International Journal of Electrical, Electronic and Communication Sciences
ISSN: 2517-9438
Vol:3, No:2, 2009

Statistical (Radio) Path Loss Modelling: For RF Propagations within localized Indoor and Outdoor Environments of the Academic Building of INTI University College (Laureate International Universities)

Emmanuel O.O. Ojakominor and Tian F. Lai

***Abstract*—**A handful of propagation textbooks that discuss radio frequency (RF) propagation models merely list out the models and perhaps discuss them rather briefly; this may well be frustrating for the potential first time modeller who's got no idea on how these models could have been derived. This paper fundamentally provides an overture in modelling the radio channel. Explicitly, for the modelling practice discussed here, signal strength field measurements had to be conducted beforehand (this was done at 469 MHz); to be precise, this paper primarily concerns empirically/statistically modelling the radio channel, and thus provides results obtained from empirically modelling the environments in question. This paper, on the whole, proposes three propagation models, corresponding to three experimented environments. Perceptibly, the models have been derived by way of making the most use of statistical measures. Generally speaking, the first two models were derived via simple linear regression analysis, whereas the third have been originated using multiple regression analysis (with five various predictors). Additionally, as implied by the title of this paper, both indoor and outdoor environments have been experimented; however, (somewhat) two of the environments are neither entirely indoor nor entirely outdoor. The other environment, however, is completely indoor.

***Keywords*—**RF propagation, radio channel modelling, statistical methods.

I. Introduction

PROPAGATION experts involved with wave propagation models, in terms of developing and/or using them, no doubt, discern the usefulness of such models, i.e., why they are somewhat indispensable in the wireless arena. Such practitioners will also have a handle on the intricacies behind the development of such models. However, folks new to this area might have doubts. This paper could help with these uncertainties.

The *DK Illustrated Oxford dictionary* defines a model, amongst other possible definitions, as 'a simplified (often mathematical) description of a system etc., to assist calculations and predictions.' It stands to reason that a radio propagation model (also referred to as path loss-, radio wave propagation-, radio frequency propagation-, or simply channel- model) is merely an equation (or set of equations) formulated to 'predict' the propagation of a radio signal within a specified environment.

Contrary to wired communication channels with characteristics that are rather easily predictable, radio channels are enormously random, resulting in their analysis being quite challenging. The radio path between a transmitter and receiver can be at variance from a simple 'line of sight' case (i.e., where there is direct unobstructed radio path between a transmitter and receiver) to one that is quite hampered by several possible objects, such as, buildings, walls, floors of multi-level buildings, vegetations, etc.

By radio-channel modelling, what is meant is that the amount of propagation path-loss obtainable within a specified environment is estimated and put forward for (future) estimation/prediction purposes, in addition to characterising the propagating channel's impulse response.

It is a generally accepted fact that no one propagation model applies in every situation. Rather, there are multitudes of models out there, each serving as a benchmark for a particular kind of environment. Accordingly, a propagation model developed by some researcher(s) or industry practitioner(s), for a particular kind of environment, couldn't only be useful to its formulator(s), but possibly also to other radio communications practitioners needful of such a model. In fact, this is why any such formulation would be referred to as a 'model' in the first place.

Customarily, [1], propagation models are concerned with predicting the expected path loss (or received signal strength) between a transmitter and a receiver, in addition to the variability of the strength of the signal in close proximity to a given position.

Propagation modelling is no doubt a very essential aspect of radio system design. It, however, [1], has traditionally been one of the most exigent parts of the design phase of radio systems deployment, and is habitually done using statistical measures, anchored in field measurements carried out

E.O.O. Ojakominor is currently at the final stages of his Master's studies (*M.Sc. in Personal, Mobile, and Satellite Communications*), conducted in full affiliation with the University of Bradford, UK, at INTI University College, (Laureate International Universities), Negeri Sembilan, Malaysia. (E-mail: ovijaks@yahoo.co.uk).

particular for a projected communication system or spectrum allocation.

The skill to accurately predict the behaviour of radio-propagation for wireless systems is essentially critical to system design [2]. Moreover, given the infeasibility in carrying out site measurements over and over again (largely due to the costs and time consumption involved), propagation models are generally seen as a low-cost, convenient, and suitable alternative [2].

The reader should note that RF propagation models are utilizable in virtually any sort of system involving the use of radio communication. They are particularly utilized (sometimes developed) during the planning and design phase of radio systems/networks. Examples of such systems include cellular networks, personal communications networks, ad hoc networks, sensor networks, WLANs, etc.

The reader can try visualising the propagation modelling concept this way; before a radio system (or some temporary radio service) is put up, why not check to see what the performance of the system could be before actually going ahead to put it up; why not make a well-versed estimate of how much loss could be incurred for possible radio paths in the radio system.

Case in point, the fact that a path loss model may be used to calculate, though approximately, the received signal level of a certain radio path in a given mobile communication system, does make it possible to envisage the signal-to-noise ratio of the system, which ultimately leads to being able to determine if the performance of the system will successfully meet service objectives and performance goals. If, at the end of the day, the performance (predicted by the propagation model) is inadequate, the system design would have to be modified accordingly, before eventually putting up the system.

Needless to say, over some years now, RF/wireless engineers have been particularly led to a concerted interest in understanding and predicting the characteristics of radio propagation in various urban as well as suburban environments, and even within buildings. This is largely the result of the immense commercial success having been attained by wireless communication systems (e.g., cellular systems) in addition to the ever increasing need for dependable wireless communication service. Now, given that this is most certainly going to continue, it is indeed imperative for RF engineers to have the necessary skills to predict/model a given environment's obtainable path loss, principally to avoid having to repeatedly conduct propagation measurements (for each particular environment), which aren't only time consuming but quite costly.

Typically, in developing a propagation model, the modeller studies the given environment, carries out relevant signal strength measurements within the specified environment (for empirical or site-specific modelling), followed by suitable measures to spawn the 'model.'

Generally speaking, radio-channel modelling may be necessary not just to estimate/predict the channel's path loss, but to characterise the propagating channel's impulse response. With mobile communications, for instance, the path loss, per se, is associated with designing base stations, given that this tells how much a transmitter ought to radiate to suitably service a specified region. On the other hand, channel characterisation deals with the dependability of the signal received at the receiver, and predominantly have to do with the make-up of he received waveform.

II. Theoretical Background

A. Introduction

This *Theoretical Background* contains information that is threefold: the first deals with EM wave propagation and the radio propagation channel; the second have to do with propagation models; and the third touches on some relevant statistical concepts. These aforementioned areas are quite broad in themselves, so only really relevant matters will be covered here.

B. Propagation & the Radio Propagation Channel

A simple 'bottom line' ought to be implicit on the subject of propagation; the propagated wave signal, in all circumstances, travels between two points, i.e., between the transmitter (that generates the wave signal) and the receiver (that receives it). These transmit and receive points are obviously the 'end systems' of the communication system in question. Now, matters relating to propagation studies will, by and large, focus on that part between the transmitter and the receiver, i.e., the communication channel, and in so doing, provisionally leaving out matters at the end systems.

Radio waves will generally propagate in one of four ways:

1. In a straight line, from the transmitter to the receiver.
2. In a curved line, where the curved line (i.e., the propagated radio wave) literally follows the earth's curvature.
3. In longer distances, due to the radio wave being somewhat trapped in the atmosphere.
4. In much longer (somewhat two-ray path) distances, where the radio wave is refracted back to the earth, off the ionosphere.

Before proceeding, the reader should recall that 'radio' propagation is fundamentally only possible within that part of the electromagnetic spectrum that extends from roughly within 30 kHz to 300 GHz.

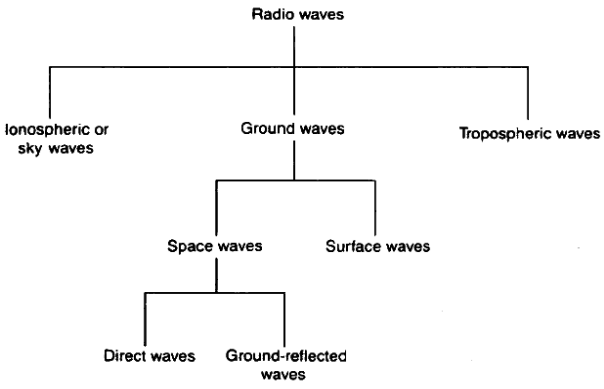

Fig. 1: Forms of radio propagation (*Source:* [3])

The Four ways in which radio waves may radiate (listed out earlier), apparently, depend largely on the operating frequency. Radio waves propagating close to the surface of the earth are termed *ground waves*; those that propagate not so high up, but much higher than the ground wave area (i.e., the lower atmosphere area – the troposphere) are referred to as *tropospheric waves*; and those that propagate very high up, via the layers of the ionosphere are *ionospheric* or *sky waves*. Fig.1 clearly depicts that ground waves are further subdivided into 'space' and 'surface' waves, wherein space waves are even further subdivided into 'direct' and 'ground-reflected' waves. The importance and usefulness of each of these wave types, in any given case, depends basically on the propagation path-length and on the operating frequency.

Having talked about the various forms of radio propagation and the way in which radio waves may radiate, more emphasis will now be put on the propagation channel itself. This is due to the rationale that this paper fundamentally concerns the radio channel, and modelling it. And, talking about modelling the radio channel, a useful first step will be the estimation of path loss obtainable after having taken RSS field measurements for various transmitter-receiver separations with the stipulated environment.

Apparently, *path loss* is the main constituent of a propagation channel [2]. It, basically, [2], is a measure of the average radio-wave attenuation experienced by a propagated signal when it reaches the receiver, after having navigated through a path of several wavelengths. Path loss is given by [2, 4]:

$$PL_{dB} = 10\log\frac{P_t}{P_r} \quad (1)$$

where P_t and P_r are the transmitted and received powers, correspondingly.

In reality, there are a number of mechanisms that may bring about radio transmission path loss (i.e., radio signal attenuation) as the radio wave navigates through the radio channel. Though these mechanisms are quite diverse, they are generally characterised into three; reflection, diffraction and scattering. See Fig. 2.

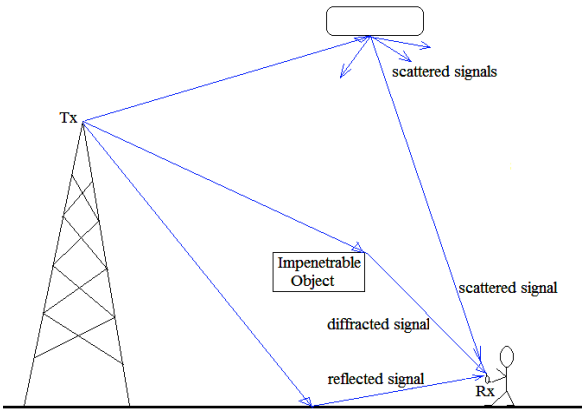

Fig. 2: A depiction of the three fundamental propagation mechanisms

On the whole, reflection results when a propagating EM wave comes in contact with an object that is enormously large in comparison to the propagating wave's wavelength. Reflection may take place from the surface of building walls, from the ground, and from furnishings. In addition to when reflection takes place, the EM wave may be (partially) refracted as well [2]. The operating frequency of the propagating wave, the polarisation of the wave, and the angle of incidence of the wave on the object upon which the reflection is to take place, all have a role to play on determining the coefficients of reflection (as well as refraction), which are generally functions of the material properties of the medium [2].

Diffraction results when the radio channel between Tx and Rx is impeded by a sharp 'knife-edge' surface. The resulting diffracted waves are present all over space [2], as well as at the rear of the obstructing object, bringing about the bending of the EM waves in the region of the obstructing object, thereby allowing EM wave receptions at locations where direct line of sight transmission path between Tx and Rx is infeasible. Moreover, as with reflection, [2], at higher frequencies, diffraction depends upon the geometry of the impeding object, in addition to the polarisation, amplitude, and phase of the incident EM wave at the diffraction point.

As for scattering, it occurs when the propagating medium of the EM wave comprise objects that are small in comparison to the wavelength, as well as where the number of impediments per unit volume is large [2]. Scattered waves are generated doe to some incident EM wave coming in contact with small objects, rough surfaces, or by other abnormalities in the channel. Entities that can induce scattering include vegetations, street signs, raindrops, stairs, lampposts, etc. Accurate prediction of scattered signal strength requires thorough knowledge of the physical details of the objects [2].

Propagations can perceptibly be done in both indoor and outdoor environments. Indoor and outdoor radio channels differ largely in terms of the transmitter-receiver separation distance covered (which is usually much smaller for indoor environments) and the variability of the environment (which is usually greater for indoor environments) [1, 2]. Propagation in indoor environments have, somewhat, more complex multipath structure than in outdoor environments, largely due to the nature of the building structures used, the room layouts, and the type of materials used in the construction of the building [2].

Outdoor environments are often categorised into three: urban, suburban, and rural. Outdoor terrain profiles, [2], need to be taken into consideration for any given outdoor environment, and may vary from a plain, curved earth to an immensely mountainous area. Possible propagation impediments such as trees, buildings, and moving vehicles also need to be taken into account when characterising RF propagation for outdoor environments [2].

Generally speaking, there are various signal fading types. The type of fading suffered by a signal propagating via a radio channel depends both on the nature of the transmitted signal and on the channel characteristics. Different transmitted

signals will generally experience dissimilar fading types, in accordance with the relation amongst the signal parameters (e.g., the bandwidth, the path loss, the symbol period, etc.) and the channel parameters (e.g., the RMS delay and the Doppler spread) [2].

Signal fading is generally split into two main types: large-scale and small-scale. Large-scale fading is impinged on, for the most part, by the presence of forests, buildings, hills, etc., lying between the transmitter and the receiver. The statistics of large-scale fading, [2], proffers a means to compute an estimate of the path loss as function of distance and other relevant factors. Small-scale fading, on the other hand, [5], is an attribute of RF propagation resulting from the presence of reflecting- and scattering- objects that cause multiple versions of the propagated signal to reach the receiver, each distorted in angle of arrival, amplitude, and phase.

It ought to be noted that, [5], the wireless channel is a function of what is transmitted over it. And, [5], to verify whether communication over the wireless channel is influenced by fading, the symbol duration of data transmission over it must be compared with coherence time, and the bandwidth of the baseband signal (i.e., fast or slow fading) compared with the coherence bandwidth of the channel (i.e., flat or frequency-selective fading). Note that all of the aforementioned four fading types (i.e., frequency-selective, flat, fast, and slow fading) are of the small-scale type.

It have got to also be apparent that, [5], specifying a channel as encompassing fast or slow fading doesn't stipulate whether it's a frequency selective or flat fading channel. These classifications are independent in their own right [5]. Flat or frequency-selective fading plain have to do with whether the relationship between the signal bandwidth and frequency range of the fading behaviour of the channel is uniform. Fast and slow fading unambiguously have to do with the rate of change of time of the channel in respect of the transmitted signal.

Apparently, a frequency-selective channel, [2], is one wherein the delay spread is larger than the symbol period, resultant each time the received multipath components of a symbol spreads out further than the symbols' time duration, which, in turn, results in a sort of inter-symbol interference (ISI) referred to as channel-induced ISI. On the contrary, a flat fading channel is purely one wherein the delay spread is lesser than the symbol period, and yields no channel-induced ISI. However, for the flat fading channel, [2], there still can be some degradation in performance, due to the irresolvable phasor components that destructively add up to yield, at the receiver, a considerable drop in signal to noise ratio.

As for fast and slow fading, they are perceptibly categorized based on how quickly the transmitted signal changes, in comparison with the rate of change of the channel's electrical-parameter [2]. Explicitly, a fast fading channel is one in which the channel's impulse response changes at a rate much faster than the transmitted baseband signal, while a slow fading channel is one wherein the channel's impulse response changes at a much slower rate than the transmitted baseband signal.

It is useful to note that fast and slow fading generally have to do with situations where the transmitter and/or receiver moves or is moveable (i.e., not entirely stationary). As a result, the speed of the mobile unit or the speed of the objects utilising the channel, through which the transmitted signal propagates, determines if a signal will be experiencing fast fading or slow fading. The faster the rate of change of the location of the mobile unit, the more likely the channel will be undergoing 'fast fading,' and vice versa.

C. Propagation Models

Evidently, there are a number of existing propagation models. The most basic of all may well be the well-known Friis free space transmission equation/model, given by:

$$P_{r,linear} = P_t G_t G_r \left(\frac{c}{4\pi f d}\right)^2 \quad (2)$$

$$P_{r,dB} = 10\log_{10} P_t + 10\log_{10} G_t + 10\log_{10} G_r + 147.558 - 20\log_{10} f - 20\log_{10} d \quad (3)$$

P_t and P_r are apparently the transmitted and received powers, respectively. G_t and G_r are, correspondingly, the transmit- and receive- antenna gains. f is the specified operating frequency, d is the Tx-Rx separation distance, and the exponent '*2*' (see linear form) connotes the rate at which the propagated signal deteriorates with increasing d, proportionate to free space propagation.

The free space model is widely known to only be useful in scenarios where the radio channel has no attenuating impediments, other than mere distance increments (in free space). That is, the radio wave propagates in a straight line from the transmitter to the receiver (in unobstructed LOS fashion). Examples of practical free space propagation applications include satellite communications and certain microwave radio links.

It should be noted that there are times when the *system loss factor* (not associated with propagation) is taken into account. In such situation, the Friis free space equation becomes:

$$P_r = P_t G_t G_r \left(\frac{\lambda}{4\pi d}\right)^2 \cdot \frac{1}{L} = P_t G_t G_r \left(\frac{c}{4\pi f d}\right)^2 \cdot \frac{1}{L} \quad (4)$$

L represents the system loss factor ($L \geq 1$). λ refers to the wavelength (in meters). Clearly, as the system loss factor increases the received power diminishes accordingly.

It must be noted here that the Friis equation doesn't hold for $d = 0$, else the received power tends to infinity, which isn't practical. As a result, [2], more than a few propagation models make use of a different representation for the *close-in distance*, d_0, referred to as the received-power *reference distance*.

In general, when modelling a given environment, [1], it is important to choose a free space reference distance that is apt for the propagation environment under consideration. For large coverage cellular systems, a reference distance of 1km is ideal, whereas in microcellular systems, much smaller reference distance of roughly 1m or 100m is appropriate [1]. Additionally, a chosen reference distance ought to always be in the *far-field* (*Fraunhofer*) *region* of the antennas –

especially noteworthy for indoor short range modelling – so as to avoid having *near-field* (*Fresnel*) *region* effects interfering with the reference received signal strength.

Apparently, the free space propagation model isn't germane for realistic radio channels, especially where the transmitter and/or receiver is movable. Path loss models will usually utilise a *path loss exponent* parameter, γ, to indicate the power-law relationship between the Tx-Rx separation distance and the received power. In view of that, path loss is expressed (in decibels) as:

$$PL_{dB}(d) = PL(d_0) + 10\gamma \log\left(\frac{d}{d_0}\right) + X_\sigma \qquad (5)$$

For eqn. (5), γ = 2 clearly characterises free space path loss. γ will usually take on a higher value when obstructions are present. X_σ symbolises a *zero-mean* Gaussian random variable of standard deviation σ, revealing, on average, the received power variation that naturally results when a path loss model of the type in eqn. (5) is used. Table I lists typical path loss exponent values for six different types of environments.

Path loss models can, in general, be classified into two main groups [2]:
1. Statistical models (also referred to as empirical models)
2. Site-specific models (also known as deterministic models)

Empirical models are generally characterised as [2]:
a. Based on statistical characterisation of the received signal,
b. Require less computational effort
c. Less sensitive to the geometry of the environment
d. Easier to implement.

Conversely, site-specific models [2]:
a. Have a certain physical basis
b. Require an immense amount of data pertaining to geometry, locations of building and of furniture in buildings, terrain profile, etc.
c. Require more computations
d. More accurate

The empirical modelling approach generally tends to reorganize a set of measured data by fitting curves or analytical expressions to the data. This (empirical) approach, no doubt, [1], has the benefit of unreservedly taking into consideration all factors of propagation, whether known or unknown, by way of real field measurements.

The reliability of an empirical model at operating frequencies and environments differing from those with which the models were originally developed can be made certain by means of additional data gathered through new field measurements within the differing environment, at the required operating frequency.

Table I: Typical path loss exponents for different environments (*Source:* [1])

#	Environment Type	Path Loss Exponent, γ
1	Free Space	2
2	Urban area cellular radio	2.7 to 3.5
3	Shadowed area cellular radio	3 to 5
4	In building line-of-sight	1.6 to 1.8
5	Obstructed in building	4 to 6
6	Obstructed in factories	2 to 3

Table II: Comparing between some of the main path loss models (*Source:* [2])

Model Name	**Suitable Environ-ment**	**Com-plexity**	**Experi mental Data**	**Details of Accuracy**	**Accuracy (Time)**
Okumura Model	Macro-cell	Simple	Based on experi-ments	No	Good (Little)
Hata Model	Macro-cell (early cellular)	Simple	No	No	Good (Little)
COST-231	Microcell (outdoor)	Simple	No	No	Good (Little)
Dual-Slope	Microcell and picocell (LOS region)	Simple	No	No	Good (Little)
Ray-Tracing	Outdoor and indoor	Com-plex	No	Yes	Very Good (Very Much)
FDTD	Indoor (small)	Com-plex	No	Every detail	Best (Very Much)
MoM	Indoor (small)	Com-plex	No	Every detail	Best (Very Much)
ANN	Outdoor and indoor	Com-plex	Yes	Detail	Very Good (Very Much)

As already accentuated, there are a number of existing path models. Discussing all of them here will only make this report bulky. Fortunately, they are well treated in the literature. These models will generally be used in indoor and/or outdoor environments. To give an idea, some of the most used path loss models have been listed out in Table II. It can be seen from the Table that the modelling approach taken proffers an incredibly different trade-off between complexity and accuracy. Of the models listed out in the Table, the Okumura, Hata, COST-231, and Dual-Slope models are empirically derived models, whereas the Ray-Tracing, Finite-Difference Time-Domain (FDTD), Method of Moments (MoM), and Artificial Neural Network (ANN) models are site-specific. Apparently, the empirical models are simple, i.e., no environmental information is used, besides the choice of the parameters [2]. Moreover, although these empirical models are well in use for path loss predictions, they aren't very accurate. Site-specific models, on the other hand, appear to be considerably more accurate than their empirically derived counterparts. However, they require a great deal 'site-specific' information on the environment under consideration (at the very least, [2], the locations of all of the objects in the environment, and conceivably the locations of large objects).

Given that this paper precisely concerns statistically modelling field measured data, a little more emphasis will be put on this. Some discussions have already been done on the four types of signal fading over a propagation channel (i.e., frequency-selective, flat fading, fast fading and slow fading). Here, flat fading and slow fading will be described a little further, in comparison with the standard empirical path loss model.

Slow and fast fading may be described as follows [6]:

- Slow fading: Fading attributable to 'shadowing' and other site-specific properties of the propagation path
- Fast fading: Fading caused by interference between signals propagating over many paths from the transmitter to the receiver

Fig. 3 presents a depiction of the typical behaviours of the slow and fast fading propagation mechanisms. Note that the abscissa here is range, i.e., Tx-Rx separation distance, in km. However, [6], these curves may well be inferred as versus time (i.e., the abscissa = time) for a mobile user, with increasing range. A typical prediction from an empirical path loss model is described in Fig. 3 (a), portraying a smooth and slowly increasing estimated path loss versus range. It should, however, be expected that variations caused by the obstacles encountered by the radio wave on a particular path, as range is increased certainly induce variations in the path loss [6]. This is particularly the phenomenon of slow fading, and is illustrated in Fig. 3 (b). Fig. 3 (c) clearly illustrates the rapid signal power variations that occur when fast fading is included. From the Figure, it is perceptible that empirical path loss is merely some straight line, which gradually increases, generally, with such a mannerism. Slow and fast fading channels will also have a generally increasing path loss. The rate of change of the path loss is then what determines if it's a slow fading or fast fading channel. The slow and fast fading models will also generally be the result of including, i.e., adding, slow or fast fading effects to the empirical path loss model, respectively.

In order to have a handle on how the slow or fast fading effects can be obtained, a little probability theory is given here. The statistics behind how the empirical (straight line) model is obtained is discussed soon after.

Probability theory is generally applied in situations where experiments taken have unknown consequences. For propagation modelling, the theory is applied to model path loss (or the total received power) when slow or fast fading is included on the propagation path at a specific time.

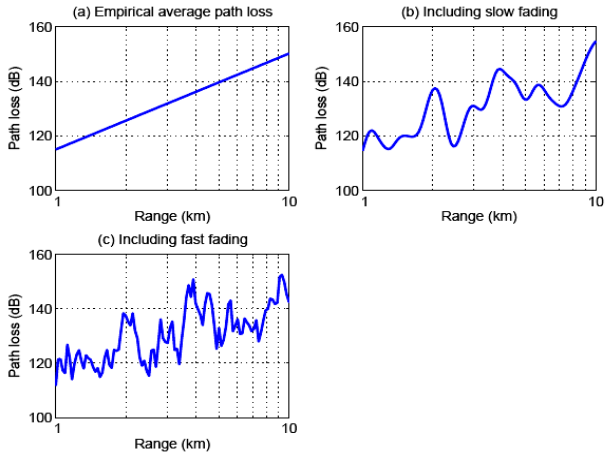

Fig. 3: Depiction of classic path loss behaviours; (a) Expected path loss w.r.t. an empirical model, (b) path loss behaviour with slow fading effects included, (c) path loss behaviour with the effects of fast fading additionally included (*Source:* [6])

Path loss is clearly an 'unknown' given the lack of absolute knowledge of the environment, as well as the complicatedness of performing accurate calculations even when such information is accessible.

The slow fading loss has been assumed here for the measurements carried out for the models to be presented later, given that the rate of change of the various Tx-Rx separation distances measured isn't so severe as to warrant the involvement of fast fading effects. In fact, for each measurement, both the transmitter and receiver were rather fixed. Accordingly, only the slow fading loss will be discussed henceforth.

The path loss due to slow fading (L_s), [6], is represented in decibels, and adds to the empirical path loss (L_E) to represent the total unknown path loss (L_T) in a given measurement. The slow fading loss (the fast fading loss, as well) is a *continuous* random variable (r.v.), seeing that the path loss on a particular path can take on a continuous range of values.

Probability theory states that, [6], *absolute knowledge* of a single r.v. is provided through knowledge of its PDF. The PDF gives a picture of the percent of time that the measured fading loss is likely to fall within a small range Δl_s of a stipulated value l_s [6]. The probability density function (PDF), $f_{L_s}(l_s)$, of the slow fading loss, Ls, can be given by [6]:

$$P(|L_s - l_s| < \Delta l_s) = f_{L_s}(l_s)\Delta l_s \tag{6}$$

where P(·) is the probability of obtaining the experiment outcome $|L_s - l_s| < \Delta l_s$. It should be noted that $0 \leq P(\cdot) \leq 1$, where $P(\cdot) = 0$ signifies that the outcome never occurs and $P(\cdot) = 1$ implies the outcome always occurs. I.e., higher values of P(·) indicate a more likely event.

The cumulative distribution function (CDF) of slow fading loss, L_s, describes the probability that $L_s \leq l_s$. Given the PDF definition, the slow fading CDF, $F_{L_s}(l_s)$, can be defined as [6]:

$$F_{L_s}(l_s) = P(L_s \leq l_s) = \int_{-\infty}^{l_s} f_{L_s}(l_s)\ dl \tag{7}$$

The generalised PDF of a Gaussian distributed r.v. X is given by:

$$f_X(x) = \frac{1}{\sigma_X\sqrt{2\pi}} e^{-\frac{(x-\mu_X)^2}{2\sigma_X^2}} \tag{8}$$

where μ_X and σ_X^2 are the expected value and variance of X, respectively.

If the local attenuations in a given slow fading channel are modelled as random variables, every one of them having identical PDF to the others, as the path is traversed; the slow fading loss L_s is then a sum of independent and identically distributed (i.i.d.) random variables [6]. It is therefore logical to model the slow fading loss (in decibels) as a Gaussian PDF r.v., given by [6]:

$$f_{L_s}(l_s) = \frac{1}{\sigma_{L_s}\sqrt{2\pi}} e^{-\frac{{l_s}^2}{2\sigma_{L_s}^2}} \tag{9}$$

where a 0 dB mean have been assumed for the slow fading loss L_s, due to the necessary separation of L_s from the empirical average path loss term L_E. That is, L_s is a zero-mean

Gaussian variable with standard deviation σ_{L_s}. l_s is the stipulated value.

D. Some Statistics Background

The mean (μ), standard deviation (σ) and variance (σ^2) are especially useful terms that will always come up in the modelling of a given set of data. So, they certainly deserve some quick talk.

Fig. 4 presents a diagrammatic illustration of the *mean* value. The Figure is rather self explanatory. At this point, it is worth mentioning that different people have changed names for the term *mean*, all however, referring to exactly the same thing. Some refer to it simply as the *average*. Statisticians refer to it as the *mean*, while engineers often refer to it as the *expectation*. They all mean precisely the same thing; i.e., the 'centre' of a given data set.

Fig. 5 clearly illustrates the standard deviation phenomenon, expressly bringing up the Gaussian/normal kind of distribution of data. The Gaussian (also called, normal) distribution is, in point of fact, the most important type of distribution (of data) in all of probability and statistics [7], and that which is used throughout this project. As portrayed in Fig. 5, the standard deviation simply measures the 'fatness' of the data distribution, i.e., how far the values, in the data set, are from the centre (i.e., the mean). The distribution is wider when the numbers, in the data set, are farther away from the mean-value of the data set. Thus, the further away the numbers are from the mean, the wider the distribution, and consequently, the larger the standard deviation.

Crucial to analysing data are scatterplots. They use a series of points to represent a set of data, on a two-dimensional coordinate system. Fig. 6 presents four possible forms. In analysing scatterplots, [8], the first thing to consider is the presence of association (i.e., a definite pattern in the scatterplot), after which the type of association should be looked at; i.e., whether the association is linear (whether the pattern follows a line) or nonlinear.

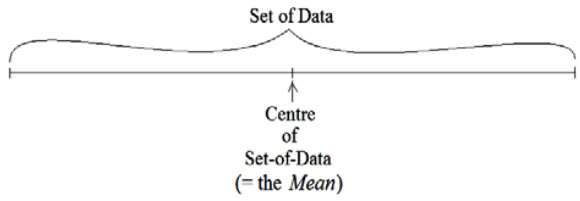

Fig. 4: Diagrammatic illustration of the *mean* value

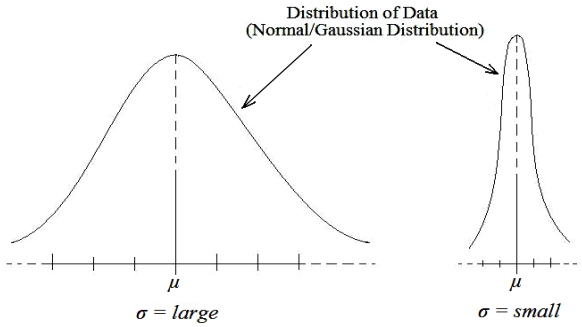

Fig. 5: Diagrammatic illustration of the *standard deviation* phenomenon

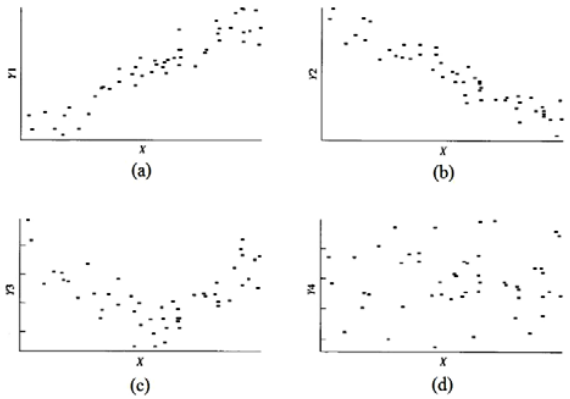

Fig. 6: Analysing scatterplots based on presence / type of association. (a) positive linear association, (b) negative linear association, (c) nonlinear association, (d) no association. (*Source:* [8])

Linear association is generally twofold; positive or negative. For nonlinear association, there are various possible forms. Furthermore, after acknowledging the type of association, another thing to check is the strength of the association [8]. A linear association is strongest when all data points stretch out on a thin straight line. The strength of the association perceptibly becomes weaker and weaker as the data points turn out more and more scattered all over the place (as in Fig. 6 (d)).

Regression analysis is a statistical concept pertinent in the *modelling* of a given set of data. For path loss modelling, in particular, it is one of the most used techniques (see, for example, [9-14]). This is plainly because, [7], it brings about the exploration of the association between two or more variables related in a nondeterministic manner; nondeterministic in the sense that, even though the variables are related, the values of one of the variables isn't completely specified just by knowing the other(s). That is, a certain level of uncertainty exists. Say $y = 2 + x$; then in this case, y is a random variable. That is, the values of variable y may turn out to be anything at any fixed value of variable x. Furthermore, with the regression analysis methodology, previously acknowledged values of x (say, data from signal strength measurements) are quite helpful in predicting what signal strength (or, path loss) observations could turn out in the future (under similar environment settings, of course).

Regression analysis may be applied to obtain, [7], the simple linear regression (SLR) model, the nonlinear model, or the multiple regression model. The SLR model will usually be used to explain variation in a linearly associated data. When the proportion of observed variation explained by the SLR model is inappropriate – i.e., some point(s) in the data set is/are too spread out, making data association not linear enough for the SLR model to be used – an analyst will habitually seek out for a substitute model (either a nonlinear or multiple regression model), which may more effectively explain the variations in the overall observed data.

The SLR model is given by [7]:

$$y = \beta_0 + \beta_1 x + \varepsilon \qquad (10)$$

where β_0 is the y-intercept, β_1 is the slope of a straight line determined by the set of pairs (x, y), ε is the random deviation (else referred to as the random error term) assumed to be normally distributed with mean = 0 and variance = σ^2. Apparently, the parameters β_0, β_1 and σ^2 are such that, for a specified value of the independent variable x, the dependent variable y is related to x by way of eqn. (10).

Generally, [7], the value of x in eqn. (10) is one that is fixed by the experimenter, referred to as the *predictor, independent,* or *explanatory variable,* while the value of y is random, and generally referred to as the *dependent* or *response variable*.

The dependent (r.v.) variable will be denoted here by Y, while y would be used to denote its *observed (measured) value*. For instance, say the RSS for a Tx-Rx separation distance of 40m is -36dBW, then y = -36dBW is the observed value of Y correlated with setting x = 40m. Now, let the predictor variable for which observations are made be denoted by x_1, x_2, ... , x_n, and let the r.v. and observed value correlated with x_i be denoted by Y_i and y_i, respectively. Thus, the (bivariate) data consists of n pairs; (x_1, y_1), (x_2, y_2), ... , (x_n, y_n).

It can be seen that without ε in eqn. (10), any observed pair (x, y) will symbolize a point falling precisely on the straight line slope $y = \beta_0 + \beta_1 x$, known as the *true* (or *population*) *regression line*.

Regression analysis merely entails 'fitting' a set of data to the true regression line (sometimes curve) – see Figure 2.8. From the Figure it can be seen that the appropriateness of the SLR model here may be advocated by theoretical deliberation (e.g., whether there's a precise linear association between the two variables, w.r.t. ε).

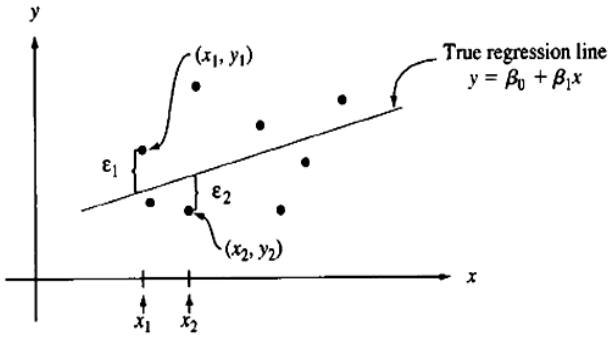

Fig. 7: Depiction of the *true regression line* inserted in a scatterplot of observations from the SLR model (*Source:* [7])

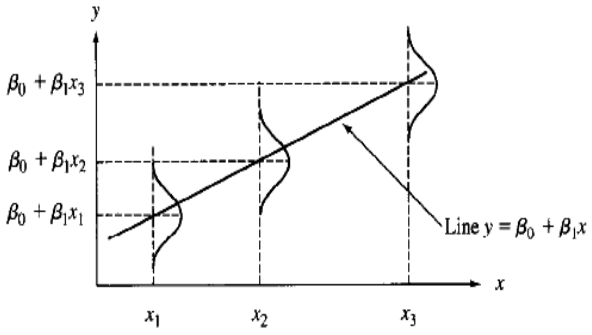

Fig. 8: Distribution of r.v. Y for different x values (*Source:* [7])

A few useful points ought to be noted from Fig. 8. First, [7], the slope β_1 of the *true regression line* has a possible interpretation as the *expected* change in Y associated with an increase (by 1-unit) in the x value. Second, [7], the amount of variability in the distribution of the values of Y is the same at every different x value (homogeneity of variance). Additionally, take the path loss measurement scenario, for example; it can be then gathered from the Figure that average path loss changes linearly with Tx-Rx separation distance, and that the amount of variability in path loss at any given Tx-Rx separation distance is the same as at any other Tx-Rx separation distance. To grasp this point, each Tx-Rx separation distance should be treated separately. Now, note that a series of path loss measurement trials are usually undertaken for each Tx-Rx separation distance. Accordingly, what the Figure implies is that the amount at which those various measurement trials vary at a particular Tx-Rx separation distance is somewhat the same as the way measurement trials would vary at any other particular Tx-Rx separation distance. Third, [7], for a particular value of x, Y is the sum of a constant ($\beta_0 + \beta_1 x$) and a normally distributed r.v. ε; as a result, Y itself is normally distributed.

Some useful statistical concepts have been presented, with most emphasis on the SLR model, though. Other relevant matters will be conferred later while discussing the models.

III. Measurement Devices and Environments

Fig. 9: Emmanuel Ojakominor carries out preliminary checks on transmit and receive devices before commencing actual measurements. Notice the transmit device (walkie-talkie, model T-388, with output power approximately 0.5W) and the receive equipment (Tektronix 2711 spectrum analyser along with a simple telescopic antenna)

Fig. 10: Receiver position for environment I

Fig. 11: Emmanuel Ojakominor takes measurements in environment I. Notice the make-up of the environment, having a number of (glass) windows at both walls, as well as objects such as metallic lockers located at adjacent positions along the passageway.

Fig. 12: Emmanuel Ojakominor at the (stationary) receiver location corresponding to measurement settings for models II and III

Fig. 13: A depiction virtually covering measurement surroundings for both models II and III. Notice the passageway for model II, which continues to the left looking straight up through the passage. This passage is approximately 64.008m long (measuring from the receiver location to the extreme of the passage, looking right up).

Fig. 14: A depiction of the ground and surrounding area corresponding to model III. Notice the foliage, which may well have shadowing effects on radio paths between a transmit spot on the ground area and the receiver located up at level four of building.

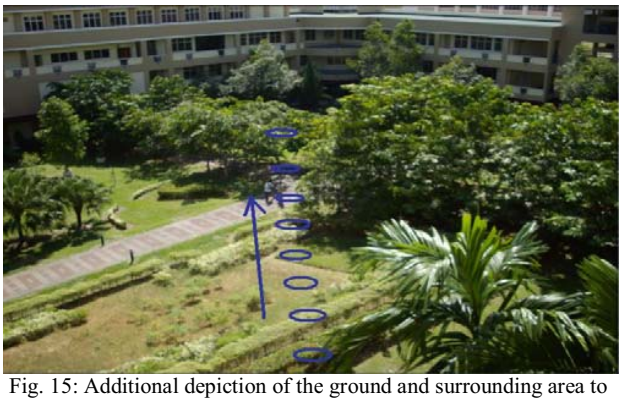

Fig. 15: Additional depiction of the ground and surrounding area to Fig. 14 The circles give an indication of the several transmit points, while the arrow points to the direction of increase in distance from the fixed receiver (i.e., the stationary receiver positioned up at level 4 of building from where this picture was taken).

IV. Results And Discussions

A. Preliminary Analyses (of Measured Data)

As earlier emphasized, a very effective way to begin looking at a set of data is to produce graphical representations of the data. In this *Preliminary Analyses* section, the measured data will be reviewed graphically in both the mean-RSS and -path loss forms.

At this point, it should be noted that, usually, after the effects of system losses and antenna gains have been controlled for, path loss may be delineated simply as the ratio between transmitter-output-power and the receiver-received-power. Consequently, for each measured RSS, path loss has been estimated using the following simple relationship:

$$\begin{aligned} PL\ [dB] &= P_{Tx}\ [dBm] - P_{Rx,\ Measured}\ [dBm] \\ &\equiv P_{Tx}\ [dB] - P_{Rx,\ Measured}\ [dB] \end{aligned} \quad (11)$$

Where PL [dB] is the estimated path loss (in decibels), P_{Tx} [dB] is RF output power generated by transmit antenna (in decibels), and $P_{Rx,\ Measured}$ [dBm] is the measured RSS (in decibel-milli-Watts).

Apparently, eqn. (11) basically says; (1) this is what was transmitted, (2) this is what has been received, and so, (3) the

path loss is the difference between the two (in dB). Path loss will normally encompass distance dissipation in addition to other channel losses such as multipath effects, penetrations, shadowing, etc.

Table III summarises field measurement results for model-I. Note that the path loss values stated there – as well as those in the rest of the other data index tables – have been obtained via conversion from RSS values, using eqn. (11). Also note that these values are mean values. The individual measurement findings from which the mean RSS values have been estimated aren't provided in this paper.

Evidently, the scatterplots of Fig. 16 and Fig. 17 each advocate the suitability of the SLR model; i.e., the association of data for either case appears to be substantially linear.

Table III: Data index corresponding to model-I

P_{Tx} = 0.5 W = [10log$_{10}$(0.5)] dBW = -3.0103 dBW				
Obs.	Tx-Rx Separation		Mean RSS	Mean Path Loss
#	[ft]	[m]	[dB]	[dB]
1	10	3.048	-6.021	3.0107
2	20	6.096	-15.041	12.0307
3	30	9.144	-33.152	30.1417
4	40	12.192	-46.061	43.0507
5	50	15.24	-50.034	47.0237
6	60	18.288	-39.172	36.1617
7	70	21.336	-43.060	40.0497
8	80	24.384	-62.075	59.0647
9	90	27.432	-58.892	55.8817
10	100	30.48	-69.069	66.0587
11	110	33.528	-72.041	69.0307

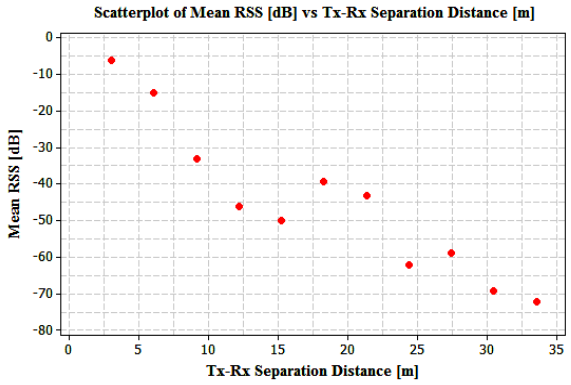

Fig. 16: Scatterplot of mean RSS vs Tx-Rx separation for model-I

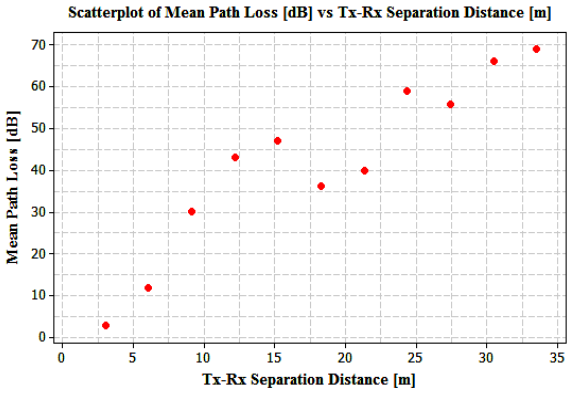

Fig. 17: Scatterplot of mean path loss vs Tx-Rx separation (model-I)

It follows that there is a rather strong predilection for RSS to reduce as Tx-Rx separation increases, and for path loss to increase as Tx-Rx separation increases. That is, larger values of RSS appear to be associated with smaller values of Tx-Rx separation (i.e., negative relationship between the variables), while larger values of path loss appear to be in line with larger values of Tx-Rx separation (i.e., positive relationship).

Table IV summarises findings for model-II. Apparently, as with observed-data for model-I, the field measured data for model-II also appears rather linear. That is, there is a generally decreasing trend for mean RSS, and a generally increasing pattern for mean path loss. Fig. 18 and Fig. 19 provide graphical depictions for data corresponding to mean RSS and mean path loss, respectively.

Unlike model-I and model-II, model-III incorporates some fair amount outdoor involvement. And, to get things clearer with the environment/condition which model-III symbolises, extra depictions are given in Fig. 20 and Fig. 21.

Table IV: Data index corresponding to model-II

P_{Tx} = 0.5 W = [10log$_{10}$(0.5)] dBW = -3.0103 dBW				
Obs.	Tx-Rx Separation		Mean RSS	Mean Path Loss
#	[ft]	[m]	[dB]	[dB]
1	10	3.048	-12.007	8.9967
2	20	6.096	-20.000	16.9897
3	30	9.144	-29.897	26.8867
4	40	12.192	-34.895	31.8847
5	50	15.24	-18.862	15.8517
6	60	18.288	-49.119	46.1087
7	70	21.336	-60.915	57.9047
8	80	24.384	-56.930	53.9197
9	90	27.432	-56.149	53.1387
10	100	30.480	-68.995	65.9847
11	110	33.528	-67.959	64.9487
12	120	36.576	-69.119	66.1087
13	130	39.624	-70.229	67.2187
14	140	42.672	-73.639	70.6287
15	150	45.720	-82.939	79.9287
16	160	48.768	-81.463	78.4527
17	170	51.816	-83.098	80.0877
18	180	54.864	-80.172	77.1617
19	190	57.912	-90.020	87.0097
20	200	60.960	-89.191	86.1807
21	210	64.008	-79.062	76.0517

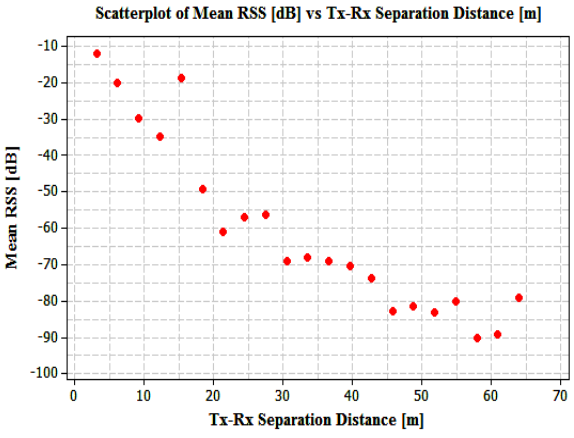

Fig. 18: Scatterplot of mean RSS vs Tx-Rx separation for model-II

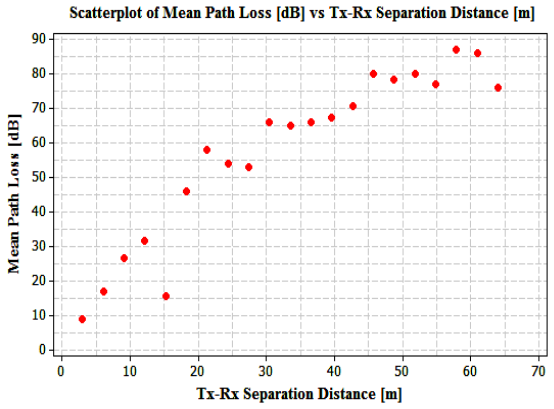

Fig. 19: Scatterplot of mean path loss vs Tx-Rx separation (model-II)

Table V: Data index corresponding to model-III

P_{Tx} = 0.5 W = [10log$_{10}$(0.5)] dBW = -3.0103 dBW				
Obs. #	Tx-Rx Separation [m]	Tx-building separation [m]	Mean RSS [dB]	Mean Path Loss [dB]
1	12.552	6.096	-82.938	79.9277
2	16.401	12.192	-72.995	69.9847
3	21.327	18.288	-64.013	61.0027
4	26.740	24.384	-76.138	73.1277
5	32.394	30.480	-63.876	60.8657
6	38.185	36.576	-73.979	70.9687
7	44.059	42.672	-81.110	78.0997
8	49.987	48.768	-72.956	69.9457
9	55.949	54.864	-84.098	81.0877
10	61.938	60.960	-82.914	79.9037
11	67.949	67.056	-82.270	79.2597
12	73.969	73.152	-84.098	81.0877
13	80.004	79.248	-79.016	76.0057

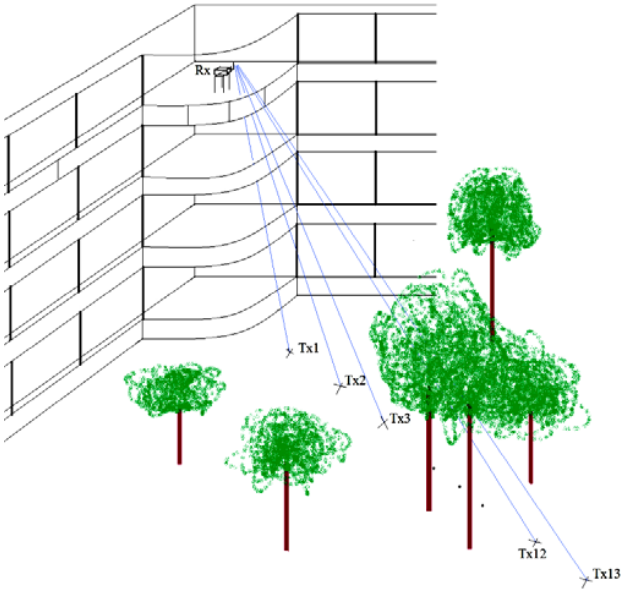

Fig. 20: 3D depiction of measurement setup/scenario for model III

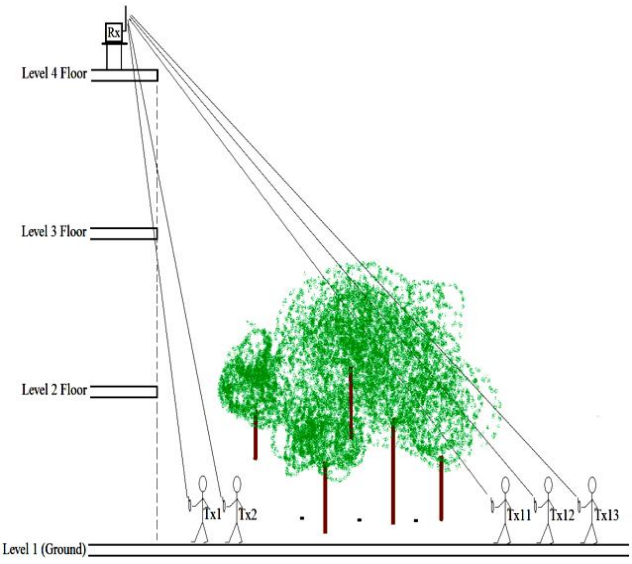

Fig. 21: 2D depiction of measurement setup/scenario for model III

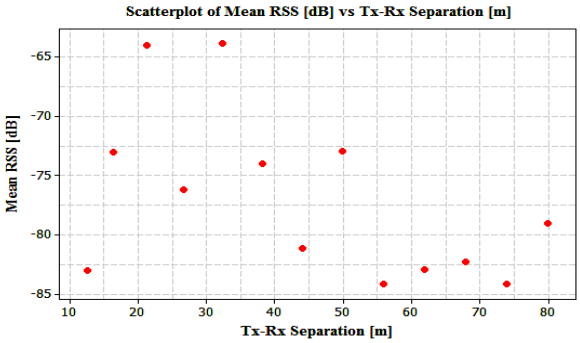

Fig. 23: Scatterplot of mean RSS [dB] vs Tx-Rx separation for model-III

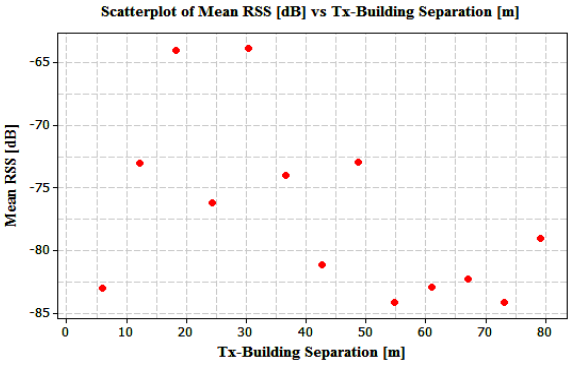

Figure 24: Scatterplot of mean RSS [dB] vs Tx-Building separation, for model-III

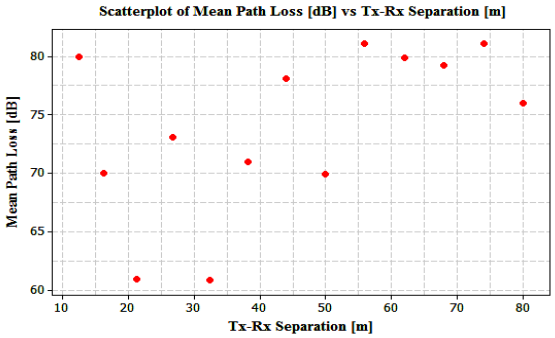

Fig. 25: Scatterplot of mean RSS [dB] vs Tx-Rx separation for model-III

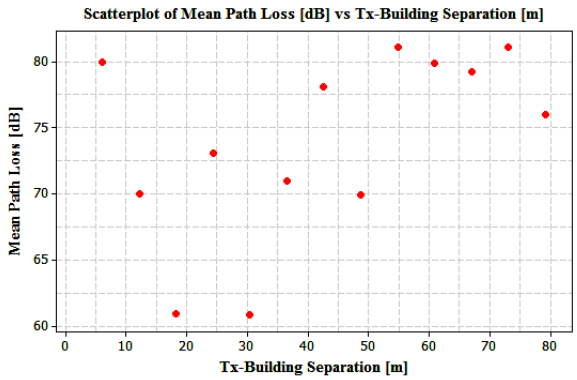

Fig. 26: Scatterplot of mean RSS [dB] vs Tx-Building separation [m] for model-III

Clearly, the scatterplots of Fig. 23 to Fig. 26 do not depict a very linear relationship between the data. More emphasis on this is given in the next section.

B. Modelling Procedures

Bear in mind that all that's originally available is purely the set of data consisting of n observed path loss pairs (Tx-Rx_1, PL_1), (Tx-Rx_2, PL_2), ... , (Tx-Rx_n, PL_n), from which a constructive educated-guess on the model parameters β_0, β_1, σ^2, and thus, the true regression line, are obtained.

The path loss variable, here, is perceptibly an r.v. PL_i, where:

$$PL_i = \beta_0 + \beta_1 + \varepsilon_i \qquad (12)$$

ε_i are the n independent random deviations ε_1, ε_2, ε_3, ... , ε_n.

Looking at Fig. 27, it can be seen that the least squares line truly provides a reasonably good fit to the observed path loss data of model-I, since the deviations (i.e., the vertical distances) from the observed points to the line (from above or below) are rather the least possible, putting together a strong likelihood of it being the best-fit line.

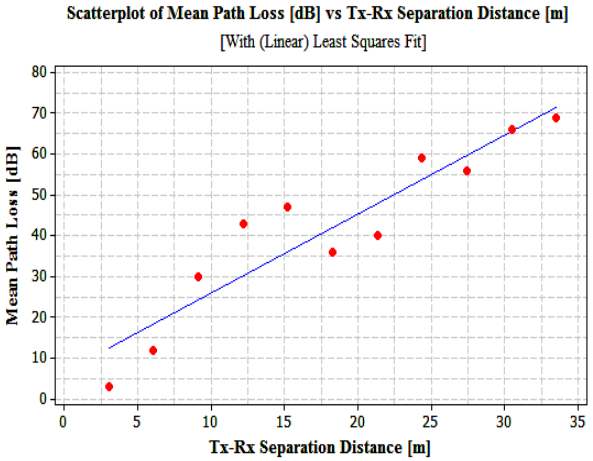

Fig. 27: Scatterplot of mean path loss [dB] vs Tx-Rx separation [m]; with least squares line superimposed (model-I)

Least squares estimates (LSEs) of the true regression line (TRL) are given by [7]:

$$\acute{\beta}_1 = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2} = \frac{\sum x_i y_i - [(\sum x_i)(\sum y_i)/n]}{\sum x_i^2 - [(\sum x_i)^2/n]}$$
$$= \text{LSE of the slope coefficient } \beta_1 \text{ of the TRL}$$

$$\acute{\beta}_0 = \frac{\sum y_i - \acute{\beta}_1 \sum x_i}{n} = \bar{y} - \acute{\beta}_1 \bar{x}$$
$$= \text{LSE of the intercept } \beta_0 \text{ of the TRL}$$

Referring to Table III, where y = path loss, x = Tx-Rx separation and n = number of observations = 11, the LSEs $\acute{\beta}_1$ and $\acute{\beta}_0$ for the ERL of model-I are thus as follows:

$$\acute{\beta}_1 = \frac{10416.6 - [(201.168)(461.505)/11]}{4700.89 - [(201.168)^2/11]} = \frac{1976.59656}{1021.929616}$$
$$= 1.93418072 \cong 1.934$$

$$\acute{\beta}_0 = 41.9550 - (1.93418072)(18.288)$$
$$= 6.582702993 \cong 6.583$$

Hence, results show that an estimate of the expected change in path loss with respect to a 1-meter increase in Tx-Rx separation is approximately 1.934 dB, with $\acute{\beta}_0 \cong 6.583$.

Thus, the equation of the estimated least squares line (LSL) (regression line) of Fig. 27 is:

$$\acute{PL}_i \text{ [dB]} = 6.583 + 1.934 \text{ Tx-Rx}_i \qquad (13)$$

Statistically, this equation can be used to make point predictions for random variable PL that falls exactly on the LSL, however strictly for Tx-Rx separations within same range as specified here.

Table VI: MINITAB output for regression of data corresponding to model-I

The regression equation is
Mean Path Loss [dB] = 6.58 + 1.93 Tx-Rx Separation Distance [m]

Predictor	Coef	SE Coef	T	P
Constant	6.583	5.251	1.25	0.242
Tx-Rx Separation Distance [m]	1.9342	0.2540	7.61	0.000

S = 8.12062 R-Sq = 86.6% R-Sq(adj) = 85.1%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	3823.0	3823.0	57.97	0.000
Residual Error	9	593.5	65.9		
Total	10	4416.5			

Obs	Tx-Rx Separation Distance [m]	Mean Path Loss [dB]	Fit	SE Fit	Residual	St Resid
1	3.0	3.01	12.48	4.58	-9.47	-1.41
2	6.1	12.03	18.37	3.95	-6.34	-0.89
3	9.1	30.14	24.27	3.37	5.87	0.80
4	12.2	43.05	30.16	2.90	12.89	1.70
5	15.2	47.02	36.06	2.57	10.96	1.42
6	18.3	36.16	41.95	2.45	-5.79	-0.75
7	21.3	40.05	47.85	2.57	-7.80	-1.01
8	24.4	59.06	53.75	2.90	5.32	0.70
9	27.4	55.88	59.64	3.37	-3.76	-0.51
10	30.5	66.06	65.54	3.95	0.52	0.07
11	33.5	69.03	71.43	4.58	-2.40	-0.36

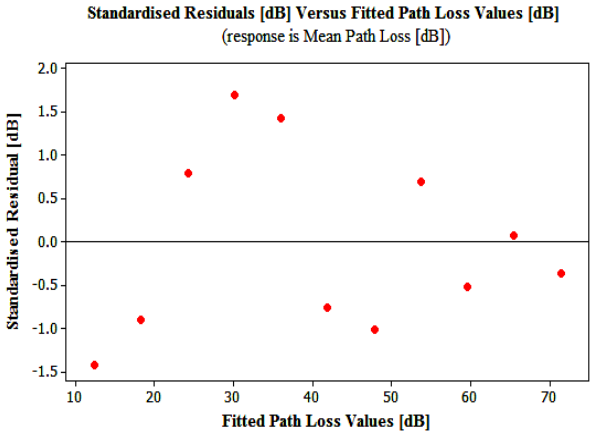

Fig. 28: Plot of standardised residuals vs. fitted PL values for model-I

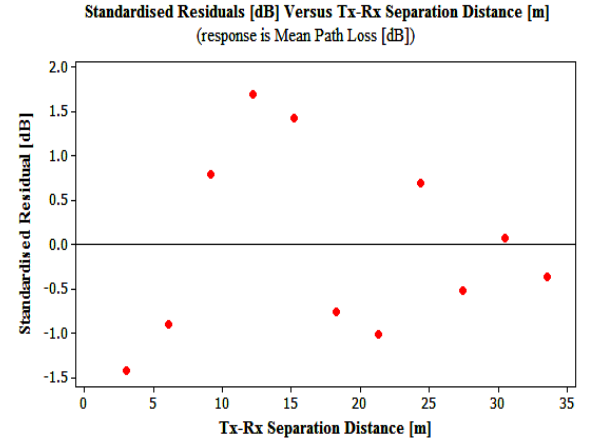

Fig. 29: Plot of standardised residuals vs. Tx-Rx separation distance for model-I

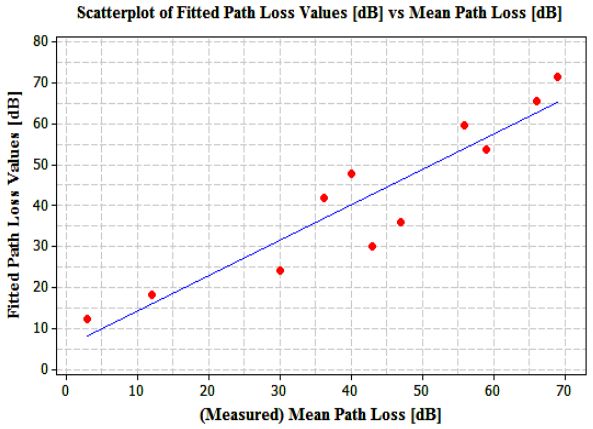

Fig. 30: Plot of fitted PL values vs. MPL for model-I

Habitually, provided a model is accurate, neither of the residual plots of Fig. 28 and Fig. 29 should exhibit distinct patterns. And, they each ought to be randomly distributed about 0 in accordance with a normal distribution, so all but very few standardised residuals should lie within 2 standard deviations of their expected value 0.

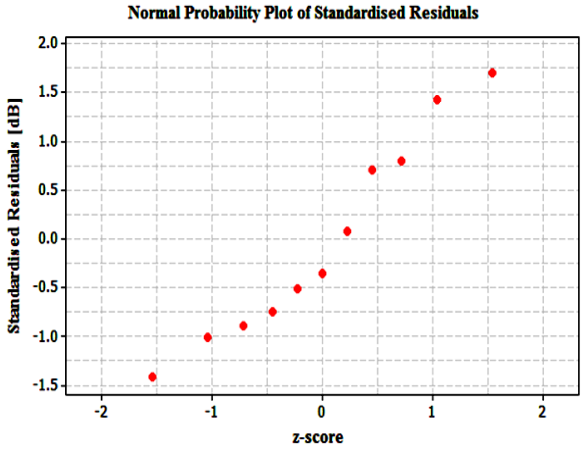

Fig.31: Normal probability plot of standardised residuals for model-I

Clearly, the plots of Fig. 28 and Fig. 29 show no distinct patterns, and the standardised residuals for either of them are well within 2 SD's of their zero-mean value, i.e., within -2 and +2. The plot of Fig. 30 yields points that are quite close to the (almost) 45° line, thereby suggesting the model effectiveness in making predictions. The straightness of the plot of Fig. 31 brings about the plausibility of the assumption that ϵ is Gaussian distributed.

Now, the next thing to consider is the amount of variability (i.e., the variance, σ^2, and thus the S.D., σ) inherent in the already-obtained regression model of eqn. (13). In other words, to get hold of an estimate of how distant the observed (Tx-Rx$_i$, PL$_i$)s are from the TRL of Fig. 27. Positive residuals in Table VI obviously correspond to (Tx-Rx$_i$, PL_i)s above the LSL, while negative ones to those lying below the line. And, note that the residuals are the differences between each observed PL_i and the corresponding fitted $\acute{P}L_i$. (I.e., residual = $PL_i - \acute{P}L_i$).

The estimate of σ^2 in regression analysis (denoted by s^2) is given by [7]:

$$\acute{\sigma}^2 = s^2 = \frac{SSE}{n-2} = \frac{\sum(y_i - \acute{y}_i)^2}{n-2} = \frac{\sum[y_i - (\acute{\beta}_0 + \acute{\beta}_1 x_i)]^2}{n-2}$$
$$= \frac{\sum {y_i}^2 - \acute{\beta}_0 \sum y_i - \acute{\beta}_1 \sum x_i y_i}{n-2}$$

Thus, the corresponding variance estimate here is given by:

$$s^2 = \frac{\sum\left(PL_i[dB] - \acute{P}L_i[dB]\right)^2}{11-2}$$
$$= \frac{\sum(PL_i[dB])^2 - \acute{\beta}_0 \sum(PL_i[dB]) - \acute{\beta}_1 \sum x_i\,(PL_i[dB])}{9}$$

where x_i = Tx-Rx separation.

Table VII lists data corresponding to the following summary quantities.

$\acute{\beta}_0 = 6.582702993$, $\acute{\beta}_1 = 1.93418072$, $\sum(PL_i[dB]) = 461.505$, $\sum(PL_i[dB])^2 = 23779.0$, $\sum x_i(PL_i[dB]) = 10416.58$

$$s^2 = \frac{23779.0 - (6.582702993)(461.505) - (1.93418072)(10416.58)}{9}$$

$$= \frac{593.50}{9} = 65.94460566$$

Table VII: Reference data index for calculating s^2 for model-I

Obs.	x_i [m]	PL_i [dB]	$(PL_i[dB])^2$	$x_i(PL_i[dB])$
1	3.048	3.0107	9.06	9.18
2	6.096	12.0307	144.74	73.34
3	9.144	30.1417	908.52	275.62
4	12.192	43.0507	1853.36	524.87
5	15.24	47.0237	2211.23	716.64
6	18.288	36.1617	1307.67	661.33
7	21.336	40.0497	1603.98	854.50
8	24.384	59.0647	3488.64	1440.23
9	27.432	55.8817	3122.76	1532.95
10	30.48	66.0587	4363.75	2013.47
11	33.528	69.0307	4765.24	2314.46

Consequently, the corresponding estimate of the standard deviation of PL for model-I is:

$$\acute{\sigma} = s = \sqrt{65.94460566} = 8.120628403 \cong 8.121$$

The next thing to consider is the coefficient of determination, given by:

$$r^2 = 1 - \frac{SSE}{SST} \tag{14}$$

$$SST = \sum(y_i - \bar{y})^2 = \sum {y_i}^2 - (\sum y_i)^2/n$$

$\sum(PL_i[dB])^2 = 23779.0$, $\sum(PL_i[dB]) = 461.505$, $n = 11$, SSE = 593.50.

Thus,

$$SST = 23779.0 - \frac{(461.505)^2}{11} = 4416.558$$

$$\therefore r^2 = 1 - \frac{593.50}{4416.558} = 1 - 0.1344 = 0.8656 \cong 0.866$$

Note that the coefficient of determination r^2 is a number between 0 and 1, i.e.:

$$0 < r^2 < 1 \tag{15}$$

It therefore follows that the SLR model chosen here for model-I is plausibly useful in explaining variation in the observed path loss data. The obtained value of r^2 connotes that roughly 86.6% of variation in the observed path loss data is attributable to the estimated linear relationship between path loss and unit Tx-Rx separation, which, obviously, is quite an imposing outcome. Table VI affirms these results.

Another useful aspect to consider is the confidence interval (CI) for β_1. Note, a 100(1 – α)% CI for the TRL is given by:

$$\acute{\beta}_1 \pm t_{\frac{\alpha}{2},n-2} \cdot s_{\acute{\beta}_1} \tag{16}$$

$$s_{\acute{\beta}_1} = \frac{s}{\sqrt{\sum x_i^2 - (\sum x_i)^2/n}}, \quad \sum x_i^2 = 4700.893824,$$

$(\sum x_i)^2 = (201.168)^2 = 40468.56422$, $s = 8.121$, $n = 11$

$$\therefore s_{\acute{\beta}_1} = \frac{8.121}{\sqrt{4700.894 - 40468.564/11}} = \frac{8.121}{31.968} = 0.254$$

With a 95% confidence level, the t critical value is $t_{\frac{0.05}{2},(11-2)} = t_{0.025,9} = 2.262$. Thus, the confidence interval is:

$$1.934 \pm (2.262 \times 0.254)$$
$$= 1.934 \pm 0.575 = (2.509, 1.359) \tag{17}$$

Accordingly, with some degree of confidence, it can be said that the average increase in path loss between 1.359 dB to 2.509 dB is associated with a 1 meter increase in Tx-Rx separation, at least with respect to the environment conditions considered here, with Tx-Rx separation roughly between 3m to 36m. The estimated CI is rather narrow, stipulating that the slope has been accurately estimated. And, the fact that the CI contains only positive values creates the confidence that path loss will generally increase with increasing Tx-Rx separation.

The subsequent test to be conducted is the *model utility test*, which brings about ascertaining that H_0 is rejected, thereby authenticating further inferences on future value predictions. This test demands H_0 be rejected for a duly small significance level, α.

Null Hypothesis: H_0: $\beta_1 = \beta_{10}$

Test statistic value: $t = \frac{\acute{\beta}_1 - \beta_{10}}{s_{\acute{\beta}_1}}$

The model utility test: test of H_0: $\beta_1 = 0$ versus H_a: $\beta_1 \neq 0$ where the test statistic value is then $t = \acute{\beta}_1/s_{\acute{\beta}_1}$.

For H_a: $\beta_1 \neq 0$, rejection region: $t \geq t_{\frac{\alpha}{2},n-2}$ or $t \leq -t_{\frac{\alpha}{2},n-2}$

Using a reasonably small significance level of α = 0.01; with (H_a: $\beta_1 \neq 0$), (H_0: $\beta_1 = 0$) will be rejected if $t = \acute{\beta}_1/s_{\acute{\beta}_1}$ satisfies either $t \geq t_{\frac{\alpha}{2},n-2} = t_{0.005,9} = 3.250$ or $t \leq -3.250$.

$$t = \frac{1.9341}{0.254} = 7.614566929 \cong 7.615$$

Evidently, $7.615 > 3.250$, so H_0 is utterly rejected. These results can also be found in the MINITAB regression output presented earlier in Table VI.

Now, since the model utility test has resulted in the rejection of H_0 for an aptly small α of 0.01, it follows that the SLR model is indeed appropriate here for further inferences concerning the estimation of the expected or true average values of path loss, or the prediction of future path loss values that will result from observations made when Tx-Rx separation falls within the range considered here (and, of course, same conditions too).

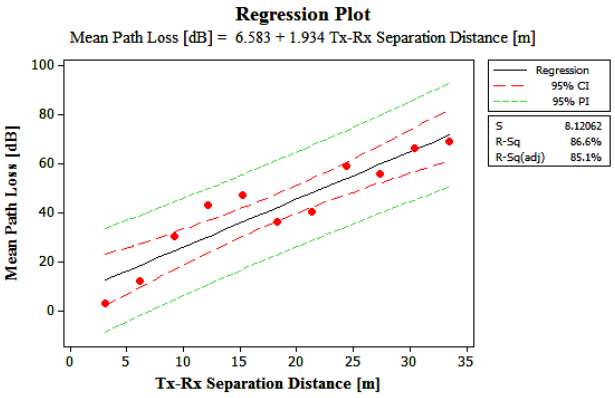

Fig. 32: Scatterplot with CIs and PIs for model-I

Table VIII: Numerical elucidation of the (95%) CIs and PIs for model-I

Predicted Values for New Observations

New Obs	Fit	SE Fit	95% CI	95% PI
1	12.48	4.58	(2.12, 22.84)	(-8.61, 33.57)
2	18.37	3.95	(9.44, 27.30)	(-2.05, 38.80)
3	24.27	3.37	(16.63, 31.90)	(4.38, 44.16)
4	30.16	2.90	(23.61, 36.72)	(10.66, 49.67)
5	36.06	2.57	(30.25, 41.87)	(16.79, 55.33)
6	41.95	2.45	(36.42, 47.49)	(22.77, 61.14)
7	47.85	2.57	(42.04, 53.66)	(28.58, 67.12)
8	53.75	2.90	(47.19, 60.30)	(34.24, 73.25)
9	59.64	3.37	(52.01, 67.28)	(39.75, 79.53)
10	65.54	3.95	(56.61, 74.47)	(45.11, 85.96)
11	71.43	4.58	(61.07, 81.79)	(50.34, 92.52)

Apparently, point estimates and predictions given by eqn. (13) single-handedly give no precise information as to how path loss has been predicted. This can be offset by means of formulating a CI for the least squares fit line, and a PI for each predicted path loss value. Fig. 32 (and Table VIII) depicts the CI and PI for model-I. Note that the CI obtained earlier (eqn. (17)) was for the slope coefficient, β_1, alone, and was based upon the properties of the distribution of $\acute{\beta}_1$. The CI here is based upon properties of the distribution of $\acute{\beta}_0 + \acute{\beta}_1 x_i$, and is obtained using the following formulation (for each CI):

$$\acute{\beta}_0 + \acute{\beta}_1 x_i \pm t_{\frac{\alpha}{2},n-2} \cdot s_{\acute{\beta}_0+\acute{\beta}_1 x_i} \tag{18}$$

where $s_{\acute{\beta}_0+\acute{\beta}_1 x_i} = s\sqrt{\frac{1}{n} + \frac{(x_i-\bar{x})^2}{\sum x_i^2 - (\sum x_i)^2/n}}$, and $(\acute{\beta}_0 + \acute{\beta}_1 x_i)$ has a normal distribution.

Whereas, the PI for future path loss observations (made under similar settings and distance range) is obtained using the formulation:

$$\acute{\beta}_0 + \acute{\beta}_1 x_i \pm t_{\frac{\alpha}{2},n-2} \cdot s\sqrt{1 + \frac{1}{n} + \frac{(x_i-\bar{x})^2}{\sum x_i^2 - (\sum x_i)^2/n}} \tag{19}$$

The '1' to the left of $\left(\frac{1}{n}\right)$ in the square root for the PI formulation makes the PI wider than the CI, although they are clearly both centred at $\acute{\beta}_0 + \acute{\beta}_1 x_i$. Table VIII has presented both the CIs and PIs for each Tx-Rx separation, along with Fig. 32 clearly providing a graphical depiction of the behaviour of intervals. It can be seen that the PI's width remains fixed, which results from the fact that even with perfect information on the values of β_0 and β_1, some ambiguity will still be present in prediction.

Now that an empirical path loss equation has been obtained (and proven suitable), the next step is to obtain the slow fading loss, which will subsequently be added to the already obtained empirical path loss, to then attain the total path loss obtainable here.

As earlier emphasized, the slow fading loss is a continuous r.v., with a CDF that describes the probability that the slow fading loss is lesser than or equal to some stipulated value. Also recall that probability theory affirms that absolute knowledge of a single r.v. is made available through knowledge of its PDF.

Let PL_T = Total (unknown) path loss within the given environment.

$$\therefore PL_T[\text{dB}] = L_E[\text{dB}] + L_s[\text{dB}] \tag{20}$$

L_E[dB] = Empirical path loss [in decibels]
L_s[dB] = Path loss resulting from slow fading and shadowing

$f_{PL_s}(l_s)$, the PDF for the r.v. L_s, is given by:

$$P(|L_s - l_s| < \Delta l_s) = f_{PL_s}(l_s)\Delta l_s \tag{21}$$

Note from Table III that the *least* mean path loss observed for model-I is 3.0107dB (Tx-Rx separation = 3.048m), and the largest is 69.0307dB (at Tx-Rx separation = 33.528m). Accordingly, L_s is an r.v. corresponding to the probability that the slow fading path loss is between 3.0107dB and 69.0307dB.

One useful point to note is that the already obtained L_E has a mean value of 0 dB, and so to have L_s suitably added to L_E, to then obtain L_T, L_s ought to also have a mean of 0 dB. This can be achieved by means of standardising L_s.

Note that if a non-standard r.v. X has a normal distribution with mean μ and S.D. σ, then

$$Z = \frac{X-\mu}{\sigma} \tag{22}$$

has a standard Gaussian normal distribution. With L_s denoting the slow fading path loss here, standardising L_s results in the following:

$$3.0107 \leq L_s \leq 69.0307$$

iff,

$$\frac{3.0107 - 41.9550}{8.121} \leq \frac{L_s - 41.9550}{8.121} \leq \frac{69.0307 - 41.9550}{8.121}$$

$$\begin{aligned}\therefore P(3.0107 \leq L_s \leq 69.0307) &= P\left(\tfrac{3.0107-41.9550}{8.121} \leq Z \leq \tfrac{69.0307-41.9550}{8.121}\right) \\ &= P(-4.80 \leq Z \leq 3.33) = \Phi(3.33) - \Phi(-4.80) \\ &= 0.9996 - 0.00000029 = 0.99959971 \cong 0.9996\end{aligned}$$

Thus, the stipulated value l_s for the slow fading loss L_s, a loss understood to be within a small range Δl_s, is approximately the 99.96th percentile of the normal distribution with μ = 41.9550 and σ = 8.121. A little illustration of this is given in Fig. 33.

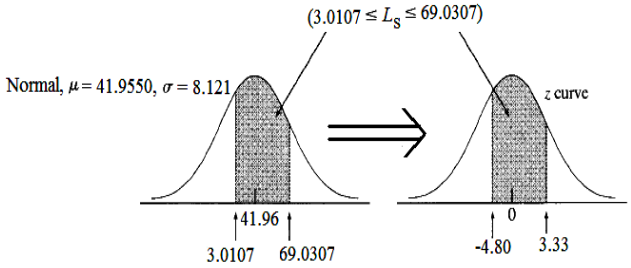

Fig. 33: Transforming from non-zero- to zero-mean for MPL distribution of model-I

In Fig. 33, note that the shaded areas under the curves are the same, except that the Gaussian normal r.v., L_s, has had its original mean value shifted to zero, through the standardising procedure elucidated earlier. Using a table of the *standard normal curve areas*, it can be shown that the 99.96th percentile of the standard Gaussian normal distribution is 3.38.

Thus, the stipulated value, l_s, which is deemed greater than or equal to the slow fading loss L_s (i.e., $L_s \leq l_s$), is obtained as follows:

$$l_s = \eta(0.9960) = \mu + Z\sigma = 41.9550 + (3.38)(8.121) = 69.40398 \cong 69.4\ dB$$

Thus, the slow fading loss (in dB), modelled here as a (zero-mean) PDF Gaussian r.v., is:

$$f_{L_s}(l_s) = \frac{1}{\sqrt{2\pi}\cdot\sigma_{L_s}} e^{-\frac{l_s^2}{2\sigma_{L_s}^2}} = \frac{1}{\sqrt{2\pi}\cdot 8.121} e^{-\left(\frac{(69.4)^2}{2(65.95)}\right)}$$

$$= 6.8067 \times 10^{-18}\ dB$$

Thus the total path loss obtainable (i.e., model-I) is given by:

$$PL_T = 6.583 + 1.934 \text{ Tx-Rx-separation} + 6.8067 \times 10^{-18} \quad [\text{dB}] \qquad (23)$$

The total loss for the second environment (i.e., model-II) has been obtained using similar steps to that of the preceding model-I, so the derivative steps for model-II will not be given here. The corresponding equation for model-II is as follows:

$$PL_T = 17.62 + 1.195 \text{ Tx-Rx-separation} + 1.656 \times 10^{-21} \quad [\text{dB}] \qquad (24)$$

Note that the slow fading loss of 6.8067×10^{-18} dB implies a very minimal loss. I.e., almost entirely no effect from objects present during measurement. The powers -18 and -21, respectively, clearly suggest that the slow fading loss for either of the environment is almost 0dB.

Moreover, it's rather conventional representing path loss as some distance power γ (i.e., the path loss exponent, also referred to as the "signal decay index") along with the Tx-Rx separation distance (which will be replaced here with *d* for easy denotation). The path loss exponent is essentially useful as it reveals how fast the signal strength, for the given environment, deteriorates with increasing distance (*d*). Larger γ's simply imply more rapidly decaying signal strength (i.e., a quicker increasing path loss) with distance, *d*.

The mean path loss for some *d* in conjunction with γ is given by [1]:

$$\overline{PL}(d)\ [dB] = \overline{PL}(d_0)[dB] + 10\gamma log_{10}\left(\frac{d}{d_0}\right) \qquad (25)$$

where d_0 is the 'close-in' 'reference' distance determined via measurements close to the transmitter [1]. Now, while a reference distance of 1m or 100m is prescribed in [1] as suitable reference distances for short-ranged systems, an obvious reference distance of 10ft (3.048m) has been used for model-I, which corresponds to the mean path loss $\overline{PL}(d_0)$ = 6.583dB. Recall that for model-II the reference distance is also 10ft. Apparently, the 3.048m distance is obviously in the far field of the antennas, so effects resulting from the antennas being in the near-field region are out of the question. Rearranging eqn. (4.14) yields:

$$\overline{PL}(d)\ [dB] = 10\gamma log_{10}\left(\frac{d}{d_0}\right) + \overline{PL}(d_0)[dB] \qquad (26)$$

which literally has the form of the generalised linear equation:

$$y = mx + c \qquad (27)$$

wherein *m* (equal to γ here) is the slope of the line and *c* is the y-axis-intercept ($\overline{PL}(d_0)$).

Clearly, the slope *m* corresponds to the path loss exponent, and so it follows that environments of the sort considered for models I and II have approximate path loss exponents equal to 1.934 and 1.195, respectively. Model-I reveals that the corresponding environment has path loss exponent almost equal to that of free space (having γ = 2). Model-II reveals much slower declining path loss as distance between transmitter and receiver is increased. In general, results show that the obstructions in the environment considered here had little effect (not much) on radio signals. This may well be as a result of the way in which the measurements have been taken; precisely, during the measurements, the transmit and receive antennas were, more often than not, in direct sight.

Moving on to deriving model-III; looking at the scatterplots of Fig. 23 to Fig. 26, the data points appear rather scattered all over the place (i.e., the association of data isn't exactly linear). However, by visualising the first two data points (w.r.t. the abscissa) as being off the plots, the association of data then seems fairly linear. These first two points clearly depict high path loss despite the shorter Tx-Rx separation, i.e., in comparison with the other observations with longer Tx-Rx separations. The remedy applied here is the utilisation of the multiple linear regression methodology, with two fundamental predictors, wherein Tx-Rx separation is one of the predictors, while Tx-Building separation is the other. The Tx-Building-separation predictor is incorporated here because, clearly, the closeness of a transmit point to the building surely has some effect on the received signal strength, and thus, the path loss. Referring to depictions provided earlier in Fig. 20 and Fig. 21, it can be seen that propagations closer to the building have to propagate through the multi-level building floors, and the nearer the transmit point is to the building the more floors the propagated signal have to penetrate through. This clearly affirms the height at which floors of a multi-level building does have on radio signals. Also, by referring to Fig. 20 and Fig. 21, it can be seen that some radio signals propagated further away from the building had to propagate through tree foliage. This is believed to be an additional reason for the randomness of the data association of Fig. 23 to Fig. 26. That is, depending on the separation distance, data points indicating high path loss ought to have resulted due to propagation through foliage (i.e., at the latter Tx-Building separations).

Table IX: MINITAB regression output for model-III

The regression equation is
Mean Path Loss [dB] = 687 + 20.4 Tx-Rx Separation [m]
- 17.9 Tx-Building Separation [m]
- 6.8 Tx-Rx Separation Squared
+ 5.0 Tx-Building Separation Squared
+ 1.83 Tx-Rx Separation times Tx-Building Building

Predictor	Coef	SE Coef	T	P
Constant	687	1918	0.36	0.731
Tx-Rx Separation [m]	20.37	27.73	0.73	0.486
Tx-Building Separation [m]	-17.92	25.77	-0.70	0.509
Tx-Rx Separation Squared	-6.85	15.64	-0.44	0.675
Tx-Building Separation Squared	5.01	16.25	0.31	0.767
Tx-Rx times Tx-Building	1.827	4.730	0.39	0.711

S = 5.25946 R-Sq = 67.9% R-Sq(adj) = 44.9%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	5	409.23	81.85	2.96	0.095
Residual Error	7	193.63	27.66		
Total	12	602.87			

Source	DF	Seq SS
Tx-Rx Separation [m]	1	160.88
Tx-Building Separation [m]	1	138.31
Tx-Rx Separation Squared	1	100.82
Tx-Building Separation Squared	1	5.09
Tx-Rx times Tx-Building	1	4.13

Obs	Tx-Rx Separation [m]	Mean Path Loss [dB]	Fit	SE Fit	Residual	St Resid
1	12.6	79.93	79.93	5.24	0.00	0.00 X
2	16.4	69.98	69.41	4.60	0.57	0.22
3	21.3	61.00	64.95	3.05	-3.95	-0.92
4	26.7	73.13	64.69	3.13	8.43	1.99
5	32.4	60.87	67.45	2.88	-6.59	-1.50
6	38.2	70.97	70.69	2.62	0.28	0.06
7	44.1	78.10	73.68	2.34	4.42	0.94
8	50.0	69.95	75.85	2.76	-5.90	-1.32
9	55.9	81.09	78.76	2.78	2.33	0.52
10	61.9	79.90	80.24	3.04	-0.33	-0.08
11	67.9	79.26	78.48	4.62	0.78	0.31
12	74.0	81.09	79.96	3.56	1.12	0.29
13	80.0	76.01	77.17	4.36	-1.17	-0.40

X denotes an observation whose X value gives it large leverage.

Using Tx-Rx-separation and Tx-Building-separation as the two fundamental predictors, the MINITAB software package has been fully utilised here to carry out the regression analysis. And, after a number of diagnostic fitting of several possible multiple regression functions to the data, the *complete second order* (*full quadratic*) *model* was found to produce the best fit to the data, i.e., with the largest *coefficient of multiple determination*. Table IX presents the regression output referred to here.

From Table IX, clearly, the least squares equation for model III is:

$$\overline{PL}[dB] = 687 + 20.37d_1 - 17.92d_2 - 6.85d_1^2 + 5.01d_2^2 + 1.827d_1d_2 \quad (28)$$

where d_1 = Tx-Rx Separation Distance [in meters]
d_2 = Tx-Building Separation Distance [in meters]
Evidently, $\acute{\beta}_0$ = 687, $\acute{\beta}_1$ = 20.37, $\acute{\beta}_2$ = -17.92, $\acute{\beta}_3$ = -6.85, $\acute{\beta}_4$ = 5.01, and $\acute{\beta}_5$ = 1.827.

Notice that three additional predictors have been used to complement the two fundamental predictors, d_1 and d_2. These additional predictors are obviously mathematical functions of d_1 and/or d_2. Eqn. (28) will usually imply that 20.37dB is the average change in path loss associated with a 1-meter increase in Tx-Rx separation, while values of the other predictors (i.e., d_2, d_1^2, d_2^2, and d_1d_2) are held fixed. However, it is obviously not feasible to increase the value of Tx-Rx separation, while, say, Tx-Building separation is held fixed. I.e., every increase in either of these predictor variables perceptibly results in the other's concurrent increase. Accordingly, a suitable alternative interpretation of the estimated regression coefficients will be used here.

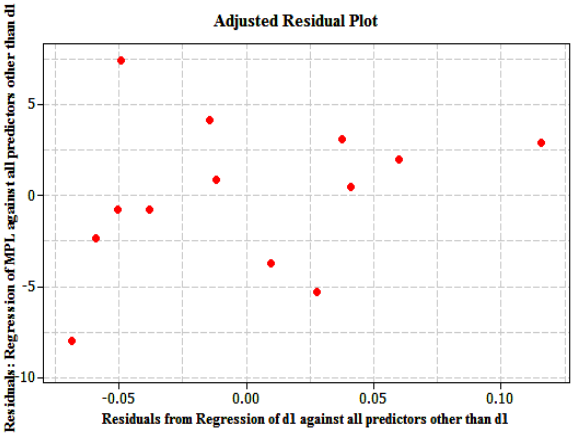

Fig. 34: *Adjusted residual plot*: [residuals from regression of mean-path-loss against d_2, d_1^2, d_2^2, and d_1d_2] vs [residuals from regression of d_1 against d_2, d_1^2, d_2^2, and d_1d_2]

The adjusted residual plot of Fig. 34 is given here in line with trying to obtain a suitable interpretation for the slope $\acute{\beta}_1$ of model III. The residuals from the regression of mean-path-loss (MPL) against d_2, d_1^2, d_2^2, and d_1d_2 represent variation in MPL after adjusting for or removing the effects of d_2, d_1^2, d_2^2, and d_1d_2. Same interpretation applies for the residuals from regression of Tx-Rx-separation (i.e., d_1) against d_2, d_1^2, d_2^2, and d_1d_2, wherein the residuals represent variation in Tx-Rx-separation after the adjusting for or removing the effects of d_2, d_1^2, d_2^2, and d_1d_2. Fig. 35 provides a plot wherein a regression fit is superimposed on Fig. 34.

The slope of the regression fit line of Fig. 35 is actually exactly $\acute{\beta}_1$, [7], i.e., the residuals from this fit line are precisely the residuals obtainable from directly applying multiple regression of MPL against d_1, d_2, d_1^2, d_2^2, and d_1d_2. Table X provides the regression output on this. It follows that $\acute{\beta}_1$ can be interpreted as the estimated changes in MPL associated with a 1-meter increase in d_1 after adjusting for or removing the effects of any of the other predictor variables d_2, d_1^2, d_2^2, and d_1d_2. The point here is that given that the slope, $\acute{\beta}_1$, of model III is exactly same as that of the corresponding adjusted residual plot, then slope $\acute{\beta}_1$ of model-III can be aptly interpreted as the estimated changes in MPL associated with a 1-meter increase in MPL after adjusting for the effects of d_2,

d_1^2, d_2^2, and $d_1 d_2$; rather than interpreting as after d_2, d_1^2, d_2^2, and $d_1 d_2$ are held fixed.

The next obvious course of action would be the model utility test. Unlike in the cases of model-I and model-II; with multivariate data, there generally isn't any initial depiction allied to a scatterplot to point out if a particular multiple regression model will be judged useful. R^2, the *coefficient of multiple regression*, tells quite a lot in terms of preliminary assessments.

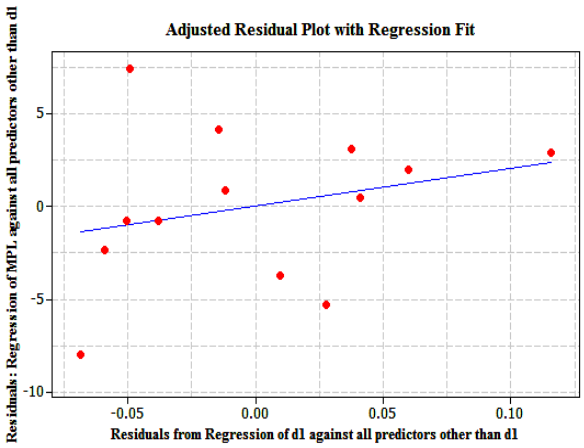

Fig. 35: Regression fit line superimposed on Fig. 34

Table X: Regression output for the regression fitting of Fig. 35

The regression equation is
[Residuals from Regression of MPL against d_2, d_1^2, d_2^2, and $d_1 d_2$] = - 0.001 + 20.35 [Residuals from Regression of d_1 against d_2, d_1^2, d_2^2, and $d_1 d_2$]

Predictor	Coef	SE Coef	T	P
Constant	-0.001	1.163	-0.00	0.999
Tx-Rx regs vs other predictors	20.35	22.12	0.92	0.377

S = 4.19485 R-Sq = 7.1% R-Sq(adj) = 0.0%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	14.89	14.89	0.85	0.377
Residual Error	11	193.56	17.60		
Total	12	208.45			

Obs	Residuals from Regression of d_1 against d_2,d_1^2,d_2^2, and $d_1 d_2$	Residuals from Regression of MPL against d_2,d_1^2,d_2^2,and $d_1 d_2$	Fit	SE Fit	Residual	St Resid
1	-0.038	-0.78	-0.78	1.44	-0.00	-0.00
2	0.116	2.93	2.36	2.82	0.57	0.18
3	0.010	-3.75	0.19	1.18	-3.94	-0.98
4	-0.049	7.43	-1.00	1.59	8.43	2.17R
5	-0.069	-7.99	-1.40	1.91	-6.59	-1.77
6	-0.051	-0.75	-1.03	1.61	0.28	0.07
7	-0.015	4.12	-0.30	1.21	4.42	1.10
8	0.028	-5.33	0.56	1.32	-5.89	-1.48
9	0.037	3.09	0.76	1.43	2.33	0.59
10	0.041	0.50	0.83	1.47	-0.33	-0.08
11	0.060	2.00	1.22	1.77	0.78	0.20
12	-0.012	0.89	-0.24	1.19	1.13	0.28
13	-0.059	-2.37	-1.20	1.75	-1.17	-0.31

R denotes an observation with a large standardized residual.

Quite a number of trials were made to ascertain the aptness of the model-form chosen here for model-III (i.e., eqn. (28)). Table XI presents the regression output of the equation that would result if only the two fundamental predictors (i.e., only Tx-Rx separation and Tx-Building separation) had been used. R^2 is obviously lower, in Table XI, which is a useful first step in seeing that the model of eqn. (28), indeed, represents a better fit to the data. Also, the RMSE, s, is larger with only d_1 and d_2 as predictors.

Table XI: Regression output if only predictors d_1 and d_2 had been used for model-III

The regression equation is
Mean Path Loss [dB] = 49.1 + 4.03 Tx-Rx Separation [m]
- 3.65 Tx-Building Separation [m]

Predictor	Coef	SE Coef	T	P
Constant	49.089	8.939	5.49	0.000
Tx-Rx Separation [m]	4.034	1.815	2.22	0.050
Tx-Building Separation [m]	-3.645	1.708	-2.13	0.059

S = 5.51063 R-Sq = 49.6% R-Sq(adj) = 39.6%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	299.19	149.60	4.93	0.032
Residual Error	10	303.67	30.37		
Total	12	602.87			

Source	DF	Seq SS
Tx-Rx Separation [m]	1	160.88
Tx-Building Separation [m]	1	138.31

Obs	Tx-Rx Separation [m]	Mean Path Loss [dB]	Fit	SE Fit	Residual	St Resid
1	12.6	79.93	77.50	4.97	2.43	1.02 X
2	16.4	69.98	70.80	2.63	-0.81	-0.17
3	21.3	61.00	68.45	2.39	-7.44	-1.50
4	26.7	73.13	68.06	2.42	5.07	1.02
5	32.4	60.87	68.64	2.34	-7.78	-1.56
6	38.2	70.97	69.78	2.16	1.19	0.23
7	44.1	78.10	71.25	1.95	6.85	1.33
8	50.0	69.95	72.94	1.80	-2.99	-0.57
9	55.9	81.09	74.76	1.79	6.32	1.21
10	61.9	79.90	76.70	1.96	3.21	0.62
11	67.9	79.26	78.72	2.30	0.54	0.11
12	74.0	81.09	80.78	2.75	0.31	0.06
13	80.0	76.01	82.90	3.30	-6.90	-1.56

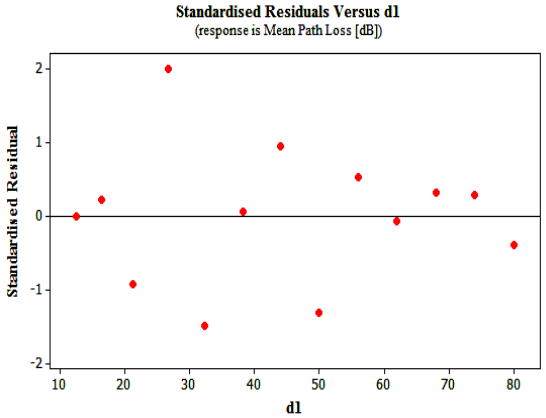

Fig. 36: Plot of Standardised residuals vs Tx-Rx separation for model-III

Now, before going on with carrying out hypotheses tests, constructing CIs, and making predictions, one ought to first inspect diagnostic plots to see if the model needs modification or whether outliers do exist in the data. The standardised residuals in multiple regression are a result of dividing each residual by its (estimated) S.D.; this isn't as straightforward as it was for SLR. The standardised residuals' normal probability plot is suitably useful in validating the normality assumption. Also, plots of the standardised residuals versus each of the predictors as well as versus fitted (i.e., predicted) path loss values ought to show no discernable pattern. The residuals ought to be randomly distributed about 0 (w.r.t. a normal distribution), therefore all but only a very few standardised residuals should lie within the region of -2 and +2, i.e., all and maybe a few residuals ought to be within 2 S.D.'s of their expected value 0. Looking at Fig. 36 to Fig. 41, they clearly all lie within the -2 to +2 range. The plot of standardised residuals vs fitted path loss values of Fig. 41 is actually a combination of the preceding residual plots, i.e., against each predictor variable, showing implicitly how residuals vary with the predictors. The plot of standardised residuals vs fitted values is apparently the single most often suggested for multiple regression analysis, and it is clearly evenly distributed about 0, with all points falling within the -2 to +2 standardised residual range. Fig. 42 portrays how the fitted values measure up with observed path loss values. As emphasised during the discussions of model-I and -II, a *normal* probability plot reveals the plausibility of the assumption that the random deviation, ε, has a normal distribution. Fig. 43 presents the normal probability plot for standardised residuals corresponding to model-III. The straightness (approximately) of the points in the plot of Fig. 43, proffers little doubt as to whether ε is normally distributed.

A 100(1 – α)% CI for β_i is given by [7]:

$$\hat{\beta}_i \pm t_{\alpha/2,n-(k+1)} \cdot s_{\hat{\beta}_i}$$

A 99% CI for β_1, the change in MPL associated with a 1-meter increase in Tx-Rx separation while the other predictors are adjusted for, requires $t_{0.005,13-(5+1)} = t_{0.005,7} = 3.499$.

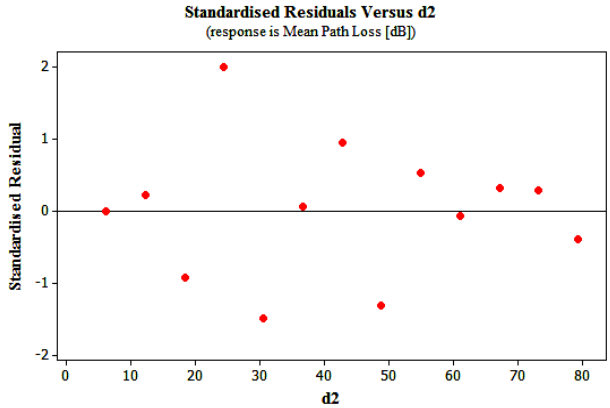

Fig. 37: Plot of standardised residuals vs Tx-Building separation for model-III

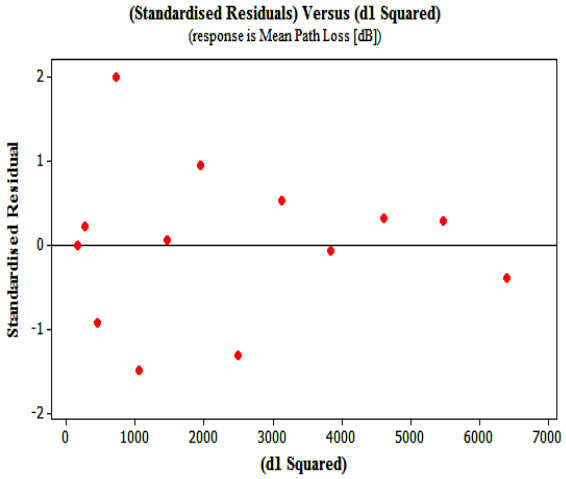

Fig. 38: Plot of Standardised residuals vs Tx-Rx separation squared, for model-III

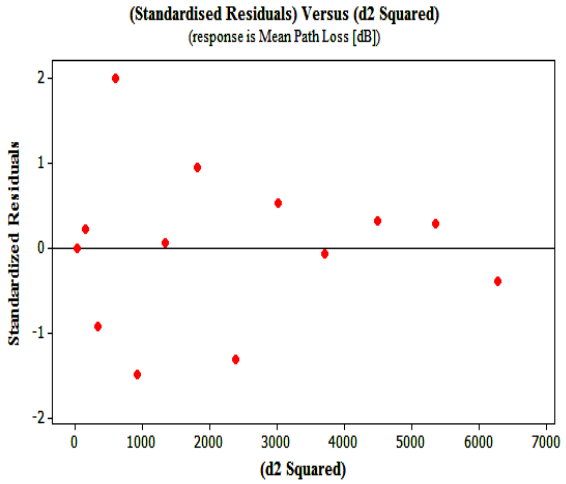

Fig. 39: Plot of standardised residuals vs Tx-Building separation squared, for model-III

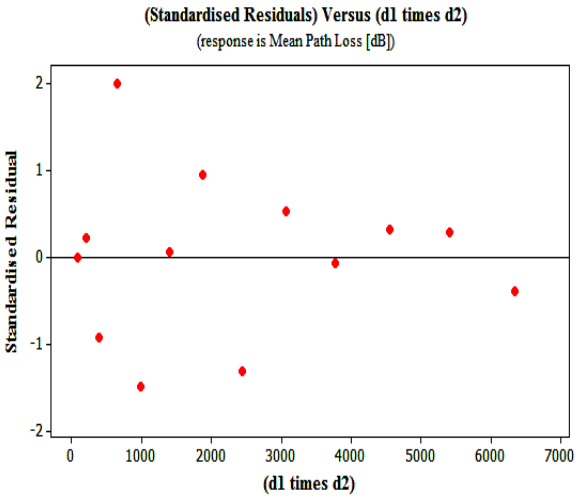

Fig. 40: Plot of Standardised residuals vs (Tx-Rx times Tx-Building separation), model-III

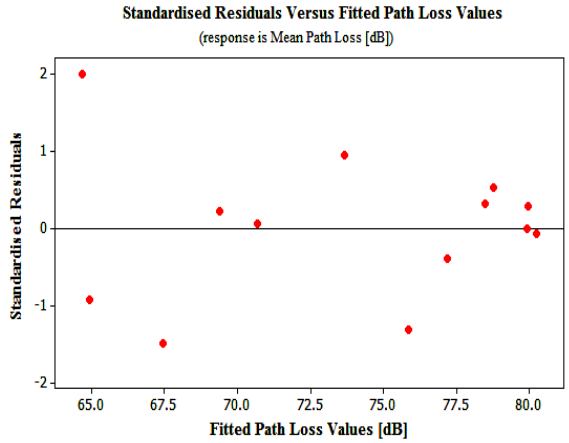

Fig. 41: Plot of Standardised residuals vs fitted path loss values, for model-III

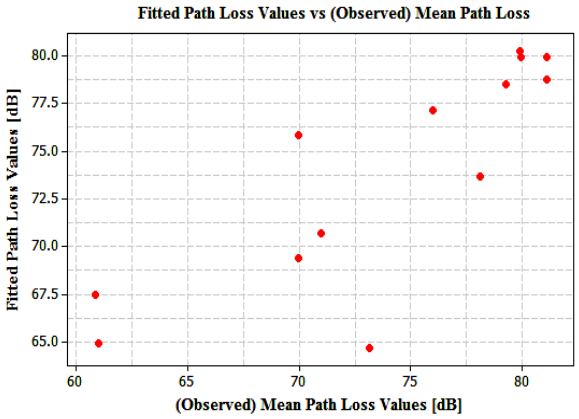

Fig. 42: Plot of fitted path loss values vs observed MPL values, for model-III

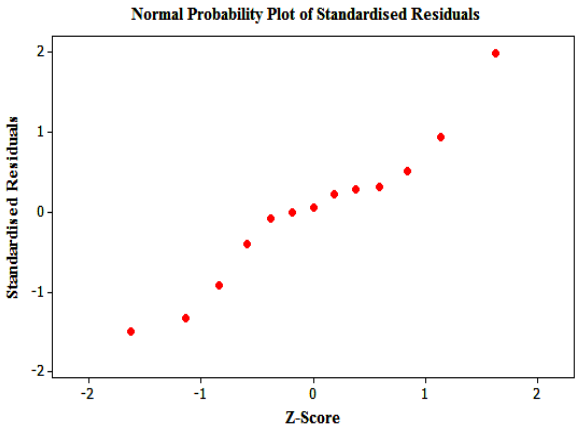

Fig. 43: Normal probability plot of standardised residuals, for model-III

With reference to Table IX for estimated values of $\acute{\beta}_i$ and $s_{\acute{\beta}_i}$, the CI for β_1 is:

$$20.37 \pm (3.499)(27.73) = 20.37 \pm 97.027$$
$$\cong (-76.657, 117.397)$$

For β_2, a 99% CI is:

$$-17.92 \pm (3.499)(25.77) = -17.92 \pm 90.169$$
$$\cong (-108.089, 72.249)$$

For β_3, 99% CI is:

$$-6.85 \pm (3.499)(15.64) = -6.85 \pm 54.724$$
$$\cong (-61.574, 47.874)$$

For β_4, 99% CI is:

$$5.01 \pm (3.499)(16.25) = 5.01 \pm 56.859$$
$$\cong (-51.849, 61.869)$$

And, for β_5, a 99% CI is:

$$1.827 \pm (3.499)(4.730) = 1.827 \pm 16.550$$
$$\cong (-14.723, 18.377)$$

For the expected path loss values of the various separation distances, CIs and PIs have been obtained using MINITAB (see Table XII). Analytically, say $d_1 = 16.4$, $d_2 = 12.2$, $d_1^2 = 269$, $d_2^2 = 149$, and $d_1 d_2 = 200$, then a 95% CI for this case would be:

$$\acute{PL}_2 \pm t_{\alpha/2, n-(k+1)} \cdot [estimated\ SD\ for\ \acute{PL}_2]$$
$$69.41 \pm t_{0.025,7} \cdot 4.60 = 69.41 \pm (2.365)(4.60)$$
$$= 69.41 \pm 10.879 = (58.53, 80.29)$$

And, the 95% PI:

$$\acute{PL}_2 \pm t_{\alpha/2, n-(k+1)} \cdot \sqrt{s^2 + [estimated\ SD\ for\ \acute{PL}_2]^2}$$
$$= 69.41 \pm t_{0.025,7} \cdot \sqrt{(5.25946)^2 + (4.60)^2}$$
$$= 69.41 \pm 16.525 = (52.89, 85.94)$$

Table XII: Numerical elucidation of the (95%) CIs and PIs for model-III

Predicted Values for New Observations

New Obs	Fit	SE Fit	95% CI	95% PI
1	79.93	5.24	(67.53, 92.33)	(62.36, 97.49)X
2	69.41	4.60	(58.54, 80.28)	(52.90, 85.93)
3	64.95	3.05	(57.74, 72.16)	(50.57, 79.32)
4	64.69	3.13	(57.30, 72.09)	(50.23, 79.16)
5	67.45	2.88	(60.65, 74.26)	(53.28, 81.63)
6	70.69	2.62	(64.49, 76.90)	(56.79, 84.59)
7	73.68	2.34	(68.14, 79.22)	(60.07, 87.29)
8	75.85	2.76	(69.32, 82.37)	(61.80, 89.89)
9	78.76	2.78	(72.19, 85.32)	(64.69, 92.82)
10	80.24	3.04	(73.05, 87.43)	(65.87, 94.60)
11	78.48	4.62	(67.55, 89.42)	(61.92, 95.04)
12	79.96	3.56	(71.56, 88.37)	(64.95, 94.98)
13	77.17	4.36	(66.86, 87.48)	(61.02, 93.33)

Values of Predictors for New Observations

New Obs	Tx-Rx Separation [m]	Tx-Building Separation [m]	Tx-Rx Separation Squared	Tx-Building Squared	Tx-Rx times Tx-Building
1	12.6	6.1	158	37	77
2	16.4	12.2	269	149	200
3	21.3	18.3	455	334	390
4	26.7	24.4	715	595	652
5	32.4	30.5	1049	929	987
6	38.2	36.6	1458	1338	1397
7	44.1	42.7	1941	1821	1880
8	50.0	48.8	2499	2378	2438
9	55.9	54.9	3130	3010	3070
10	61.9	61.0	3836	3716	3776
11	67.9	67.1	4617	4497	4556
12	74.0	73.2	5471	5351	5411
13	80.0	79.2	6401	6280	6340

With multiple regression analysis, the model utility test is based on a statistic that has a specified F distribution when H_0 is true, where [7]:

Null Hypothesis: $H_0: \beta_1 = \beta_2 = \cdots = \beta_k = 0$

Alternative hypothesis: H_a: at least one $\beta_i = 0$ $(i = 1, \cdots, k)$

Test statistic value: $f = \frac{R^2/k}{(1-R^2)/[n-(k+1)]} = \frac{SSR/k}{SSR/[n-(k+1)]} = \frac{MSR}{MSE}$

SSR = SST – SSE

Rejection region for a level α test: $f \geq F_{\alpha,k,n-(k+1)}$

Now, based on model-III of eqn. (28), with essentially five predictors, the relevant hypotheses are:

H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$

H_a: at least one of these five βs is not 0

Table IX presented the (MINITAB) regression output for model-III. From this table, the values of R^2, adjusted R^2, and s (Root Mean Square) have been found from the best fit to the data, and they all suggest a useful model.

The model utility F ratio:

$$f = \frac{R^2/k}{(1-R^2)/[n-(k+1)]} = \frac{0.679/5}{0.321/(13-6)} = \frac{0.1358}{0.045857142} = 2.961$$

This value also appears in Table IX (MINITAB regression output for model-III), precisely in the F ratio column of the ANOVA table. Now, from a *Table of Critical Values for F Distributions*, at $\alpha = 0.100$, $f \geq F_{\alpha,k,n-(k+1)}$. I.e., $F_{0.100,5,7} = 2.88 \leq f$.

So, the F critical value applicable here, for numerator-df = 5 and denominator-df = 7, is 2.88, which captures an upper-tail area of 0.100. Thus P-value < 0.100. The ANOVA table in the MINITAB output of Table IX shows that P-value = 0.095, which is < 0.100. Thus, it can be concluded that there is a useful linear relationship between *path loss* and one of the five predictor variables in model-III. It should be, particularly, noted that this does not imply that all five predictors are useful; results only suggest that H_0 is successfully rejected, and that *at least* one of the five predictors is linearly associated to the response (i.e., linearly associated to the *path loss* predicted by model-III).

Referring to Table V, the *least* mean path loss observed for model-III is 60.8657dB (at Tx-Building separation = 30.480m), and the *largest* is 81.0877dB (at Tx-Building separation = 54.864m and 73.152m). Accordingly, the slow fading loss L_s, estimated here as a Gaussian PDF function, is an r.v. approximately between 60.8657dB and 81.0877dB. Apparently, the mean of L_s ought to be transformed to zero to allow for an apt addition to the zero-mean empirical path loss. From Table IX, it can be found that the *standard deviation* of *mean path loss* here is 5.25946. The *mean*, as usual, is obtained by taking the average of the thirteen *mean path loss* observations made for model-III, which turns out to be equal to 73.9436230769231 ≅ 73.9436 dB.

Standardising to obtain zero-mean slow fading loss here, results in:

$$60.8657 \leq L_s \leq 81.0877$$

if and only if,

$$\frac{60.8657 - 73.9436}{5.259} \leq \frac{L_s - 73.9436}{5.259} \leq \frac{81.0877 - 73.9436}{5.259}$$

$$\therefore P(60.8657 \leq L_s \leq 81.0877)$$

$$= P\left(\frac{60.8657-73.9436}{5.259} \leq Z \leq \frac{81.0877-73.9436}{5.259}\right)$$

$$= P(-2.487 \leq Z \leq 1.358) = \Phi(1.36) - \Phi(-2.49)$$
$$= 0.9131 - 0.0064 = 0.9067$$

Thus, the stipulated value, l_s, is approximately the 90.67th percentile of the Gaussian distribution, with μ = 73.9436 and σ = 5.259. Using a table of the *standard normal curve areas*, it can be shown that the 90.67th percentile of the standard Gaussian normal distribution is approximately 1.32.

$$\therefore l_s = \eta(0.9067) = \mu + Z\sigma$$
$$= 73.9436 + (1.32)(5.259) = 80.885 \text{ dB}$$

Consequently, the slow fading loss (in dB), modelled here as a PDF Gaussian r.v., is:

$$f_{L_s}(l_s) = \frac{1}{\sqrt{2\pi}\cdot\sigma_{L_s}} e^{-\frac{l_s^2}{2\sigma_{L_s}^2}} = \frac{1}{\sqrt{2\pi}\cdot 5.259} e^{-\left(\frac{(80.885)^2}{2(27.657)}\right)}$$
$$= 3.258 \times 10^{-53}\ dB$$

Thus, the total loss/equation for model-III is:

$$L_T = 687 + 20.37d_1 - 17.92d_2 - 6.85d_1^2 + 5.01d_2^2 + 1.827d_1d_2 + 3.258 \times 10^{-53} \text{ [dB]}$$

where d_1 = Tx-Rx separation distance

d_2 = Tx-Building separation distance

V. Conclusions

Fundamentally, the essence of any model would be to proffer valuable information with reference to the relationship between the response (dependent) variable and the predictor (independent) variable(s). If the decision is to obtain a model empirically, then all the modeller essentially has to do is determine the right number of predictor variables applicable to the given context, upon which the dependent r.v.(s) will be exemplified. Of course, the dependent variable in propagation modelling will always be path loss (or received signal strength), depending on the modeller preference.

An empirically derived propagation model will, at the end of the day, comprise constant values (i.e., the y-intercept and regression slope coefficient(s)) and the predictor variable(s), proportionate to the specified environment. Utmost care certainly ought to be taken to obtain the right number of predictors. Though, more often than not, empirically derived propagation models will only encompass Tx-Rx separation distance as the predictor variable. However, as with model-III of this paper, the more predictors a model has, the more information the model bids.

It ought to be comprehended that every path loss model (not just those proposed here) is subject to inconsistency, i.e., not 100% accurate, of course. Moreover, no model (including the widely accepted ones) deserves being presupposed to having high unlimited accurateness. This is somewhat impossible given the highly random nature of the radio signal and channel. Here, the accuracy level of the models have been made implicit by way of providing confidence and prediction intervals for each predicted path loss amplitude, in an attempt to proffer some general idea as to how much variability is obtainable, for each particular Tx-Rx separation distance.

After having gone though the *Results and Discussions* section, it may well have become obvious that developing an empirical model basically involves having a look at an available data (usually graphically) and then deciding on one or more *candidate models*, wherein some modification is then carried out to sooth the particular circumstance at hand.

Three empirically/statistically derived propagation models are proposed here. The first two (i.e., models -I and -II) only utilise the distance between transmitter and receiver as the predictor variable, to predict the behaviour of path loss in typical indoor and somewhat-outdoor passageways, respectively. Model-III incorporates, in addition to Tx-Rx separation distance, other predictors, such as Tx-Building separation distance, as well as other predictor functions, composed of Tx-Rx separation distance and/or Tx-Building separation distance.

It must be noted that while the models proposed here may well predict path loss under similar settings, more polished and expanded models are obtainable with supplementary field measurements and trials, thereby reducing the standard deviation and thus yielding more precise predictions. Note that the standard deviations of mean path loss corresponding to the models are 8.121dB, 9.206dB and 5.259dB, respectively, for models -I, -II and –III. Indeed, more precise models may be obtained, possibly through more measurements, as just accentuated.

Unmistakably, one path loss attenuation-factor that seems to immensely impinge on radio signal strength is that of the floor thickness of the multi-storey building. Its effect is well observed and exemplified, here, by data corresponding to model-III, where signal strength is seen to be severely weaker at an even (much) shorter propagation path length, barely as a result of more floor(s) being between the weaker, shorter channel. This can be observed from the path loss plot of Fig. 25, portraying high path loss despite the shorter (first few) Tx-Rx separation distances; clearly resulting from the RF signals in question having to propagate through one or two floors.

To sum up, as can be seen from the path loss plots of Fig. 25 and Fig. 26, the behaviour of path loss, on the whole, (for the environment corresponding to model-III) tends to show some rather nonlinear reduction, which may well be as a result of the vegetations between the transmitter and receiver for certain propagation paths, in addition to the multi-level building floor attenuation factor.

ACKNOWLEDGMENT

Emmanuel Ojakominor would like to thank Mr. Tian Fat Lai, his supervisor on this project. Mr. Lai is no doubt an expert in this area (i.e., RF propagation). Mr. Lai was helpful all the way; assisting in clearing certain doubts that Emmanuel came across while conducting the research.

Emmanuel Ojakominor would also like to extend his sincere gratitude to certain colleagues of his who assisted him during the data collection phase of this project. Carrying out signal strength (path loss) field measurements, surely, isn't a one man's job; a couple of people have to work collectively towards that. Very good friends of his, Prabu Jayakuma and Mohamed Niyaz, both helped with this part of the project. Emmanuel acknowledges the fact that he couldn't have done this research without their help.

His final thankfulness goes to the many authors and publishers (see References) from whom he has learned a great deal and acquired the necessary knowledge required in carrying out this research successfully.

REFERENCES

[1] T.S. Rappaport, *Wireless Communications: Principles and Practice*, Prentice Hall PTR, Prentice-Hall, Inc., A Simon & Schuster Company, Upper Saddle River, New Jersey 07458, 1996.

[2] T.K. Sarkar, Z. Ji, K. Kim, A. Medouri, and M. Salazar-Palma., "A Survey of Various Propagation Models for Mobile Communication," IEEE Antennas and Propagation Magazine, vol. 45, no. 3, June 2003, pp. 51-82.

[3] J.D. Parsons, *The Mobile Radio Propagation Channel*, 2nd ed., John Wiley & Sons, Ltd, Baffins Lane, Chichester, West Sussex PO19 1UD, England, 2000, pp. 1-218, 362-401.

[4] H.L. Bertoni, *Radio Propagation for Modern Wireless Systems*, Prentice Hall PTR, Prentice-Hall, Inc., Upper Saddle River, New Jersey, 2000, pp. 90-92.

[5] S. Mathur, Department of Electrical Engineering, Rutgers University, Piscataway, NJ 08904, 16:332:546 Wireless Communication Technologies Spring 2005, "Small Scale Fading in Radio Propagation," December 2008, Obtained from : http://www.winlab.rutgers.edu/~narayan/Course/Wless/Lectures05/lect3.pdf

[6] "Chapter 8: Empirical path loss and signal fading models; Chapter 9: Groundwave propagation," October 2008, Obtained from: http://www.ece.osu.edu/~johnson/713/ch7a.pdf

[7] J.L. Devore, *Probability and Statistics for Engineering and the Sciences*, 7th ed., Thomson Brooks/Cole Publishing Company, International Student Edition, 2008, pp. 130-161, 446-452, 454-464, 468-475, 477-482, 500-505, 508-517, 519-525, 528- 542.

[8] J.D. Petruccelli, B. Nandram, and M. Chen, *Applied Statistics for Engineers and Scientists*, Prentice-Hall, Inc., Simon & Schuster / A Viacom Company, Upper Saddle River, New Jersey, 1999, pp. 352, 354-355, 357, 380-384,

[9] N.C. Rogers et al., "A Generic Model of 1-60 GHz Radio Propagation – Final Report," project by QinetiQ, for the UK Radiocommunications Agency, project report date: May 2002, reference AY3880/510005719, October 2008, Obtained from: http://www.ofcom.org.uk/static/archive/ra/topics/research/topics/propagation/vegetation/vegetation-finalreportv1_0.pdf

[10] V. Erceg et al., "An Empirically Based Path Loss Model for Wireless Channels in Suburban Environments," IEEE Journal on Selected Areas in Communications, vol. 17, no. 7, July 1999, pp. 1205-1211.

[11] J. Medbo and J.-E. Berg, "Simple and Accurate Path Loss Modelling at 5 GHz in Indoor Environments with Corridors," Vehicular Technology Conference, 2000., IEEE 52nd VTS-Fall VTC 2000., vol. 1, 24-28 Sept. 2000, pp. 30-36.
[12] S.Y. Seidel and T.S. Rappaport, "914MHz Path Loss Prediction Models for Indoor Wireless Communications in Multifloored Buildings," IEEE Transactions on Antennas and Propagation, vol. 40, no. 2, February 1992, pp. 207-217.
[13] A.A. Moinuddin and S. Singh, "Accurate Pathloss Prediction in Wireless Environment," IE(I) Journal-ET, vol. 88, July 2007, pp. 9-11.
[14] B.J. Singh, K.K. Aggarwal, and S. Kumar, "Characterization of the Propagation Environment by Field Measurements," IE(I) Journal-ET, vol. 88, July 2007, pp. 22-25.

Emmanuel Ovie O. Ojakominor officially received the B.Eng. degree (with Honours) in Electronic and Communication Engineering from Curtin University of Technology, Australia, in May 2008, and is currently at the final stages of his Master's studies (M.Sc. in Personal, Mobile and Satellite Communications) with the University of Bradford, UK.

His research interests include propagation modelling, prediction and system design for wireless communications, cell planning and base station design, and wireless sensor networks.