Modelling of radio transmissions into and within multistorey buildings at 900, 1800 and 2300 MHz

A.M.D. Turkmani
A.F. de Toledo

Indexing terms: Indoor radiowave propagation, Regression analysis

Abstract: A multiple regression analysis method is used to predict the signal path loss encountered by radio transmissions into and within buildings at 900, 1800 and 2300 MHz. It has been found that in addition to the initial free space path loss, the floor mean signal level for radio transmissions *into* buildings is also dependent on the logarithm of the floor area and on the condition of transmission, which is represented in this study by the number of building sides that have a direct line of sight to the transmitter. The path loss for radio transmissions *within* buildings has been found, however, to be dependent on the logarithm of the distance, the floor area, the number of floors between the transmitter and the receiver, and on two other factors identified in the study as sight and ground. The first factor accounts for signals leaving and returning to the building, complemented by some considerations on the ability of signals to propagate on the floor where the transmitter was located. The ground factor accounts for the observed tendency for the signal strength to be higher on the first two floors of a building. Statistical methods have also been used to determine the reliability of the prediction.

1 Introduction

The move towards personal communications has led to the realisation that not enough is known about radio propagation either into or within buildings. In this context, *into* is used to identify the propagation scenario where a base station, often located on a hilltop site or rooftop of a high building, communicates with a radio receiver that is inside another building, and *within* is used to identify the case when both transmitter and receiver are inside the same building. Propagation models that adequately describe the signal in open and urban areas are no longer adequate, since there will be a building penetration loss associated with the indoor environment [1–3]. This additional loss will depend on a large number of factors having various degree of importance. Among them are the transmission frequency, the distance between the transmitter and the receiver, the building construction material and the nature of the surrounding buildings.

Several researchers have studied the problem of receiving radio signals into buildings and model it as the distance–dependency of the path loss when the mobile is outside a building, plus a building loss factor, which is included in the model to account for the increase in attenuation of the received signal observed when the mobile is moved from outside a building to inside. This model was first proposed by Rice [4] and has been used in most subsequent investigations. However, it is necessary to remember that outdoor propagation models such as those of Okumura [5], Hata [6] and Ibrahim [7] has been developed for large cells, whereas for personal communications the suitable cell diameter could be even less than 500 m. Therefore those models cannot be trusted when used for the indoor environment without further investigation. In addition, predicting first the signal outside the building of interest and then, from the result, determining the signals inside the building yields an inevitable reduction in accuracy.

It was therefore concluded that prediction of the path loss for radio transmissions into buildings should be undertaken directly and not merely as an extension of outdoor propagation models. A similar approach was adopted by Barry and Williamson [8] to analyse measurements undertaken in New Zealand at 851 MHz. The propagation into buildings was modelled considering three variables which are often found to be influential in modelling outdoor path loss (namely, base station and mobile antenna heights and path length) and four additional variables which are related to the building layout. These variables were: floor area, number of rooms on a floor, angle of illumination (Θ) of the building relative to the base station direction, and the presence or otherwise of mirror-glass windows. The model has been tested and the results are presented in Section 4.1.

One of the earliest approaches to statistical modelling of propagation totally within buildings was reported by Alexander [9], who stated that the path loss within buildings at 900 MHz can be predicted using the simple distance–power law.

$$path\ loss = 10n \log_{10} d \qquad (1)$$

Alexander's experiments were undertaken in a building with internal steel partitioning, and yielded a path-loss attenuation factor $n = 5.7$. Motley and Keenan [10] have also undertaken a series of indoor measurements at 900 and 1700 MHz in a building of standard steel frame construction with brick external walls and plasterboard internal partitions. Motley and Keenan's modelling

Paper 9684I (E8), first received 3rd February and in revised form 7th June 1993

The authors are with the Department of Electrical Engineering and Electronics, The University of Liverpool, P.O. Box 147, Liverpool L69 3BX, United Kingdom

results have shown that a better fit to the experimental data can be achieved by introducing, to the Alexander model, a correction factor (F_{floor}) representing the signal attenuation per floor. The resultant model was then expressed as

$$path\ loss = S + k_{floor} F_{floor} + 10n \log_{10} d \qquad (2)$$

where S is the path loss at 1 m and k_{floor} is the number of floors traversed. The values of F_{floor} were found to be 10 and 16 dB at 900 and 1700 MHz, respectively. The values of n were found to be of the same order at both frequencies (4 at 900 MHz and 3.5 at 1700 MHz) and S was found to be 16 dB at 900 MHz and 5 dB at 1700 MHz.

The primary emphasis of this study is to examine the aforementioned (into and within buildings) models and to determine their relative accuracy in modelling the indoor environments. The experimental data of a previously reported series of field trials [2, 3] are used in this paper. The validity of the models is assessed by evaluating the root mean square error (RMSE) between the predicted and measured path loss values. In addition to the evaluation of the previous models, the emphasis of this study is also to investigate and develop ways of improving these models, such as their RMSE values are reduced to acceptable levels. This is undertaken by introducing as many additional factors as necessary.

2 Background

Measurements of the received signal strength were undertaken within buildings in the University of Liverpool precinct. The buildings were Blocks A and B of the Department of Electrical Engineering & Electronics and the Departments of Computer Science and Life Sciences. For the into building measurements the transmitter was located on the roof of the Mechanical Engineering building, at the height of 40 m and at distances of 170, 220, 300 and 320 m from the other buildings. No line of sight existed between Electrical Engineering Block B and the transmitter. Partial line of sight existed, however, to the other buildings. A general description of the buildings is given in Reference 2. The cumulative distribution of the small- and large-scale signal variations were discussed in References 11 and 2. The penetration loss was found to decrease with height by 1.38, 1.36 and 1.5 dB per floor for signals at 900, 1800, 2300 MHz, respectively. The average penetration loss at ground level was found to be around 14.2, 13.4 and 12.8 dB, respectively, at 900, 1800 and 2300 MHz. In general, the penetration loss decreased with an increase in the transmission frequency [3]. Perhaps it is important to point out that experiments were conducted in every single room and corridor of the four buildings. Enough data samples were collected in each experiment to determine accurately the mean signal level in every room/corridor. The path loss for a given floor of a certain building is the mean of the path loss values of all rooms and corridors in that floor. In this study only the mean floor path loss values will be used in developing the into building model.

The within buildings measurements were undertaken within the same four buildings. Twenty-two survey measurements at 1800 and 2300 MHz, and six survey measurements at 900 MHz were conducted within the four buildings. In every survey, experiments were conducted in every room and corridor of the building for a certain base station location. To determine the transmitter location which gives the best signal coverage, six different rooms in the Electrical Engineering Block B, and three in the Computer Science building were selected to house the base station. The significant conclusions relating to the small and large-scale signal variations, the rate of change of mean signal level per floor for signals travelling within building, the best signal coverage according to the transmitter location, and the path loss attenuation factors are fully described in References 2 and 3.

The prediction of indoor signal levels requires the undertaking of an extensive series of field trials in all buildings of interest. However, this is not possible because of the very large number of required experiments and the inevitable expense. It is therefore necessary to perform the prediction using information based on what may seem to be incomplete data. Such information may therefore have an element of uncertainty and a risk of being incorrect. Using appropriate statistical techniques it is possible to generalise from a given set of data to a more broadly applicable statement and for that purpose specific and rigorous techniques should be applied to estimate the degree of uncertainty [12, 13]. In this study, the statistical technique of regression analysis is applied. A detailed discussion of the usefulness of regression analysis was reported previously in Reference 11. For convenience and completeness, a summary of the main equations of the regression analysis is given in the Appendix. Refer to Reference 11 for further details.

In summary, the multiple regression analysis requires a rigorous and careful estimation of its quality, which can be tested using the F-test and the coefficient of determination r^2. Partial quality test of the individual variables should also be made using the t-test. In addition to these three essential quality tests, a close examination of the correlation between any two variables is also necessary. Even if all the statistical tests have shown that the assumed model was useful in predicting the dependent variables, it cannot be concluded that such a model is the best predictor model without further data and analysis. It is also necessary to consider that the calculated prediction equation, based on a set of independent variables, is appropriate only over the range of the sample values used in the analysis. Extrapolation beyond this range may lead to errors.

3 Modelling

3.1 General discussion

Propagation into and within buildings involves a more complex multipath structure than that of the outdoor land-mobile radio channel which is dependent on path length, effective base station antenna height, frequency, mobile antenna height, environment local to the mobile. In addition to these variables, indoor propagation is also affected by other empirically observed variables such as building structure and layout of rooms. Hence an investigation of the relationships between all these variables is of interest.

One type of analysis primarily concerned with the association between variables is regression analysis. Simple regression analysis is used when the measured values are related to one variable $[Y = f(X_1)]$, whereas for multiple variable $[Y = f(X_1, X_2, \ldots)]$, multiple regression analysis can be applied. The most challenging task in regression analysis is, however, choosing which of the possible variables to include in the final regression model. In what follows, the possible variables which can affect propagation into and within buildings will be discussed and decisions whether or not to include some of these variables will be made.

3.2 *Factors affecting propagation into buildings*

The area of the external windows is expected to be an influential variable which could affect the path loss. However, because of the four buildings tested in the University precinct all had windows areas restricted to a fairly narrow range *viz.* 35 to 40% of the external wall areas, for most of the floors, it was decided to discard this variable, only because of the narrow range of 5% (40–35).

It has been found previously [2] that propagation into buildings is highly affected by the condition of transmission, i.e. the presence or absence of a line-of-sight path. It was therefore decided to include this as a variable. The line-of-sight variable was subjected to two different treatments. The first approach was to consider it as a qualitative factor S_I of value 0 or 1, depending on whether or not a line of sight exists between the transmitter and the receiver. The second approach was to consider the number of building sides seen by the transmitter on each floor of the building housing the receiver. If the transmitter could see only one side, then the line-of-sight quantity S_Q was valued at 0.25 (i.e. 1 divided by 4). Obviously, when no line of sight existed or two sides of the building were seen by the transmitter, S_Q was given the value 0.0 or 0.5.

The path length d (i.e., the distance between the antennas), area of the floor A_f, average room area on each floor a_R, receiver height h_{RX}, number of rooms N_R on each floor, the relation between the number of rooms which contain external windows and the number of rooms with no windows R_{win}, the relation between the length and the width of the floor R_l, and the angle of illumination Θ were some of the other independent variables considered. Some of these variables, however, were considered in both linear and logarithm terms. Overall, 15 variables were considered for the multiple regression analysis. However, the number of predictor variables to be used simultaneously should not be greater than one quarter of the sample size [12]. Therefore, recalling that in the into building measurements the sample size was equal to 30 (seven floors in the Electrical Engineering Block A, four floors in the Electrical Block B, nine floors in the Computer Science and ten floors in the Life Sciences), the regression model should not have more than seven independent variables at any one time.

3.3 *Factors affecting propagation within buildings*

As in the case for propagation into buildings, the path length d, area of the floor A_f, room area a_R, number of rooms N_R on each floor, and the receiver height h_{RX}, were also considered as independent variables for the modelling of the within buildings measurements. Fig. 1 shows an example of the isometric projections of the mean signal levels on the floors of a building. In this particular example, the building was that of the Life Sciences with the transmitter located on the fourth floor (room 411). It is evident that the transmitted signal spreads more or less uniformly in all direction. This means that floors just above and below the transmitter will receive stronger signals than, for instance, a region far away from the transmitter on the same floor. It is evident also that the signal levels on all floors are, in general, higher near the side of the building where the transmitter was located. This observation indicates that part of the signal propagates from inside the building to the outside and then returns back to the other floors of the same building.

The first approach to account for this propagation situation was to introduce a new qualitative variable S_{sig_1} which is called sight in this study. S_{sig_1} can take the value 1 or 0, depending on whether or not the surveyed rooms were located at the same side of the building where the transmitter was located. The second approach was to consider the variable sight (represented by S_{sig_2}) as a quantitative factor: for rooms located at the same side of the building where the transmitter was located

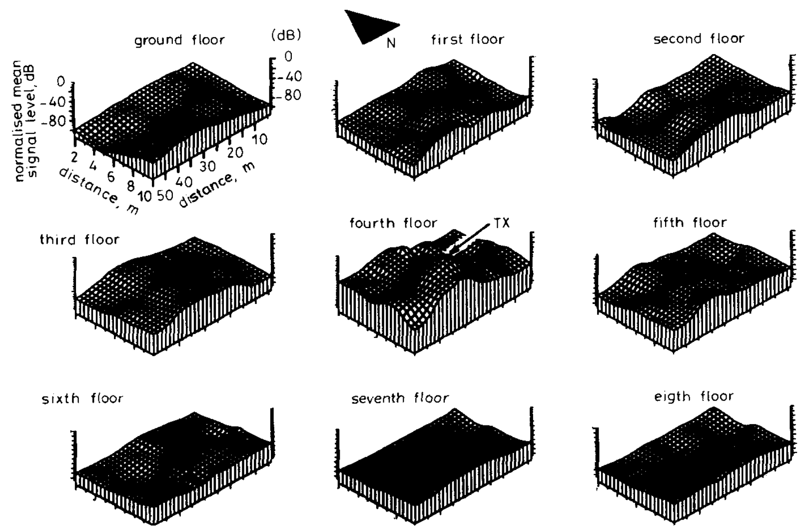

Fig. 1 *Isometric projection views of normalised mean signal levels*
Transmitter within Life Sciences, room 411, 2300 MHz

$S_{sig\ 2}$ was given the value 1. For rooms located in the lateral sides (i.e. the sides which are perpendicular to the side where the transmitter was located) $S_{sig\ 2}$ was given the value 0.5. The $S_{sig\ 2}$ factor was made equal to 0.25 for the rooms located on the opposite side of the building to where the transmitter was located. Finally, for internal rooms where no windows existed towards the outside, $S_{sig\ 2}$ was considered equal to 0. In fact, this treatment is similar to the line-of-sight factor discussed in Section 3.2.

Because of the signal levels were consistently higher on the floors where the transmitter was located, a third approach was investigated for the variable sight. In addition to the specification used for S_{sig_2}, in areas, rooms and corridors which are close to the room housing the transmitter, and where the barriers between the transmitter and the receiver include only wooden doors, the variable sight (identified now simply by S_{sig}) was made equal to 1, 0.5 or 0.25 depending on the proximity and the number of corners which have to be turned in the free path joining the transmitter and the receiver: for rooms located in front of the room housing the transmitter, S_{sig} was considered equal 1. If there was one corner (or two corners) in the corridor joining the transmitting and receiving rooms, S_{sig} was made equal to 0.5 (or 0.25). S_{sig} was considered equal to 0 for any other condition.

Life Sciences was described in Reference 2 as an 11-floor reinforced concrete building with an annex structure (in its south direction) from the ground floor up to the second floor. Therefore, a new variable B_{annex} was created mainly for this building to deal with this specific situation, and it was treated as a qualitative factor. $B_{annex\ 1}$ was made equal to 1 for all rooms located below the third floor. Obviously for rooms located above the second floor $B_{annex\ 1}$ was equal 0. In the second approach, only the rooms located in the half partition of the building, in a southerly direction, where the annex structure is present, had the factor $B_{annex\ 2}$ valued at 1.

Some buildings include main corridors which provide access to most of the rooms on both sides of the building. A new qualitative indicator variable identified by C_{main} can be used in this building and made equal to 1 for the main corridors or 0 for every other area, room or secondary corridor.

The analyses of past measurements, such as those reported in Reference 2, have shown that the signal levels in the first two floors of a building tend to be relatively higher than expected. That was due to signals propagating from the inside of the building at higher floors and then returning back to the building at the lower floors. A new qualitative variable, identified here as ground factor G_G and which defines whether the rooms are located on the first two floor ($G_G = 1$), or are positioned on any other floor ($G_G = 0$), was introduced to account for this situation.

4 Model implementation

There are three multiple regression methods, namely stepwise, forward and backward [13]. Although the three methods are equivalent, forward procedure is, perhaps, more common. However, for convenience and because there was an upper limit to the number of variables that could be handled simultaneously, the multiple regression analysis in this paper is performed using first the backward procedure, then the remaining possible independent variables are tested using the forward procedure. Both the overall and partial quality tests are applied to the regression analysis of the experimental measurements at 900, 1800 and 2300 MHz. The level of significance α for the F and t-tests was set to 1%.

4.1 Analysis of into-building meaurements

In total, over 60 models have been considered: some of them failed completely, while others show a reasonable degree of acceptance. Table 1 shows only 13 examples of the analysis conducted with the data at 2300 MHz. The first two models of Table 1 compare a number of possible combinations using six independent variables. These included two of the four variables generally found to be influential in modelling path loss (mobile antenna height h_{RX} and path length d), three variables which were used in the Barry and Williamson model [8] (floor area A_f, angle of arrival of the signal Θ, and number of rooms in the tested floor N_R), and finally a sixth independent variable relating the number of rooms without and with windows R_{win}. There is a general improvement in the overall quality of the multiple regression equation, i.e. with a few exceptions, the coefficient of determination r^2 and the F-value have increased.

After the second model, the variable R_{win} was replaced by the variable R_1 yielding slightly better results. The results obtained for the fourth model, considering the level of significance $\alpha = 1\%$ (e.g. the tabulated values of $F_{6,\,23;\,1\%}$ and $t_{23;\,1\%}$ are 3.7102 and 2.807, respectively), are all significant (i.e. $F > 3.7102$) and it was also possible to conclude that the model for the floor means signal level was dependent on each of the six independent variables tested (i.e. all $|t_j| > 2.807$). However, strong correlation between the variables Θ and R_1 and the variables $\log A_f$ and R_1 (e.g. 81 and 54%, respectively) were found. Dropping the variable R_1 for the fifth model, the results show that the overall quality of the multiple regression has worsened and the t-values of three independent variables have fallen below the acceptable $t_{23;\,1\%} = 2.807$ and consequently could not be considered any longer to influence the regression model significantly.

In the sixth model, when the variable room area a_R was replaced by $\log_{10} a_R$ there was a slight improvement in the general quality of the regression. However, the

Table 1: Multiple regression models for into-buildings measurements at 2300 MHz (t_j between brackets)

Model	Chosen set of independent variables											Analysis	
	$\log A_f$	a_R	$\log a_R$	θ	d	$\log d$	h_{RX}	N_R	R_{win}	R_1	S_Q	r^2	F
1	26.2 (5.6)	—	—	−0.04 (−1.9)	—	9.0 (1.8)	−0.24 (−5.0)	−0.23 (−2.4)	4.1 (0.9)	—	—	0.890	30.9
2	26.6 (5.7)	—	—	−0.05 (2.0)	16.9 (1.8)	—	−0.24 (−5.1)	−0.23 (−2.5)	4.2 (1.0)	—	—	0.891	31.3
3	34.3 (6.9)	—	—	−0.1 (−2.0)	29.5 (2.4)	—	−0.24 (−5.4)	−0.25 (−2.8)	—	−1.54 (−1.3)	—	0.894	32.5
4	39.0 (6.9)	0.03 (3.0)	—	−0.21 (−3.8)	57.8 (3.7)	—	−0.24 (−5.5)	—	—	−4.4 (−3.1)	—	0.898	33.9
5	22.7 (8.5)	0.01 (1.2)	—	−0.04 (−2.0)	16.9 (1.7)	—	−0.28 (−5.6)	—	—	—	—	0.855	28.2
6	33.0 (7.1)	—	5.8 (3.1)	−0.16 (−3.3)	45.2 (3.4)	—	−0.23 (−5.4)	—	—	−3.1 (−2.6)	—	0.900	34.4
7	18.7 (7.1)	—	0.2 (0.1)	—	74.3 (5.5)	—	−0.07 (−1.4)	—	—	−0.45 (−1.0)	−24.3 (−5.3)	0.934	54.2
8	20.9 (7.5)	—	0.8 (0.6)	—	—	42.4 (5.5)	−0.07 (−1.5)	—	—	−0.64 (−1.4)	−23.3 (−5.2)	0.934	54.0
9	20.4 (7.8)	—	—	—	—	42.7 (5.7)	−0.07 (−1.4)	—	—	−0.6 (−1.4)	−24.3 (−6.1)	0.933	66.8
10	17.9 (9.2)	—	—	—	—	34.0 (6.9)	−0.09 (−1.9)	—	—	—	−21.6 (−6.0)	0.926	78.1
11	16.1 (9.7)	—	—	—	—	37.8 (7.9)	—	—	—	—	−26.4 (−11.0)	0.919	97.7
12	17.6 (9.2)	—	—	—	—	—	−0.07 (−1.5)	—	—	—	−24.5 (−10.2)	0.964	231.0
13	16.1 (9.8)	—	—	—	—	—	—	—	—	—	−27.3 (−18.0)	0.961	331.1

variable R_1 could not be considered important for the model. Replacing the variable angle Θ by line of sight (not shown in the Table) in its qualitative form (i.e. S_I), the important of R_1 worsened still further ($t_j = 0.62$) and the variable $\log_{10} a_R$ failed to hold its significance ($t_j = 1.7$). In the seventh model, when the variable line of sight was modified from qualitative to quantitative (i.e. S_I was replaced by S_Q), the only independent variables to hold their significance were $\log_{10} A_f$ and d which added to the good level of significance of the new introduced variable S_Q. There was also an improvement in the general quality of the regression model.

Models from the eighth to eleventh show improved results when the considered poor quality variables were gradually dropped and the path length d was replaced by $\log_{10} d$. Attempts have also been made using distance between buildings instead of distance between antennas, and/or height of the building as two of the independent variables considered without, however, achieving complete success. Model ten is one of these examples, where the variable h_{RX} represents the difference between transmitter and receiver heights. Although the overall quality of the model was acceptable, the variable h_{RX} yielded an unacceptable $t_j = 1.9$.

The interesting fact about the eleventh model was the coefficient 37.8, calculated for the independent variable $\log_{10} d$, significantly close to 40 dB per decade, which is the predicted coefficient in the theoretical model for path loss, when two waves are combined: a direct wave and a reflected wave [14]. Coefficients close to 40 were always observed after replacing d by $\log_{10} d$ (e.g. see models 8 to 11). For the following models, the coefficient of the independent variable $\log_{10} d$ was assumed equal to 40 and the measured path loss was adjusted accordingly. For example, the measured path loss on the ground floor of the Life Sciences building was found equal to 136.5 dB, which is the difference between the transmitted and received signals. Considering that part of this loss could be attributed to the factor $40 \log_{10} d = 40 \log_{10}(322.5) = 100.3$ dB, the new measured path loss, i.e. the amount of variation left to be explained in the regression analysis, was adjusted to 36.2 dB.

Models 12 and 13 show the results when the effect of the variable h_{RX} which was found significant in the first models, was again assessed. The best of all results (13th model) was obtained when only two variables were present in the regression equation and, the resulting model for the path loss, at 2300 MHz, was found to be

$$Y_{2300} = -7.9 + 40.0 \log_{10} d + 16.1 \log_{10} A_f - 27.3 S_Q \quad (3)$$

where the RMSE determined was 1.7 dB. Although the coefficient of determination r^2 was found to be slightly better for the 12th model when comparing it to the 13th model (e.g. 96.4 against 96.1%, respectively), the independent variable h_{RX} of the 12th model yielded an unacceptable t-coefficient (e.g. $t = 1.5$), and it was dropped for the 13th model. In contrast to the coefficient of determination, the F-value improved substantially when h_{RX} was dropped (e.g. the F-value increased from 231.0 to 331.1).

A similar analysis carried out for the measurements at 1800 and 900 MHz yielded, respectively, the following equations:

$$Y_{1800} = -27.9 + 40.0 \log_{10} d + 23.3 \log_{10} A_f - 20.9 S_Q \quad (4)$$

and

$$Y_{900} = -37.7 + 40.0 \log_{10} d + 17.6 \log_{10} A_f - 27.5 S_Q \quad (5)$$

with RMSEs equal to 2.2 and 2.4 dB, respectively. The appropriateness of the models was verified graphically. Fig. 2 shows the results for the 2300 MHz measurements. Fig. 2*a* compares the proposed Barry and Williamson models with the measured values. The path loss between antennas for Electrical Engineering Block A are shown as cases 1 to 7, Electrical Engineering Block B as cases 8 to 11, Computer Science as cases 12 to 20 and Life Sciences as cases 21 to 30. Examination of this graph reveals a very good agreement between the proposed model (i.e. eqn. 3) and the measured values. All generalisations made from the regression relationship are based on the assumption that the residual term ε [11] is a random variable having a mean value equal to 0 and constant variance.

This normality has been verified by plotting the cumulative distribution function (CDF) of the residuals for the proposed model on normal graph paper, and is shown in Fig. 2*b*. The validity of the proposed model was also assessed by determining the building penetration loss as a function of floor level and comparing it with the measured values, as shown in Fig. 2*c*. The good agreement between measured and predicted values is clear from the graph. Both the measured and predicted penetration loss values were referred to the same mean signal levels measured outside the buildings. Figs. 3 and 4 show similar analysis carried out for the 1800 and 900 MHz, with the results revealing strong similarities.

4.2 Analysis of within-building measurements

Both the overall and partial quality tests were applied to the regression analysis of the 28 within building survey measurements. Initially, each survey was examined separately. In a subsequent stage, all the results for one particular frequency were joined together to allow the investigation of global models.

Table 2 shows some of the successful models for the within buildings measurements conducted in the Life Sciences building at 2300 MHz. The first model corresponds to the well known distance/power law proposed by Alexander. The linear regression is defined by regression coefficient (gradient) 5.17 dB/decade. The RMSE was found to be 12.4 dB, and the coefficient of determination r^2 was 0.44; i.e. 44% of the path loss is explained by the line defined by the gradient 51.7 dB/decade. The F-values are not shown in this table since only the successful models are presented. In general, the F-values were always higher than 100.

The Motley and Keenan model, i.e. the second model of Table 2, displays improved RMSE and r^2: the RMSE decrease by 0.3 dB and the coefficient of determination increased by 3%. The coefficients F_{floor} and $10n$ of eqn. 2 were found to be 2.7 dB/floor and 43.5 dB/decade, respectively.

The three possibilities for the variable sight (indicated in Section 3.3) are represented in Table 2 by models 3, 4 and 5. It is clear that there is an improvement (e.g. the RMSE values decrease) towards model 5 which considers S_{sig} as a quantitative factor complemented with the special specification for the floor housing the transmitter. There is an obvious improvement in the overall model when compared with the models by Alexander (or Motley and Keenan). The RMSE decreased from 12.4 (or

Table 2: Multiple regression analysis for within-building measurements in life sciences department (*Tx*: room 411, 2300 MHz, t_j between brackets)

Model	Chosen set of independent variables									Analysis	
	$\log_{10}d$	k_{floor}	S_{sig_1}	S_{sig_2}	S_{sig}	B_{annex}	C_{main}	G_G	cnst	RMSE	r^2
1	51.7 (12.5)								8.8	12.4	0.44
2	43.5 (9.6)	2.7 (3.8)							13.0	12.1	0.47
3	46.0 (10.4)	2.3 (3.4)	−7.0 (−4.0)						15.4	11.6	0.51
4	44.1 (12.4)	2.4 (4.5)		−17.9 (−11.5)					21.2	9.4	0.68
5	41.5 (13.4)	1.8 (3.7)			−21.2 (−15.3)				28.2	8.2	0.76
6	42.0 (13.8)	2.1 (4.4)			−22.3 (−15.9)	−4.1 (−3.0)			28.5	8.1	0.77
7	41.3 (14.4)	2.2 (4.9)			−21.4 (−16.7)	−9.5 (−6.1)			28.9	7.6	0.79
8	36.1 (13.1)	2.5 (6.0)			−23.1 (−19.2)	−9.0 (−6.3)	−16.1 (−6.3)		36.3	6.9	0.83
9	36.4 (13.5)	2.9 (6.7)			−23.7 (−19.7)	−7.3 (−4.7)	−16.3 (−6.5)	−4.2 (−2.8)	35.9	6.8	0.84

12.1) to 8.2 dB, and the coefficient of determination r^2 increased from 44 (or 47) to 76%. Model 5 implies the 76% of the path loss in the Life Sciences building is explained by the variables $\log_{10} d$, F_{floor} and S_{sig} with coefficients 41.5 dB/decade, 1.8 dB/floor and −21.2 dB, respectively.

Table 2 shows a slight improvement in model 6 ($B_{annex} = B_{annex\_1}$), and a better improvement in model 7 ($B_{annex} = B_{annex\_2}$). Additional slight improvements were obtained in the values of RMSE and r^2 in models 8 and 9, where the other two variables C_{main} and G_G are used. In these two models $B_{annex} = B_{annex\_2}$. Compared with the Alexander model, the last model decreased the RMSE by 45% (from 12.4 to 6.8 dB), and improved r^2 by 91% (from 0.44 to 0.84). It is also interesting that the improvement of the Motley and Keenan model in relation to the Alexander model was only 2% for the RMSE and 7% for r^2. The resulting model for the path loss for the Life Sciences building with the transmitter in room R411 (located in the middle of the building), is

$$
\begin{aligned}
Y_{room,\,2300,\,LSc(R411)} &= 35.9 + 36.4 \log_{10} d + 2.9 k_{floor} \\
&\quad - 23.7 S_{sig} - 16.3 C_{main} - 7.3 B_{annex} - 4.2 G_G \qquad (6)
\end{aligned}
$$

Introducing more appropriate independent variables could bring some more improvement to the model; however, the amount of refinement gained after every new implementation would tend to be less significant than the previous one.

Another important consideration regarding the modelling of radio transmission within buildings is the degree of universality exhibited by each considered variable. For example, the number of floors is a very general concept which can be applied to virtually every multistorey building. However, the annex structure is not a very common feature, and is applied only for a limited number of cases. It was therefore decided to perform the analyses for the other buildings using universal features which can be applied to most multistorey building cases. Models obtained in locations which yielded the best signal coverage inside the other three buildings (Section 5.2 of Reference 2), at 2300 MHz are given below

$$
\begin{aligned}
Y_{room,\,2300,\,EEA(R302)} &= -4.5 + 35.2 \log_{10} d + 10.8 k_{floor} \\
&\quad - 9.0 S_{sig} - 7.2 G_G \qquad (7)
\end{aligned}
$$

$$
\begin{aligned}
Y_{room,\,2300,\,EEB(R101)} &= 5.5 + 24.8 \log_{10} d + 12.8 k_{floor} \\
&\quad - 5.8 S_{sig} - 6.3 G_G \qquad (8)
\end{aligned}
$$

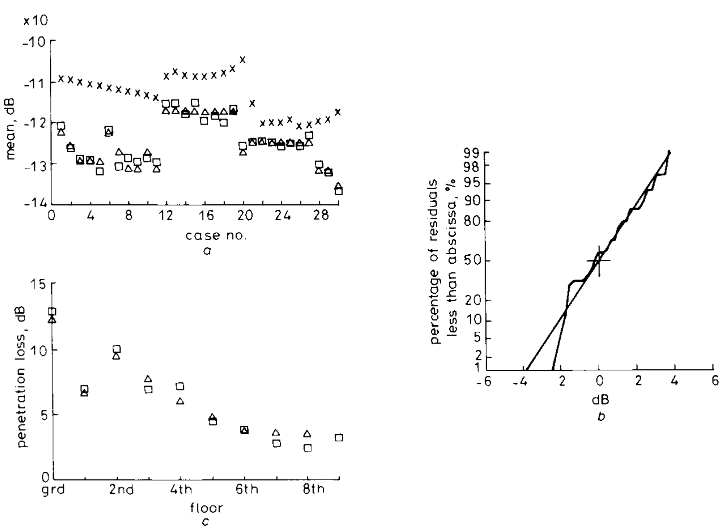

Fig. 2 *Into-buildings measurements at 2300 MHz*
a measured and predicted signal strength
b CDF of residuals (dB)
c penetration loss results
△ proposed model
× Williamson model
□ measured values

$$Y_{room,\,2300,\,CSc(R413)} = 11.0 + 36.9\log_{10}d + 9.0k_{floor} - 9.6S_{sig} - 6.0G_G \qquad (9)$$

with RMSE equal to 6.4, 5.9 and 8.6 dB, respectively.

It is evident that good predictions can be made for radio transmissions within buildings where the features and transmission condition are well known because, in that case, particular models such as those analysed previously can be applied yielding smaller errors. In the absence of this specific information, there will be a need for global models that are more dependent on global/universal variables. Three such models were obtained by collating all the data of all the survey measurements at the three different frequencies. The models were found to be

$$Y_{room,\,2300} = 21.6 + 39.1\log_{10}d + 3.8k_{floor} - 17.8S_{sig} - 0.01A_f - 8.8G_G \qquad (10)$$

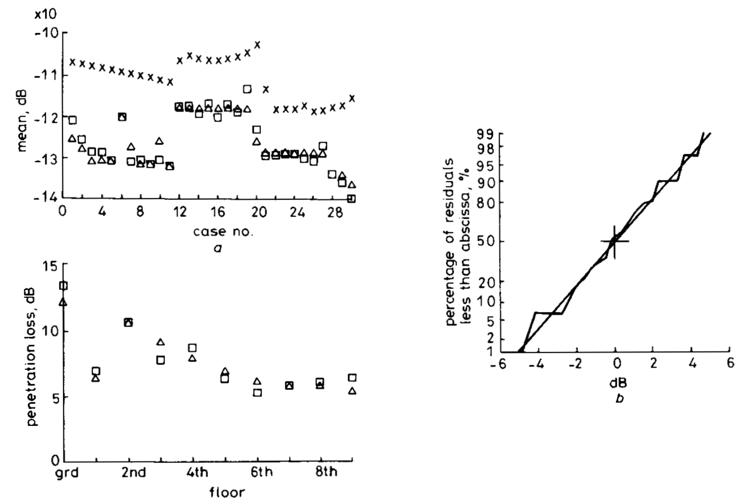

Fig. 3 *Into-buildings measurements at 1800 MHz*
a measured and predicted signal strength
b CDF of residuals (dB)
c penetration loss results
△ proposed model
× Williamson model
□ measured values

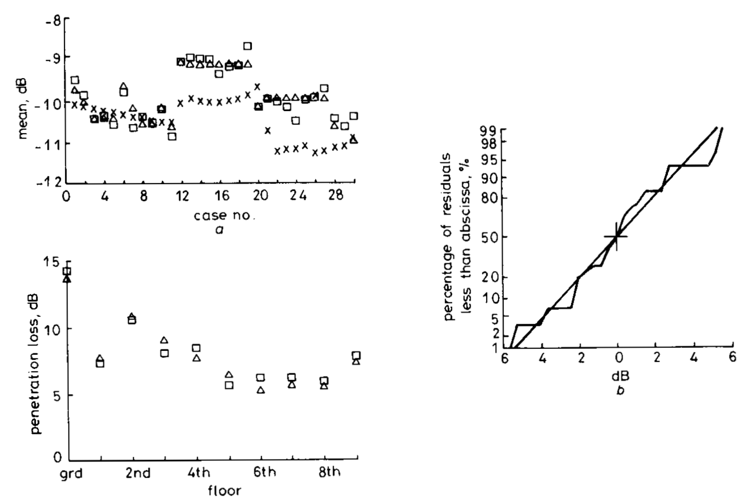

Fig. 4 *Into-buildings measurements at 900 MHz*
a measured and predicted signal strength
b CDF of residuals (dB)
c penetration loss results
△ proposed model
× Williamson model
□ measured values

$$Y_{room,\,1800} = 24.5 + 33.8\log_{10} d + 4.0k_{floor} - 16.6S_{sig} - 0.017A_f - 9.8G_G \tag{11}$$

$$Y_{room,\,900} = 18.8 + 39.0\log_{10} d + 5.6k_{floor} - 13.0S_{sig} - 0.024A_f - 11.0G_G \tag{12}$$

where the RMSE determined were 12.4, 10.9 and 11.6 dB for 2300, 1800 and 900 MHz, respectively.

5 Conclusion

It has been shown that the path loss of radio transmissions into buildings was found to be linearly dependent on the logarithm of the floor area, on a relation representing the number of building sides seen by the transmitter and on the free-space path-loss equation. The path loss for radio transmissions within buildings has been found to be linearly dependent on the logarithm of the distance, on the floor area, on the number of floors between transmitter and receiver, and on two factors identified in this study as sight and ground. It has been shown that different conditions of signal transmission and reception are more important in determining the values of each model coefficient than the frequency variation. Nevertheless, better predictions are possible in buildings where the features and transmission conditions are well known because, in that case, particular models can be applied yielding, in consequence, smaller errors.

6 References

1 COX, D.C., MURRAY, R.R., and NORRIS, A.W.: 'Measurements of 800 MHz radio transmission into buildings with metallic walls', *Bell System Tech. J.*, 1983, **62**, (9), pp. 2695–2717

2 TURKMANI, A.M.D., and TOLEDO, A.F.: 'Radio transmission at 1800 MHz into and within buildings', *IEE Proc., I*, 1991, **138**, (6), pp. 577–584

3 TOLEDO, A.F., and TURKMANI, A.M.D.: 'Propagation into and within buildings at 900, 1800 and 2300 MHz'. 42nd IEEE Conference on *Vehicular technology*, Denver, CO, USA, 1992, pp. 633–636

4 RICE, L.P.: 'Radio transmission into buildings at 35 and 135 MHz', *Bell System Tech. J.*, 1959, **38**, (1), pp. 197–210

5 OKUMURA, Y., OHMORI, E., KAWANO, T., and FUKUDA, K.: 'Field strength and its variability in VHF and UHF land mobile radio service', *Rev. Elec. Commun. Lab.*, 1968, **16**, pp. 825–873

6 HATA, M.: 'Empirical formula for propagation loss in land mobile radio services', *IEEE Trans.*, 1980, **VT-29**, pp. 317–325

7 IBRAHIM, M.F., and PARSONS, J.D.: 'Signal strength prediction in built-up areas, part 1: median signal strength', *IEE Proc. F*, 1983, **130**, (5), pp. 377–384

8 BARRY, P.J., and WILLIAMSON, A.G.: 'Statistical model for UHF radio-wave signals within externally illuminated multistorey buildings', *IEE Proc. I*, 1991, **138**, (4), pp. 307–318

9 ALEXANDER, S.E.: 'Radio propagation within buildings at 900 MHz', *Electron. Lett.*, 1982, **18**, (21), pp. 913–914

10 MOTLEY, A.J., and KEENAN, J.M.P.: 'Personal communication radio coverage in buildings at 900 MHz and 1700 MHz', *Electron. Lett.*, 1988, **24**, (12), pp. 763–764

11 TOLEDO, A.F.: 'Narrowband characterisation of radio transmissions into and within buildings at 900, 1800 and 2300 MHz'. PhD thesis, Department of Electrical Engineering and Electronics, University of Liverpool, 1992

12 KENNEDY, J.B., and NEVILLE, A.M.: 'Basic statistical methods for engineers and scientists' (Harper & Row, New York, USA, 1986, 3rd edn.)

13 LAPIN, L.L.: 'Probability and statistics for modern engineering' (PWS–Kent Publishing Co., Wadsworth, Inc., 1990, 2nd edn.)

14 LEE, W.C.Y: 'Mobile communications design fundamentals' (Howard W. Sams, Indianapolis, IN, USA, 1986)

7 Appendix

Suppose that n measurements of a certain quantity, say Y, have been undertaken and that m variables are thought to affect all the measured values of Y. A multiple regression model that relates an individual value of Y to the m variables can then be expressed by

$$Y_i = b_o + \sum_{j=1}^{m} b_j X_{ij} + \varepsilon_i \tag{13}$$

for $i = 1, 2, \ldots, n$ and X_{ij} is the ith value of the jth variable and ε_i is the error or residual. The parameters $b_1, b_2, \ldots, b_m$ are the regression coefficients corresponding to the m variables; b_o is a constant which can account for all other unconsidered variables. The predicted value of the dependent variable Y_i, can be expressed by

$$\hat{Y}_i = b_o + \sum_{j=1}^{m} b_j X_{ij} \tag{14}$$

An important issue in regression analysis is assessing its overall quality. An analysis of residuals or errors is a useful way of evaluating how good the regression is. One possible statistical test dealing with residuals is the F-test, which for m independent variables (i.e. $b_1, b_2, \ldots, b_m$) and n observations, is represented by

$$F = \frac{\sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2/m}{\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2/v} \tag{15}$$

where $\bar{Y}$ is the mean value of Y. With the number of degrees of freedom m for the numerator and $v = n - m - 1$ for the denominator, the tabulated value of $F_{m,\,v:\,\alpha\%}$ for a given level of significance α is compared to the calculated F. If the computed F is greater than the tabulated value it can then be concluded that the regression results are significant, and the level of significance α, represents the probability that the wrong conclusion has been drawn. The overall quality of the regression analysis can be assessed using the coefficient of determination which measures the degree of association between the dependent variable and all the independent variables taken together and is

$$r^2 = \frac{\sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^{n} (Y_i - \bar{Y})^2} \tag{17}$$

The ratio r^2 must lie between zero and unity. When $r^2 = 1$, all variation has been explained but if it is equal to zero, the regression model does not explain anything, that is Y is not a function of the variables $X_1, X_2, \ldots, X_m$. After determining r^2, the assumed model is useful in describing the dependent variable Y (i.e. the path loss between the antennas), it is necessary to check the significance of the partial regression coefficients b_j, which were considered to have some practical importance. This is done by means of the t-test. The appropriate statistic of the test for the jth independent variables, is

$$t_j = \frac{b_j}{\sqrt{\left[e_{jj} \sum_{i=1}^{n} (Y_i + \hat{Y}_i)^2/v \right]}} \tag{18}$$

where $j = 1, 2, \ldots, m$ and e_{jj} is the jth diagonal element of the matrix $(X^T X)^{-1}$, with the first element e_{00} corresponding to the constant term of the model being ignored. The quality.

$$\sqrt{\left[\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 / v\right]} \tag{19}$$

is often referred to as the root mean square error (RMSE) value. As in the F-test, the computed t_j should be greater than the tabulated $t_{v;\,\alpha\%}$ for a level of significance α.

One of the problems of multiple regression analysis is the possibility of selecting highly correlated independent variables which indicates that the variables tend to change together. The correlation coefficient between any two variables, say X_k and X_l, is

$$r_{X_k X_l} = \frac{\sum_{i=1}^{n} (X_{ik} - \bar{X}_k)(X_{il} - \bar{X}_l)}{\sqrt{\left[\sum_{i=1}^{n} (X_{ik} - \bar{X}_k)^2 \sum_{i=1}^{n} (X_{il} - \bar{X}_l)^2\right]}} \tag{20}$$

where X_{ik} and X_{il} are the ith component of X_k and X_l, respectively. $\bar{X}_k$, $\bar{X}_l$, are the average values of two variables. A high correlation coefficient does not prevent finding an estimated regression equation; however, different sets of sample observations could result in very different estimated regression coefficients, reflecting an inherent instability in the regression relationship. The elimination of one of the highly correlated variables from the regression breaks that pattern and yields different results for the interrelation among the remaining variables.