) + \mu \mathbf{x}(n) e(n) \quad (\text{A.28})$$

where μ is the learning rate parameter. The feedback loop around the weight vector $\hat{\mathbf{w}}(n)$ in the LMS algorithm behaves like a low-pass filter, passing the low frequency components of the error signal and attenuating its high frequency components [Haykin, 96]. The average time constant of this filtering action is inversely proportional to the learning rate parameter μ . Hence, by assigning a small value to μ , the adaptive process will progress slowly.

In equation (A.28) it has been used $\hat{\mathbf{w}}(n)$ in place of $\mathbf{w}(n)$ to emphasize the fact that the LMS algorithm produces an estimate of the weight vector that would result from the use of method of steepest descent. As a consequence, in using the LMS algorithm a distinct feature of the steepest descent algorithm is sacrificed. In the steepest descent algorithm the weight vector $\mathbf{w}(n)$ follows a

well-defined trajectory in weight space for a prescribed μ . In contrast, in the LMS algorithm the weight vector $\hat{\mathbf{w}}(n)$ traces a random trajectory. For this reason, the LMS algorithm is sometimes referred to as a “stochastic gradient algorithm”. As the number of iterations in the LMS algorithm approaches infinity, $\hat{\mathbf{w}}(n)$ performs a random walk (Brownian motion) about the Wiener solution \mathbf{w}_0 . The important point is the fact that, unlike the method of steepest descent, the LMS algorithm does not require knowledge of the statistics of the environment.

A summary of the LMS algorithm is presented in Table A.1 that clearly illustrates the simplicity of the algorithm.

Table A.1 Summary of the LMS algorithm

Training sample: Input signal vector = $\mathbf{x}(n)$

 Desired response = $d(n)$

User-selected parameter: μ

Initialization. Set $\hat{\mathbf{w}}(0) = 0$

Computation. For $n = 1, 2, \dots$, compute

$$e(n) = d(n) - \hat{\mathbf{w}}^T(n)\mathbf{x}(n)$$

$$\hat{\mathbf{w}}(n+1) = \hat{\mathbf{w}}(n) + \mu \mathbf{x}(n)e(n)$$

References

- [**Akeyama, 82**] A. Akeyama, T. Nagatsu, Y. Ebine, "Mobile radio propagation characteristics and radio zone design method in local cities", *Review Elec. Commun. Lab.*, 30, 1982, pp. 308-317
- [**Andersen, 94**] J. B. Andersen, "Transition zone diffraction by multiple edges", *IEE Proc. Microwaves, Antennas and Propagation*, vol. 141, no. 5, October 1994, pp. 382 - 384
- [**Andersen, 96**] J. B. Andersen, J. Wiart, "UTD transition zone diffraction", COST 231 TD(96)3, Belfort, January 1996
- [**Anderson, 58**] L. J. Anderson, L. G. Trolese, "Simplified method for computing knife-edge diffraction in the shadow region", *IRE Trans. Ant. and Prop.*, 6, 1958, pp. 281 - 287
- [**Aurand, 85**] J. F. Aurand, R. E. Post, "A comparison of prediction models for 800 MHz mobile radio propagation", *IEEE Transactions on Vehicular Technology*, VT-34, no. 4, 1985, pp. 149 – 153
- [**Balandier, 95**] T. Balandier, A. Caminada, V. Lemoine, F. Alexandre "170 MHz Field Strength Prediction in Urban Environment Using Neural Nets" *Personal, Indoor and Mobile Radio Communications, 1995. PIMRC'95. Wireless: Merging onto the Information Superhighway. Sixth IEEE International Symposium on, Volume: 1, 1995*, pp: 120 –124.
- [**Bargallo, 98**] J. M. Bargallo, "A comparison of Neural Network models for prediction of RF propagation loss", *Vehicular Technology Conference, VTC 98, 48th IEEE*, pp: 445 –449
- [**Beale, 72**] E. M. L. Beale, "A derivation of conjugate gradients", F. A. Lootsma, ed. *Numerical methods for nonlinear optimization*, London: Academic Press, 1972
- [**Bergljung, 89**] C. Bergljung et al., "Micro-cell Radio Channel", *Preliminary Report on Indoor Field-strength Measurements at 900 and 1700 MHz, RMTP/RB/J031*
- [**Bertoni, 94**] H. L. Bertoni, W. Honcharenko, L. R. Maciel, H. H. Xia, "UHF propagation prediction for wireless personal communications", *Proceedings of the IEEE*, vol. 82, no. 9, pp. 1333 – 1359, 1994
- [**Broomhead, 88**] D. S. Broomhead, D. Lowe, "Multivariable functional interpolation and adaptive networks", *Complex Systems*, vol. 2, pp. 321 - 355
- [**Causebrook, 71**] J. H. Causebrook, B. Davies, "Tropospheric radio wave propagation over irregular terrain: the computation of field strength for UHF broadcasting", *BBC Research Report No. 43*, 1971
- [**Causebrook, 90**] J. H. Causebrook, "Signal attenuation by buildings and trees", COST 231 TD(90)106, Darmstadt, 1990
- [**Chang, 97**] P. R. Chang, W. H. Yang, "Environment-Adaptation Mobile Radio Propagation Prediction Using Radial Basis Function Neural Networks" *IEEE Transactions on VT*, vol. 46, No. 1, Feb. 1997.
- [**Chen, 95**] S. Chen, "Nonlinear time series modelling and prediction using Gaussian RBF networks with enhanced clustering and RLS learning", *Electronic Letters*, vol. 31, No. 2, pp. 117 – 118, 1995

- [**Chinrungrueng, 94**] C. Chinrungrueng, C. H. Sequin, "Optimal adaptive k-means algorithm with dynamic adjustment of learning rate", *IEEE Transactions on Neural Networks*, vol. 6, pp. 157 – 169, 1994
- [**Christodoulou, 01**] C. Christodoulou and M. Georgiopoulos, "Applications of Neural Networks in Electromagnetics", Artech House, 2001
- [**Cichon, 94a**] D. J. Cichon, T. C. Becker, W. Wiesbeck, "A ray launching approach for indoor and outdoor applications", COST 231 TD(94)32, Lisbon, Portugal, January 1994
- [**Cichon, 94b**] D. J. Cichon, W. Wiesbeck, "Indoor and outdoor propagation modeling in picocells", *Proc. Personal, Indoor and Mobile Radio Conference PIMRC'94*, pp. 491-495, The Hague, The Netherlands, September 1994
- [**Cichon, 95**] D. J. Cichon, T. Zwick, J. Lähteenmäki, "Ray optical indoor modeling in multifloored buildings: simulations and measurements", *Proc. IEEE Intern. Antennas and Propagation Symposium AP-S'95*, Newport Beach, USA, 1995, pp. 522-525
- [**Constantinou, 94**] P. Constantinou, "Properties of Wireless RF Channels", *Wireless LAN systems*, Artech House, pp. 129-163, 1994
- [**COST 207**] "COST 207 – Digital land mobile radio communications", Final report, Commission of the European Communities, 1989
- [**COST231, 99**] "COST Action 231 – Digital mobile radio towards future generation systems", Final report, European Communities, 1999
- [**Cox, 84**] D. C. Cox, R. R. Murray, A. W. Norris, "800 MHz Attenuation Measures in and Around Suburban houses", *AT&T, BLTJ*, vol. 63, July–August 1984
- [**Delisle, 85**] G. Y. Delisle, J-P. L. Lefèvre, M. Lecours, J-Y. Chouinard, "Propagation loss prediction: a comparative study to the mobile radio channel", *IEEE Transactions on Vehicular Technology*, vol. VT-34, no. 2, May 1985, pp. 86 - 96
- [**Deygout, 66**] J. Deygout, "Multiple knife-edge diffraction of microwaves", *IEEE Trans.*, AP-14, no. 4, pp. 480-489, 1966
- [**Dougherty, 64**] H. T. Dougherty, L. J. Maloney, "Applications of diffraction by convex surfaces to irregular terrain situation", *Radio Science*, 68 – D, no. 2, 1964, pp. 284 - 305
- [**Duda, 73**] R. O. Duda, P. E. Hart, "Pattern classification and Scene Analysis", New York: Wiley, 1973
- [**Egli, 57**] J. J. Egli, "Radio propagation above 40 Mc over irregular terrain", *Proc. IRE*, 45, No. 10, pp. 1383 - 1391, 1957
- [**Feuerstein, 94**] M. J. Feuerstein, K. L. Blackard, T. S. Rappaport, S. Y. Seidel, H. H. Xia, "Path Loss Delay Spread, and Outage Models as Functions of Antenna Height for Microcellular System Design", *IEEE Transactions on Vehicular Technology*, Vol. 43, No. 3, pp. 487-498, August 1994
- [**Fraile, 97**] R. Fraile, N. Cardona, "Fast Neural Network Method for Propagation Loss Prediction in Urban Environments", *ELECTRONICS LETTERS* 20th November 1997, vol. 33, No. 24 pp. 2056-2058.

- [**Fraile, 98**] R. Fraile, N. Cardona, “Macrocellular Coverage Prediction for all Ranges of Antenna Height using Neural Networks”, Universal Personal Communications, 1998. ICUPC '98. IEEE 1998 International Conference on, vol. I, pp: 21 -25 vol.1
- [**Gschwendtner, 93**] B.E. Gschwendtner, F. M. Landstorfer, “An application of Neural Networks to the prediction of terrestrial wave propagation”, IEE 8th International Conference on Antennas and Propagation, April 1993, pp.804-807
- [**Gschwendtner, 96a**] B. E. Gschwendtner, F. M. Landstorfer, “Adaptive propagation modeling using a hybrid neural technique”, Electronics Letters, Feb. 1996, vol. 32, No. 3, pp. 162-164.
- [**Gschwendtner, 96b**] B. Gschwendtner, F. Landstorfer, “Adaptive Propagation Modelling Based on Neural Network Techniques”, Vehicular Technology Conference, 1996. Mobile Technology for the Human Race, IEEE 46th, vol. 2, pp: 623 –626.
- [**Hata, 80**] M. Hata, “Empirical formula for propagation loss in land mobile radio services”, IEEE Trans. Veh. Technol., vol. VT-29, no. 3, pp. 317-325, Aug. 1980.
- [**Hacking, 68**] K. Hacking, “Propagation over rounded hills”, BBC Research Report RA-21, 1968
- [**Harley, 89**] P. Harley, “Short distance attenuation measurements at 900 MHz and 1.8 GHz using low antenna heights for microcells”, IEEE Jurnal on Selected Areas in Communications, vol. 7, no. 1, Jan. 1989
- [**Haykin, 94**] S. Haykin, “Neural Networks: A Comprehensive Foundation”, IEEE Press, McMillan College Publishing Co., 1994
- [**Haykin, 96**] S. Haykin, “Adaptive filter theory”, Prentice-Hall, 1996
- [**Haykin, 99**] S. Haykin, “Neural Networks. A Comprehensive Foundation”, Prentice Hall, 1999
- [**Honcharenko, 92**] W. Honcharenko, H. L. Bertoni, J. L. Dailing, J. Qian, H. D. Yee, “Mechanisms governing UHF propagation on single floors in modern office buildings”, IEEE Trans. on Vehicular Technology, vol. 41, no. 4, pp. 496-504, 1992
- [**Huang, 91**] S. C. Huang, Y. F. Huang, “Bound on the number of the hidden neurons in multilayer perceptrons”, IEEE Trans. On Neural Networks, vol.2, pp. 47-55, 1991
- [**Huschka, 94**] T. Huschka, “Ray-tracing models for indoor environments and their computational complexity”, Proc. PIMRC'94, pp. 486-490, 1994
- [**Hush, 89**] D. R. Hush, “Classification with neural networks, a performance analysis”, Proc. IEEE International Conference on Systems Engineering, pp. 277-280, 1989.
- [**Ikegami, 84**] F. Ikegami, S. Yoshida, M. Umehira, “Propagation factors controlling mean field strength on urban streets”, IEEE Trans. On Antennas and Propagation, vol. 32, no.8. August 1984, pp. 822-829
- [**ITU, 94**] ITU Recommendation ITU – R 526 – 2, 1994 PN Series Volume “Propagation in non – ionized media”, pp. 129 – 149, 1994
- [**Jakoby, 95**] R. Jakoby, U. Liebenow, “Modelling of radiowave propagation in microcells”, Proc. Intern. Conference on Antennas and Propagation ICAP'95, Eindhoven, The Netherlands, pp. 377 - 380
- [**Jakes, 74**] W. C. Jakes, “Microwave mobile communications”, John Wiley & Sons, Inc., 1974

- [**Kaji, 85**] M. Kaji, A. Akeyama, "UHF – band propagation characteristics for land mobile radio", AP-S Int. Symp. 1985, Symposium digest, Vol. 2, pp. 835 – 838
- [**Kanatas, 99**] A. Kanatas, N. Moraitis, C. Steriadis, N. Papadakis, E. Angelou, P. Constantinou, D. G. Xenikos, A. Vorvolakos, "Measurements and channel characterization at 1.89 GHz in urban and suburban environments", The Second Internat. Symp. on Wireless Personal Multimedia Commun., WPMC 99, Amsterdam, The Netherlands, pp. 448 – 454, Sept. 1999.
- [**Kauschke, 94**] U. Kauschke, "Propagation and system performance simulations for the short range DECT system in microcellular urban roads", IEEE Transactions on Vehicular Technology, vol. 44, no. 2, pp. 253 – 260, November 1994
- [**Kearns, 96**] M. Kearns, "A bound on the error of cross validation using the approximation and estimation rates, with consequences for the training-test split", Advances in Neural Information Processing Systems, vol. 8, pp. 183-189, Cambridge, MA:MIT Press
- [**Keller, 62**] J. B. Keller, "Geometrical theory of diffraction", J. Opt. Soc. Am., vol. 52, pp. 116-130, 1962
- [**King, 77**] B. G. King, "A 60 GHz short – hop 50 Megabit digital radio system", Eascon 77, pp. 24-1A-24-1D, 1977
- [**Kohonen, 90**] T. Kohonen, "The Self-Organizing Map", Proceedings IEEE, Vol. 78, No. 9, 1990, pp. 1464-1480
- [**Kouyoumijan, 74**] R. G. Kouyoumijan and P. H. Pathak, "A uniform theory of diffraction for an edge in a perfectly conducting surface", Proc. IEEE, vol. 62, Nov. 1974, pp. 1448-1461
- [**Kreuzgruber, 93**] P. Kreuzgruber, P. Unterberger, R. Gahleitner, "A ray splitting model for indoor radio propagation associated with complex geometrics", Proc. IEEE VTC'93, NJ, USA, 1993, pp. 227-230
- [**Kreuzgruber, 94**] P. Kreuzgruber, T. Bründl, W. Kuran, R. Gahleitner, "Prediction of indoor radio propagation with the ray splitting model including edge diffraction and rough surfaces", COST – 231 TD(94)50, Prague, Czech republic, April 1994
- [**Lafortune, 90**] J. F. Lafortune, M. Lecours, "Measurements and Modeling of Propagation Losses in a Building at 900 MHz", IEEE Transactions on Vehicular Technology, vol. 39, no. 2, May 1990
- [**Lähteenmäki, 90**] J. Lähteenmäki, S. Ruiz – Boqué, "Indoor propagation models for 1.7 – 1.9 GHz", COST 231(90)114, Darmstadt, December 1990
- [**LeCun, 93**] Y. LeCun, "Efficient learning and second-order methods", A Tutorial at NIPS 93, Denver, 1993
- [**Lee, 93**] W. C. Y. Lee, "Mobile Communications Design Fundamentals", John Wiley, 1993
- [**Leros, 98**] A. P. Leros, A. A. Alexandridis, K. Dangakis, P. Kostarakis, "Evaluation of Radio Propagation Parameters for Field Strength Prediction using Neural Networks", Conference on Antennas and Propagation for Wireless Communications, 1998. IEEE-APS, 1998, pp. 17 –20.
- [**Li, 96**] J. Li, J. F. Wagen, E. Lachat, "On the ITU model for multi – knife edge diffraction", COST 231 TD(96)25 Belfort, January 1996

- [**Lippmann, 87**] R. Lippmann, "An introduction to computing with neural networks", IEEE Accoustics, Speech and Signal Processing, vol. 4, no. 2, pp. 4-22, April 1987.
- [**Lotse, 90**] F. Lotse and A. Wejke, "Propagation measurements for microcells in central Stockholm", IEEE Vehic. Tech. Conf., 1990
- [**Luebbers, 84**] R. J. Luebbers, "Finite conductivity uniform GTD versus knife-edge diffraction in prediction of propagation path loss", IEEE Trans. Ant. and Prop., vol. AP-32, Jan. 1984, pp. 70-72
- [**Makhoul, 89**] J. Makhoul, A. El-Jaroudi, R. Schwartz, "Formation of disconnected decision regions with a single hidden layer", Proc. Int. Joint Conf. on Neural Networks, vol. 1, pp. 455-460, 1989.
- [**McNamara, 90**] D. A. McNamara, C. W. I. Pistorius, J. A. G. Malherbe, "Introduction to the Uniform Geometrical Theory of Diffraction", Artech House Norwood, MA, USA, 1990
- [**Mitchelli, 86**] C. A. Mitchelli, "Interpolation of Scattered Data: Distance Matrices and Conditionally Positive Definite Functions", Constructive Approximation, Vol. 2, 1986, pp. 11-22
- [**Moody, 89**] J. E. Moody, C. J. Darken, "fast Learning in Networks of Locally Tuned Processing Units", Neural Computation, Vol. 1, 1989, pp. 281-294
- [**Neskovic, 97**] N. Neskovic, "The field strength prediction for urban microcells (band 900 MHz)", Master's Thesis, Faculty of Electrical Engineering – Belgrade, 1997
- [**Nešković, 98**] Aleksandar Nešković, Natasa Nešković, Dorde Paunovic, "A Field Strength Prediction Model Based on Artificial Neural Networks", Proc. 9th IEEE Med. Electro. Conf. MELECON, Tel-Aviv, Israel, May 1998, pp. 420-424
- [**Neskovic, 00a**] N. Neskovic, A. Neskovic and D. Paunovic, "A new microcell prediction model based on the arrangement of the streets and their types", Proc. 10th IEEE Med. Electrotech. Conf.-MELECON, Limassol, Cyprus, May 2000
- [**Neskovic, 00b**] A. Neskovic, N. Neskovic and D. Paunovic, "Indoor electric field level prediction model based on the artificial neural networks", IEEE Communications Letters, vol. 4, no. 6, June 2000, pp. 190-192
- [**Okumura, 68**] Y. Okumura et al., "Field strength and its variability in VHF and UHF land-mobile radio service", Review. Elec. Commun. Labs., vol. 16, Sept.-Oct.1968, pp. 825-837.
- [**Papadakis, 98**] N. Papadakis, A. G. Kanatas, E. Angelou, N. Moraitis, P. Constantinou, "Indoor mobile radio channel measurements and characterization for DECT picocells", Third IEEE Symposium on Computers and Communications, Athens, Greece, ISCC '98.
- [**Park, 93**] J. Park, I. W. Sandberg, "Approximation and Radial Basis Function Networks", Neural Computation, Vol. 5, 1993, pp. 305-316
- [**Parsons, 92**] J. D. Parsons, "The Mobile Radio Propagation Channel", John Wiley & Sons, New York, 1992.
- [**Popescu, 00a**] I. Popescu, M. Nafornită, A. Voulgarelis, P. Constantinou: "Applications of Neural Networks to the Prediction of Propagation Path Loss for Mobile Communication Systems", 2nd Symposium on Applied Electromagnetism of Trans Black Sea Region, Xanthi, Greece, 27-29 June, 2000

- [**Popescu, 00b**] I. Popescu, I. Nafoarniță, P. Constantinou, A. Kanatas, N. Moraitis, "Prediction of outdoor propagation path loss with neural networks", Proceedings of 8th Telecommunications Forum Telfor'2000, 21-23 Nov. 2000, Belgrade, Serbia, Yugoslavia, pp. 140-143
- [**Popescu, 01a**] I. Popescu, I. Nafoarniță, P. Constantinou, A. Kanatas, N. Moraitis, "Neural Networks Applications for the Prediction of Propagation Path Loss in Urban Environment", The IEEE Semiannual Vehicular Technology Conference, VTC 2001 Spring, Rhodes, Greece, May 6-9.
- [**Popescu, 01b**] I. Popescu, I. Nafoarniță, P. Constantinou, A. Kanatas, N. Moraitis, "Neural model for path loss prediction in suburban environment", accepted at IASTED International Conference on Advances in Communications, July 3-6, 2001, Rhodes, Greece.
- [**Popescu, 01c**] I. Popescu, I. Nafoarniță, A. Kanatas, P. Constantinou, "Hybrid model based on neural network for path loss prediction", accepted at IASTED International Conference on Advances in Communications, July 3-6, 2001, Rhodes, Greece.
- [**Powell, 77**] M. J. D. Powell, "Restart procedures for the conjugate gradient method", Mathematical Programming, vol. 12, pp. 241-254, 1977
- [**Rappaport, 96**] T. S. Rappaport, "Wireless Communications. Principles and practice", Prentice Hall PTR, 1996
- [**Reudink, 72**] D. O. Reudink, "Comparison of radio transmission at X-band frequencies in suburban and urban areas", IEEE Trans. AP-20, 1972, pp. 470 - 473
- [**Rizk, 94**] K. Rizk, J. F. Wagen, S. Bokhari, F. Gardiol, "Comparison between the diffraction theory and the moment method for the scattering by surface irregularities in a building wall", Journées Internationales de Nice sur les Antennes (JINA'94), Nice, France, Nov. 1994
- [**Rosenblatt, 58**] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain", Psychological Review, vol. 65, pp. 386-408
- [**Rumelhart, 85**] D. E. Rumelhart, D. Zipser, "Feature discovery by competitive learning", Cognitive Science, vol. 9, 1985, pp. 75 - 112
- [**Rumelhart, 86**] D. E. Rumelhart, J. L. McClelland, "Parallel distributed processing: Explorations in the microstructure of cognition", 1986, vol. 1, Cambridge, MA: MIT Press
- [**Rustako, 89**] A. J. Rustako, N. Amitay, G. J. Owens, R. S. Roman, "Propagation measurements at microwave frequencies for microcellular mobile and personal communications", 93th IEEE Vehicular Technology Conference, San Francisco, 1989, pp. 316 - 320
- [**Saunders, 91**] S. R. Saunders, F. R. Bonar, "Explicit multiple building diffraction attenuation function for mobile radiowave propagation", Electron. Letters, 1991, 27 (14), pp. 1276-1277.
- [**Schalkoff, 97**] R. J. Schalkoff, "Artificial Neural Networks", McGraw-Hill, 1997
- [**Schaubach, 92**] K. R. Schaubach, N. J. Davis, T. S. Rappaport, "A ray tracing method for predicting path loss and delay spread in microcellular environments", Proc. IEEE VTC'92, pp. 932-935, Denver, 1992

- [Seidel, 92] S. Y. Seidel, T. S. Rappaport, "914 MHz path loss prediction models for indoor wireless communications in multifloored buildings", IEEE Transactions on Antennas and Propagation, vol. 40, no. 2, pp. 207-217, February 1992
- [Sejnowski, 77a] T. J. Sejnowski, "Strong covariance with nonlinearly interacting neurons", Journal of Mathematical Biology, vol. 4, 1977, pp. 303 – 321
- [Sejnowski, 77b] T. J. Sejnowski, "Statistical constraints on synaptic plasticity", Journal of Theoretical Biology, vol. 69, 1977, pp. 385 - 389
- [Stocker, 92] K. E. Stocker, F. M. Landstorfer, "Empirical prediction of radiowave propagation by neural network simulator" Electronics Letters, April 1992, Vol. 28, No. 8, pp. 724 - 726
- [Stocker, 93] K. E. Stocker, B.E. Gschwendtner, F. M. Landstorfer, "Neural Network approach to prediction of terrestrial wave propagation for mobile radio", IEE Proceedings-H, Vol. 140, No. 4, August 1993
- [Yang, 99] Seomin Yang, "Feature Extraction for Neural Network Wave Propagation Loss Models from Field Measurements and Digital Elevation Map", IEICE Trans. Electron., vol. E82-C, No. 7, July 1999, pp. 1260 – 1266.
- [Tan, 96] S. Y. Tan and H. S. Tan, "A microcellular communications propagation model based on the Uniform Theory of Diffraction and Multiple Image Theory", IEEE Trans. Ant. and Prop., vol. 44, no. 10, Oct. 1996
- [Tou, 74] J. T. Tou, R. C. Gonzalez, "Pattern Recognition Principles", Reading, MA: Addison-Wesley, 1974
- [Törnevik, 93] C. Törnevik, J. E. Berg, F. Lotse, "900 MHz propagation measurements and path loss models for different indoor environments", Proc. IEEE VTC 1993, New Jersey, USA, 1993.
- [Vogler, 82] L. Vogler, "An attenuation function for multiple knife edge diffraction", Radio Science, 17, 1982, pp. 1541 – 1546
- [Wagen, 93] J. F. Wagen, M. Keer, "Comparison of diffraction coefficients for propagation prediction in microcells", COST – 231, TD(93)80, Grimstad, Norway, May 1993
- [Walfisch, 88] J. Walfisch, H. L. Bertoni, "A theoretical model of UHF propagation in urban environments", IEEE Trans. On Antennas and Propagation, vol. 36, no. 12, pp. 1788-1796, Dec. 1988.
- [Wettscherek, 92] D. Wettscherek, T. Dietterich, "Improving the Performance of Radial Basis Function Networks by Learning Center Locations", Advances in Neural Information Processing Systems 4, J. E. Moody, S. J. Hanson and R. P. Lippman (eds.), San Mateo, CA: Morgan Kaufmann, 1992, pp. 1133-1140.
- [Wuart, 95] J. Wuart, P. Metton, "Terrain influence on line of sight microcellular propagation", COST 231 TD(95)78, Florence, Italy, April 1995
- [Widrow, 60] B. Widrow, M. A. Lehr, "30 years of adaptive neural networks: Perceptron, madaline and backpropagation", Proceedings of the Institute of Electrical and Electronics Engineers, vol. 78, 1960, pp. 1415 - 1442

- [Williamson, 84] A. G. Williamson, B. Egan, J. W. Chester, "Mobile radio propagation in Auckland at 851 MHz", *Electron. Lett.*, 20, No. 12, 1984, pp. 517 – 518
- [Whitaker, 88] J. H. Whitaker, "Measurements of path loss at 910 MHz for proposed microcell urban mobile system", *IEEE Trans. VT-37*, No. 3, pp. 125 – 129, 1988
- [Wolfe, 97a] G. Wolfe, F. M. Landstorfer "A recursive model for the field strength prediction with neural networks" *IEE 10th Conference on Antennas and Propagation (ICAP) 1997*, Edingburgh
- [Wolfe, 97b] G. Wolfe, F. M. Landstorfer, "Field strength prediction in indoor environments with neural networks" *IEEE 47th Vehicular Technology Conference (VTC)*, May 1997, Phoenix, Arizona
- [Wolfe, 97c] G. Wolfe, F. M. Landstorfer "Field strength prediction with dominant paths and neural networks for indoor mobile communications", *MIOP/'97*
- [Wolfe, 97d] G. Wolfe, F. M. Landstorfer, R. Gahleitner, E. Bonek, "Extensions to the field strength prediction technique based on dominant paths between transmitter and receiver in indoor wireless communications, *EPMCC*, September 1997
- [Wolfe, 98] G. Wolfe, F. M. Landstorfer, "Dominant paths for the field strength prediction", *48th IEEE International Conference on Vehicular Technology*, Ottawa, May 1998.
- [Wolfe, 99] G. Wolfe, F. M. Landstorfer "Prediction of the field strength inside buildings with empirical, neural and ray-optical prediction models", *COST 259*, 1999
- [Xia, 93] H. H. Xia et al., "Radio propagation characteristics for line-of-sight microcellular and Personal Communications", *IEEE Trans. Ant. and Prop.*, vol. 41, no. 10, Oct. 1993
- [Xia, 97] H. H. Xia, "A simplified analytical model for predicting path loss in urban and suburban environments", *IEEE Transactions on Vehicular Technology*, vol. 46, no. 4, November 1997, pp. 1040 – 1046
- [Zhou, 95] Q. Zhou, A. K. Y. Lai "Characterization of Indoor Radio Propagation Environments by Cluster Analysis as Preprocessing for Neural Network Model" 1995, *IEEE*

1. I. Popescu, A. Kanatas, E. Angelou, I. Naforniță, Philip Constantinou, “Applications of Generalized RBF-NN for Path Loss Prediction”, Proc. PIMRC 2002 - The 13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, Lisbon, Portugal, September 15-18, 2002, pp. 484 – 488.
2. I. Popescu, I. Naforniță, A. Kanatas, P. Constantinou, “Applications of General Regression Neural Networks for Path Loss Prediction”, International Workshop “Trends and Recent Achievements in Information Technology”, 16 – 18 May 2002, Cluj – Napoca, Romania
3. I. Popescu, I. Naforniță, Gh. Gavriloaia, P. Constantinou, C. Gordan “Field Strength Prediction in Indoor Environment with a Neural Model”, published in FACTA UNIVERSITATIS, Electronics and Energetics series, vol. 14, No. 3, December 2001, pp. 329-336
4. I. Popescu, I. Naforniță, Gh. Gavriloaia, P. Constantinou, C. Gordan “Field Strength Prediction in Indoor Environment with a Neural Model”, 5th International Conference on Telecommunications in Modern Satellite, Cable and Broadcasting Services – TELSIS 2001, Yugoslavia, Niš, 19–21 September 2001 and published in FACTA UNIVERSITATIS, Electronics and Energetics series, vol. 14, No. 3, December 2001, pp. 329-336
5. Gh. Gavriloaia, I. Popescu, “A New Approach to Mixed Radar Signal Separation”, 5th International Conference on Telecommunications in Modern Satellite, Cable and Broadcasting Services – TELSIS 2001, Yugoslavia, Niš, 19–21 September 2001
6. I. Popescu, I. Naforniță, A. Kanatas, P. Constantinou, “Hybrid model based on Neural Network for Path Loss Prediction”, Proceedings of the IASTED International Conference ADVANCES IN COMMUNICATIONS, July 3 – 6, 2001, Rhodes, Greece, pp. 106 – 109
7. I. Popescu, I. Naforniță, P. Constantinou, A. Kanatas, N. Moraitis, “Neural Model for Path Loss Prediction in Suburban Environment”, Proceedings of the IASTED International Conference ADVANCES IN COMMUNICATIONS, July 3 – 6, 2001, Rhodes, Greece, pp. 110 – 113
8. I. Popescu, I. Naforniță, P. Constantinou, A. Kanatas, N. Moraitis, “Neural Networks Applications for the Prediction of Propagation Path Loss in Urban Environment”, The

IEEE Semiannual Vehicular Technology Conference, VTC 2001 Spring, Rhodes, Greece, May 6-9 2001

9. I. Popescu, I. Nafoarniță, P. Constantinou, A. Kanatas, N. Moraitis, "Prediction of Outdoor Propagation Path Loss with Neural Networks", Proceedings of the 8th Telecommunications Forum Telfor'2000, 21 - 23 November 2000, Belgrade, Serbia, Yugoslavia, pp. 140 – 143
10. A. Voulgarelis, I. Popescu, I. Nafoarniță, P. Constantinou, "Implementation of a Direction-of-Arrival Finding Antenna using Neural Networks", 2nd Symposium on Applied Electromagnetism of Trans Black Sea Region, Xanthi, Greece, 27-29 June 2000
11. I. Popescu, M. Nafoarniță, A. Voulgarelis, P. Constantinou, "Applications of Neural Networks to the Prediction of Propagation Path Loss for Mobile Communication Systems", 2nd Symposium on Applied Electromagnetism of Trans Black Sea Region, Xanthi, Greece, 27-29 June 2000
12. I. Popescu, M. Nafoarniță, A. Voulgarelis, P. Constantinou, "Neural Networks Implementation for Path Loss Prediction Models", May 2000, Oradea, Romania
13. A. Voulgarelis, I. Popescu, I. Nafoarniță, P. Constantinou, "A Neural Networks Application to the Direction – Finding Systems, May 2002, Oradea, Romania

CURRICULUM VITAE

First name: Ileana
Last name: Popescu
Nationality: Romanian
Date and place of birth: 12.01.1969, Oradea, Romania

Education/degrees:

- 1997-Present PhD student at the Faculty of Electronics and Telecommunications of Timisoara, under the supervision of Mr. prof. dr. eng. Nafornită Ioan. The subject of the thesis is: "Neural Networks Applications for Radiocoverage Studies in Mobile Communication Systems"
- 1994 "Novell Education Certificate" from a Standard Course for Novell NetWare v 3.1x Administration, organized by Novell Inc. Utah, USA.
- 1993 Bachelor in Engineering, at the Technical University of Timioara, Faculty of Electronics and Telecommunications, specialized in "Telecommunications".
- 1987 Graduate of High school of Mathematics and Physics "Emanuil Gojdu" Oradea, Romania

Scholarships

- 1999-2000 Socrates scholarship at the Mobile Radiocommunications Laboratory, National Technical University of Athens, Greece

Working experience:

- 1999-Present Visitor Researcher at Mobile Radiocommunications Laboratory, National Technical University of Athens, Greece
- Responsibilities:
- Scientific authority on radiation Hazard Measurements
 - Responsible for the project on field strength measurements from MW antennas in Kerkyra and Rhodes, Greece
 - Participating in the project on measurements of the radiation environment due to indoor monitors in the Greek police headquarters

- Participation on the evaluation of interference from approaching Radars at the International Airport of Athens
- Participation on statistical analysis of the measurements on evaluating the power density from base station antennas at 900 and 1800 MHz
- Participation in COST 281

1994-1999

Teaching assistant at the Departement of Applied Electronics, Faculty of Electrotechnics and Informatics, University of Oradea, Romania

1993-1994

Electronics engineer at S.C. "Mobila Alfa" S.A. Oradea, Romania

Languages:

- Romanian
- English
- Greek – working ability

Area of interest

- Neural Networks applications in Telecommunications such as:
 - o Automatic Modulation Recognition
 - o Direction of Arrival Estimation
 - o Call Admission Control
- Neural Networks applications for the prediction of delay spread and of several parameters related to fast fading (time delay, Doppler effect,...)
- Non-Ionizing Radiation Hazard

Eng.

Ileana Popescu