Distance Dependence of Path Loss Models with Weighted Fitting

Aki Karttunen[1] *Member, IEEE*, Andreas F. Molisch[1] *Fellow, IEEE*, Rui Wang[1],
Sooyoung Hur[2] *Member, IEEE*, Jianzhong Zhang[2] *Senior Member, IEEE*, and Jeongho Park[2]

[1] University of Southern California, Dept. of Electrical Engineering, Los Angeles, CA, USA,
Email: {karttune, molisch, wang78}@usc.edu,
[2] Samsung, Email: {sooyoung.hur, jeongho.jh.park, jianzhong.z}@samsung.com

***Abstract*—The path loss model describing the power-law dependency on distance plus a log-normally distributed shadowing attenuation, is a staple of link budgets and system simulations. Determination of the parameters of this model is usually done from measurements and ray tracing. We show that the typical least-square fitting to those data points is *inherently* biased to give the best fitting to the link distances that happen to have more evaluation points; this bias might be highly undesirable in various types of simulations that use the resulting model. In this paper we present a weighted fitting method to address this issue. While it is unavoidable that fits are better for one distance range than another, we argue that such a decision should be made *consciously*, and adjusted to the type of simulation for which the path loss model should be used. We discuss the weighting functions for different purposes, and show their impact on prediction accuracy of signal level, interference level, and capacity in a hexagonal cellular grid simulation. As examples, weighted fitting models are presented for 28 GHz channels in urban macrocells, and it is shown that the fitting accuracy can be improved by our approach.**

I. Introduction

Path loss (PL) is the most fundamental parameter of the wireless propagation channel, determining the signal-to-noise (SNR) and signal-to-interference (SIR) ratios, and thus the coverage range as well as data rates [1]. Accurate path loss models are thus a *conditio sine qua non* for wireless system simulations. In order to properly reflect reality, path loss models should be extracted from measurements (or ray-tracing results), as has been done for many decades. Widely used models based on such parameter extraction are the Okumura-Hata models [2], the COST 231 path loss models [3], and many more, see, e.g., [4], [5], and references therein.

Most path loss models assume a power-law dependence on distance on a logarithmic scale $10 \cdot \alpha \cdot \log_{10}(d) + \beta$, which is a straight line on a "power in dB" - vs - $\log_{10}(d)$-plot. Scatter plots of measurement data can thus be fitted in an extremely simple way as a least-square fit; an approach that has been used for decades. However, as we will demonstrate in this paper, the fitting results are sensitive to the choice of measurement locations, i.e., how many measurement points exists at each distance. In other words, those distances at which the more measurement results are available, inherently provide the largest impact on the fitting result. We stress that this is different from the issue that the parameters are different in different environments (in that case the propagation physics are responsible for the different parameters); rather the dependence of the measurement locations is a *statistical artifact* that stems from: (i) a deviation of the model from the actual propagation law, and (ii) a finite number of measurement points. Such a bias should be avoided as much as possible, as it distorts the results of system simulations done with the biased path loss model.

While the location bias effect seems obvious in hindsight, it has, to our knowledge, not yet been recognized and discussed in the literature.[1] Besides understanding the nature of the effect, an important question is if, and how, it can be remedied. The current paper is intended to address these questions.

The main contributions of this paper are thus: (i) we point out the issue of the distance bias, (ii) we present a method for compensating for the bias, through weighting of the measurement results by their measurement point density and (iii) we provide examples of this effect based on extensive ray-tracing results at millimeter-wave frequencies. The method outlined here is being adopted by an industry standardization group for millimeter-wave channels, though we stress that it is generally applicable, not just for mm-wave channels.

The rest of the paper is organized as follows: the path loss data, used as an example in this paper, are introduced in Section II, the path loss models and model fitting is presented in Section III, and the derived models are compared to the original path loss data in Section IV.

II. Path Loss Data

The principles of the location bias will be explained by means of example ray-tracing data for mm-wave propagation channels in an urban environment. Specifically, in this work, we model the New York University (NYU) campus area in Manhattan, NY, USA, based on 3D building models. Fig. 1 shows the geometry used for the ray-tracing simulation, which illustrates the NYU campus model in an area that is 920 m

[1]The effect is also related to, though different from, the effect of *selection* bias recently discussed in [6]; this selection bias occurs because only measurement points are used at which the received power is large enough to actually allow quantitative measurements.

in length and 800 m in width.[2] We included vegetation in the building model, and the area of Washington Square Park in the center of Fig. 1 is modeled with trees in it. The trees are modeled as 40 dB loss, and the vegetation is placed from 4∼5m over the ground and up to 13 m in height. In the geographical model, all the buildings and the ground are assumed to be made of concrete and wet earth, respectively. Similar environment settings are used as in ITU-R M.2135 and 3GPP SCM for UMa.

In the ray-tracing simulations, the signal power samples are collected with 5 m resolution within the observation area. The transmitter is placed 5 m above the rooftop.

The parameters for ray-tracing simulations are the same as used in [7], [8]. All paths are modeled by reflection, diffraction, and penetration based on geometrical optics (GO) and uniform theory of diffraction (UTD) using the ray-tracing software Wireless In-site designed by REMCOM [9]. In each ray, at most twelve reflections, two penetrations, and a single diffraction are considered because attenuation by multiple diffractions and combination of reflection and diffraction at mm-wave frequencies are very severe. At each combination of TX and RX, at most 40 rays and their signal power, phase, propagation time, direction of departures (DoDs) and direction of arrivals (DoAs) for both azimuth and elevation are collected in descending order of received power within 250 dB per-path path loss. The received power of each RX point is calculated as a sum of all the path powers, and only this channel parameter is used for further analysis. Furthermore, each RX location is categorized as line-of-sight (LOS) or non-line-of-sight (NLOS) by the definition of visual LOS between the TX an RX.

In order to validate the ray-tracing approach, we performed ray-tracing simulation on the spots where the NYU Wireless team conducted 28 GHz-band field measurement campaign in the same area [10]; (these TX locations marked with red dots in Fig. 1). The comparison of the measured channels and the ray tracing is presented in [11].

A total of eleven transmitters (TXs), i.e., base stations (BS), are simulated in order to eliminate geometry dependency of the statistical results. The receiver (RX), i.e., mobile station (MS), height is set to be 1.5 m above the ground, and only outdoor RX points are simulated.

III. Path Loss Models with Weighted Fitting

Available path loss data are typically unevenly distributed across the link distances: often the path loss data distribution is determined by, or at least limited by, practical limitations in the measurements or the ray-tracing environment. Thus, when performing a fit of a path loss model to the measurement results, the distances that *by accident* have more data points, have a bigger impact on the fitting parameters. With appropriate weighing it is possible to "equalize" this effect, i.e., give

[2]Note that this is the area for which the building database is available; the actual observation area (in which TX and RX locations can be) is the smaller 560 m by 550 m area within the red square in Fig. 1. The reason for this difference is that all relevant interacting objects (i.e., buildings that can give rise to reflections, scattering, and diffraction) for all observation points have to be contained in the database.

Fig. 1. Models for ray-tracing simulation

relatively more weight to distances with fewer data points, and thus improve fitting accuracy for those distances. On the other hand, we might also want to selectively improve the modeling accuracy for the shortest or the longest link distances.

Using distance dependent weighted fitting can also be seen as an alternative to using a more complicated model with larger number of free parameters, e.g., dual-slope model in [8]. Well parameterized models with a sufficiently large number of optimized parameters will fit the data for every part of the distance range with any weighing. On the other hand, a simpler model is a compromise and does not fit the data equally for every part of the dataset link distance range.

The distribution of data points within the 28-GHz ray-tracing dataset is illustrated in Fig. 2. As can be seen, both the LOS and NLOS datasets are quite unevenly distributed. Note that later we use a discrete approximation to the point density pdf; the choice of the "bin size" for the distance bins has to be a compromise between a sufficiently fine distance resolution, and including a sufficient number of measurement points in each bin (see also discussion below).

The path loss model that we consider is the classical "power law": define the small-scale-averaged (SSA) path gain as the instantaneous (local) channel gain averaged over the small-scale fading, and the large-scale averaged (LSA) channel gain as the SSA channel gain averaged over shadow fading. Then the path loss is the inverse of the LSA channel gain. Then we model the path loss on a dB-scale as

$$PL_m(d) = 10\cdot\alpha\cdot\log_{10}(d) + \beta, \tag{1}$$

and the variations of the SSA path loss around the mean is modeled as a zero-mean log-normal distribution, so that on a dB scale, it is described by a normally distribution random variation $N(0,\sigma)$ from the mean-value line given by (1). The standard deviation can be either a constant or a distance dependent function:

$$\sigma(d) = a\cdot\log_{10}(d) + b. \tag{2}$$

Define now a weighted log-likelihood function

$$LLF = \sum_{i=1}^{N_s} w(d_i)\log\frac{e^{-\frac{1}{2\sigma^2}(PL_m(d_i)-PL_d(d_i))^2}}{\sqrt{2\pi}\sigma} \tag{3}$$

where N_s is the total number of data samples, σ is the standard deviation (std) of the distribution, w is the weight, PL_m is the path loss model and PL_d is the path loss data at distance d_i given in dB.

The path loss parameters are determined as the minimum of the negative of the log-likelihood function as

$$\arg\min_{\alpha,\beta,\sigma}\{-LLF\} \tag{4}$$

or with $\sigma(d_i) = a\cdot\log_{10}(d_i) + b$:

$$\arg\min_{\alpha,\beta,a,b}\{-LLF\}. \tag{5}$$

We note that in many papers, the determination of the path loss parameters has not been done by joint minimization of the log-likelihood function, but instead by first performing a least-squares fit of the path loss Eq. (1), and then determining a (distance-independent) σ as the mean-square deviation from the fit. This approach, while popular, gives slightly sub optimum results. In any case, this distinction is not relevant for the main point of this paper, namely the impact of the density of measurement points.

We now analyze different weighting functions. We consider four different approaches:

- a) equal weight to each data point. This is the default solution, used in many previous papers. We stress that while it was not recognized at the time, using all points with equal weighting is in itself a form of weighting, and it *will* bias the results. The weights are determined by the placement of measurement points, which might not be meaningful for any later system simulations that use the model.
- b) equal weight to N bins over link distance d. This is an "intuitively pleasing" weighting, which gives slightly better fit for both short and long distance that have relatively few data points in measurements.
- c) equal weight to N bins over $\log_{10}(d)$. This weighting is most closely aligned with a "least square" fitting on a dB-vs-$\log_{10}(d)$ plot. Each equal-sized "bin" on the $\log_{10}(d)$ axis obtains equal importance for the fitting. As a consequence of this approach, we obtain a better fit for short distances and worse fit for largest link distances, compared to option b).
- d) equal weight to N bins over d^2. This approach is most appealing as the basis for system simulations with uniform random "drops" (i.e., choice of mobile station location) in a geographical area. Then the weights become proportional to the number of MSs that will be dropped in a particular distance bin. As a consequence, we obtain a better fit with the largest link distances and worse fit for short distances.

The weights w for (3) are inversely proportional to number of data points in a bin:

$$w = \frac{1}{N_b}\frac{N_s}{N}, \tag{6}$$

where N_b is number of data points in the bin and N_s is the total number of points. The fitting accuracy is improved for $w(d_i) > 1$ and data points with $w(d_i) < 1$ have reduced effect on the fitting as compared to weighing option a) which has the $w(d_i) = 1$ for each point. In our examples, the number of bins is selected as $N = 30$ for each weighing option.

In order to illustrate the distance dependent weights define:

$$W = N_b \cdot w, \tag{7}$$

where W is the apparent data point density, N_b is number of data points on a linear scale (or logarithmic etc.), and w are the weights that are used in the weighted LLF-equation. The apparent data point density W averaged over 20 m is presented in Fig. 2. The W is approximately constant over

TABLE I
LOS PARAMETERS FOR (1)-(2)

Method	α	β	σ
a	1.60	66.62	2.10
	1.69	64.78	1.52· $\log_{10}(d)$-1.23
b	1.55	67.48	2.36
	1.67	64.88	1.54· $\log_{10}(d)$-1.20
c	1.60	66.56	2.00
	1.68	64.93	1.45· $\log_{10}(d)$-1.17
d	1.46	69.58	2.53
	1.49	68.83	1.07· $\log_{10}(d)$-0.10

TABLE II
NLOS PARAMETERS FOR (1)-(2)

Method	α	β	σ
a	8.37	-53.94	22.58
	7.77	-39.88	14.97· $\log_{10}(d)$-14.44
b	8.10	-46.78	22.64
	7.46	-31.14	15.46· $\log_{10}(d)$-15.93
c	7.81	-40.10	20.48
	7.13	-24.38	16.57· $\log_{10}(d)$-18.95
d	8.11	-47.24	24.26
	7.71	-37.07	15.03· $\log_{10}(d)$-14.46

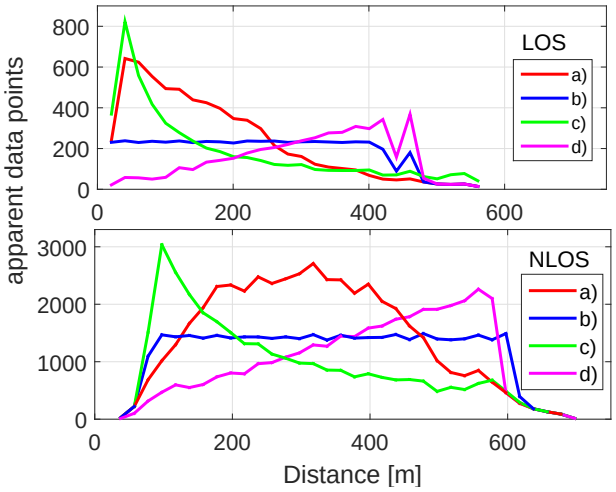

Fig. 2. Apparent data point density per 20 m in LOS and NLOS, i.e., the fitting weights w multiplied by number of points averaged over 20 m. The red lines (a) are the true data point densities.

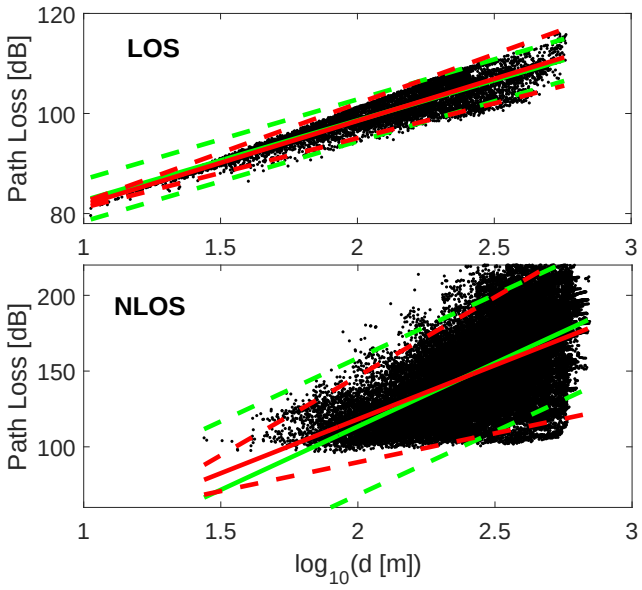

Fig. 3. LOS and NLOS PL-model fitting examples; constant σ and a) equal weight to each data point (green), distance dependent σ and c) equal weight to each bin as function of $\log_{10}(d)$ (red). Dashed lines indicate the $PL(d) \pm 2\sigma(d)$-lines.

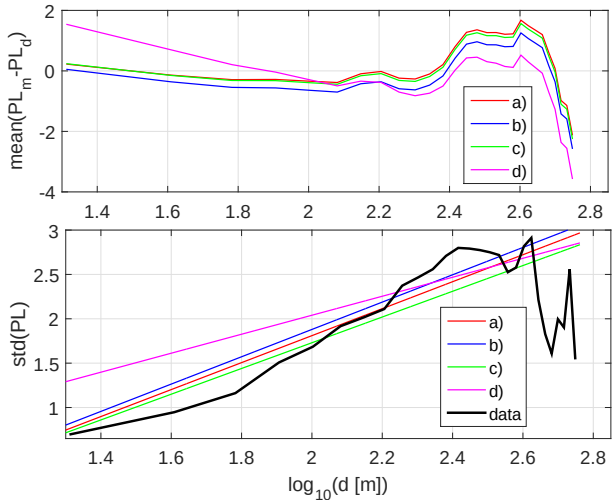

Fig. 4. Comparison of data and models for LOS with distance-dependent σ. Average of $PL_m[dB] - PL_d[dB]$ and std of the models and data over 20 m.

distance d, $\log_{10}(d)$, and d^2 with weighing options b), c), and d), respectively.

We are imposing as an additional constraint that the bins with the smallest number of data points, in total including about 2% of the total number of data points, are limited to $w = 1$. This prevents giving huge weights to severely undersampled distance ranges with very small number of points associated with high variance of the realizations from the mean. The limits on number of bins and the definition of the weights affect model matching results, but for simplicity and brevity, these effects are not studied in this paper.

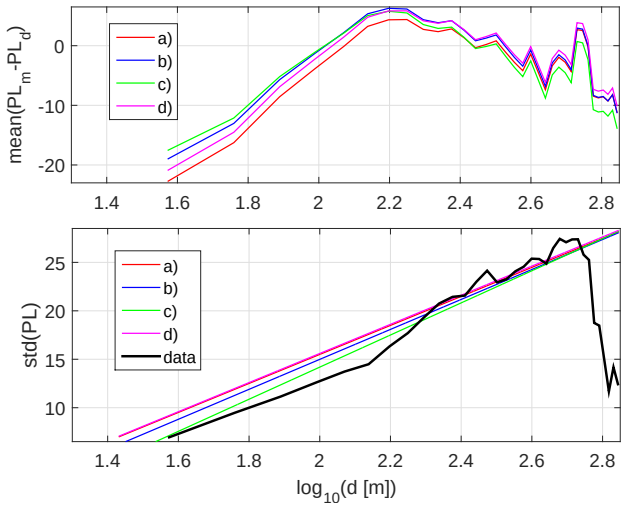

Fig. 5. Comparison of data and models for NLOS with distance-dependent σ. Average of $PL_m[dB] - PL_d[dB]$ and std of the models and data over 20 m.

The path loss model parameter values with different weighings are presented in Tables I and II.[3] Examples of PL data and fitting results are given in Fig. 3. In the LOS-case, the model parameters α and β are quite similar for the different weightings and the main differences are in the σ values. The α and β with constant σ models are different than with the distance dependent σ-models. This clearly shows the importance of optimizing the power-law model (1) together with the σ-model.

In general, comparing the weighings a)-d) for both LOS and NLOS, it can be noted that the c)-models shows the strongest differences from the others. This means that the PL data for the shortest link distances follow a slightly different distribution. In cases when the relatively short link distances are important, using these fits could improve PL-model accuracy.

Statistics of the difference between models and data (averaged with a 20 m sliding window) are presented in Figs. 4-5 for the distance-dependent σ-models. In general it can be observed that in the LOS-case, the models fit the data better than in the NLOS-case. In LOS, the average difference $PL_m[dB] - PL_d[dB]$ is close to zero for most of the distance range, with all the weighing options, which indicates that the LOS data fits the PL-model (1) very well. When analyzing the original ray-tracing data, we see that the standard deviation (std) of the path loss ($std(PL_d[dB])$) is increasing as a function of the distance; thus a distance-dependent model of the shadowing variation provides better results. When a constant σ is used, the σ of the PL models is selectively adjusted to fit the standard deviation of the underlying raw data either for the shortest distances or the longest, depending on the weighting option (see Tables I and II). For example, the weighing option c) makes σ fit well to the $std(PL_d[dB])$ occurring at the

[3]Note also that the derived models are valid only within the distance range of the underlying data (a point that seems obvious, yet is often forgotten in the application of path loss models).

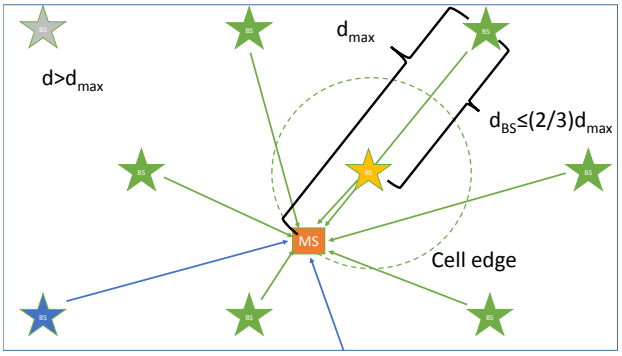

Fig. 6. Simulation setup: (i) hexagonal grid of base stations, (ii) cell edge at radius $d_{BS}/2$, (iii) signal from base station with smallest PL and interference from others, and (iv) nothing from BS that are too far away (no extrapolation). If d_{max} is the validity range limit of the PL-model then maximum distance between base stations is defined as $d_{BS} = (2/3)d_{max}$, as shown in the illustrated geometry.

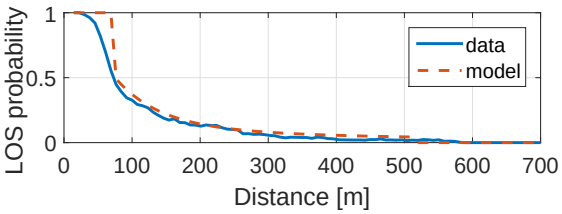

Fig. 7. LOS probability.

shortest link distances; these values are smaller, so a smaller σ is fitted. The distance dependent σ-models correctly model the distance dependency of the standard deviation from the local average path loss for most of the distance range.

IV. Comparison in Hexagonal Cellular Grid

In this section the channel models are compared in terms of prediction accuracy of signal level, interference level, and capacity in a hexagonal cellular grid simulation. The simulation setup is illustrated in Fig. 6. The users are dropped randomly inside the cell. The signal level from each base station is calculated with the path loss models derived in Section III. For comparison, data points with same distribution of link distances are drawn from the path loss data.

The base station is either in LOS or NLOS based on LOS probability. The same LOS probability is used for the ray tracing model and for selecting the data points from the ray-tracing data. This is done to avoid influencing the comparison of the *path loss fitting* models to the reality by the accuracy of the *LOS probability model*. Specifically, the LOS probability is modeled by the model of 3GPP (3GPP TR 36.873 UMa, without antenna height)

$$p(d) = \min(1, \frac{d_1}{d})(1 - e^{-d/d_2}) + e^{-d/d_2}, \quad (8)$$

with limitations based on available ray-tracing data for comparison:

$$p(d) = 1, d < d_3 \quad (9)$$

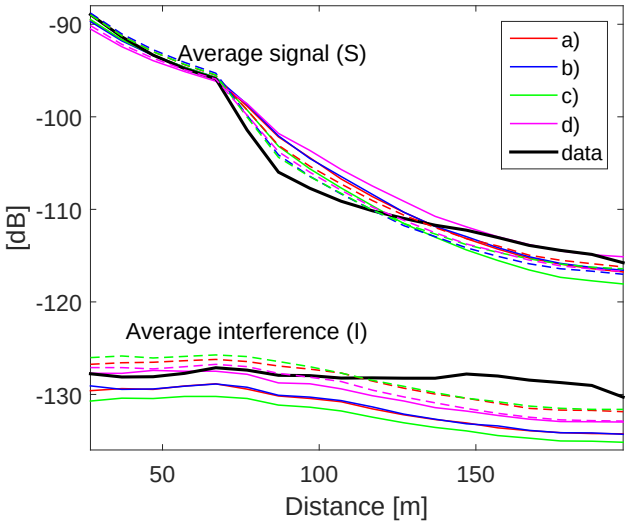

Fig. 8. Average signal and interference over 10 m as a function of distance to the closest BS. Constant σ with solid lines and distance dependent σ with dash lines.

$$p(d) = 0, d > d_4 \quad (10)$$

where $d_1 = 22$ m, $d_2 = 61$ m, $d_3 = 76$ m, $d_4 = 512$ m. LOS probability data and the model are illustrated in Fig. 7. Less than 1% of the NLOS data are at distances less than d_3 and less than 1% of the LOS data are at distances greater than d_4. Similarly the shortest simulated link distance is limited to 22 m, because less than 1% of the LOS data are at shorter distances. The longest simulated link distance is limited to 622 m, because less than 1% of the NLOS data are at longer distances.

We denote as the "signal level" $S = -min(PL)$ the strongest received power (i.e., we assume that each MS is associated with the BS to which it has the smallest SSA path loss), and the interference level is the sum of received signals from all other base stations.

The capacity with interference-plus-noise C_{IN} are calculated as: (the unit for the capacity values is [bits/s/Hz])

$$C_{IN} = \log_2(1 + SINR), \quad (11)$$

where SINR is the signal-to-interference-plus-noise ratio.

Simulation results as a function of distance to the closest base station are presented in Figs. 8-9. Here the distance between base stations in the hexagonal cellular grid is 400 m. The noise level corresponds to PL = 130 dB, e.g., if PL = 130 dB then $S = -130$ dB and SNR = 0 dB. The noise level of 130 dB is selected for this example so that the cell averaged SNR and SIR are approximately equal.

The signal and interference level modeling accuracy is presented in Fig. 8. In general, distant dependent σ gives better match. The weighing option c), which improves the model fitting for the shortest distances, is good for predicting the signal level at those distances. The weighing option c), with the distant dependent σ, predicts the average signal level very well and also the interference level quite accurately for the

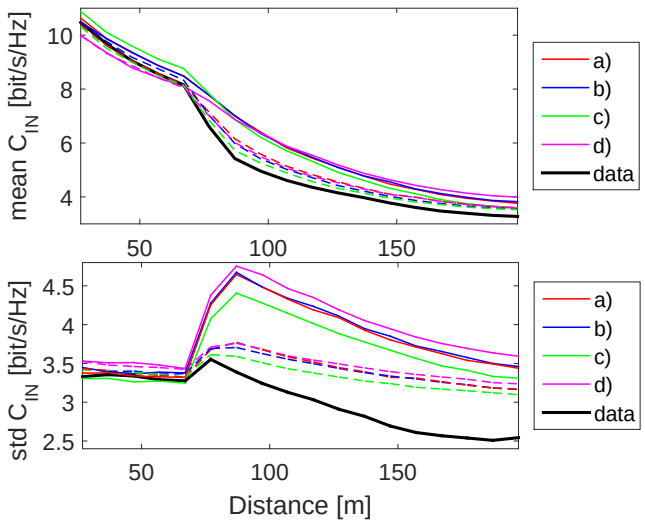

Fig. 9. Average and standard deviation of C_{IN} and over 10 m as a function of distance to the closest BS. Constant σ with solid lines and distance dependent σ with dash lines.

whole cell. However, option c) with distance independent σ is poor at predicting signal power and interference at larger distances, as is to be expected from the definition.

The capacity with interference plus noise C_{IN} is affected by both LOS and NLOS with both short and long link distances. Therefore, the modeling accuracy of the average and standard deviation of C_{IN}, in Fig. 9, can be considered as a good measure of how much the differences in the models really matter. Again, it can be noted that the distant dependent σ gives better match. Also it can be noted that the weighing option c) usually outperforms others. The weighing option c) with the distant dependent σ gives the best match. This is also related to the fact that average cell capacity is largely determined by the very high capacities at the cell center.

The results in Figs. 8-9 show that the option a), with the *unintended* distant dependent weighing, rarely gives the best match. In general, the weighing should be chosen based on the intended purpose of the model.

V. Conclusion

The typical path loss data is measured (or simulated) at locations whose density is unevenly distributed over the link distances. This can cause unintended bias in the path loss model fitting, favoring those link distances with more data samples. We present a path loss model fitting with distance dependent weighing that provides better model fitting evenly across the distances or selectively to improve accuracy for the shortest or longest link distances. The criteria for the weights should be adapted to the goals of subsequent system simulations.

As an example, we discuss path loss models for 28 GHz channels in an urban macrocellular scenario. The path loss data are simulated with ray tracing with 11 different base station locations. The model parameters are given with different weightings and the modeling accuracy is examined as a function of the distance. It is shown that the modeling accuracy can be improved for the weighted link distance range. The models and the path loss data are compared with signal and interference level analysis in a hexagonal cellular grid. It is shown that by appropriate weighting the accuracy at cellular grid level analysis can be improved. Another interesting insight is that a distance-dependent modeling of the variance significantly improves the fit between the raw ray-tracing results and the model.

The results presented in this paper show that the presented method can improve path loss modeling accuracy. It is shown that even in the absence of any conscious weighing there is an implicit weighting, that thus impacts the results. However, care must be taken to choose the weights in a manner that most closely reflects the later application of the path loss model.

Acknowledgment

A. Karttunen would like to thank the Walter Ahlström Foundation and the ”Tutkijat maailmalle”-program for financial support. Part of the work was supported by the National Science Foundation. We thank Prof. Fredrik Tufvesson for helpful discussions.

References

[1] A. F. Molisch, *Wireless Communications*, 2nd ed. IEEE Press - Wiley, 2010.

[2] Y. Okumura, E. Ohmori, T. Kawano, and K. Fukuda, “Field strength and its variability in vhf and uhf land-mobile radio service,” *Rev. Elec. Commun. Lab*, vol. 16, no. 9, pp. 825–873, 1968.

[3] E. Damosso and L. M. Correia, *COST Action 231: Digital Mobile Radio Towards Future Generation Systems: Final Report*. European Commission, 1999.

[4] C. Phillips, D. Sicker, and D. Grunwald, “A survey of wireless path loss prediction and coverage mapping methods,” *Communications Surveys & Tutorials, IEEE*, vol. 15, no. 1, pp. 255–270, 2013.

[5] K. Haneda, “Channel models and beamforming at millimeter-wave frequency bands,” *IEICE Transactions on Communications*, vol. 98, no. 5, pp. 755–772, 2015.

[6] C. Gustafson, D. Bolin, and F. Tufvesson, “Modeling the cluster decay in mm-wave channels,” in *Antennas and Propagation (EuCAP), 2014 8th European Conference on*, April 2014, pp. 804–808.

[7] S. Hur, S. Baek, B. Kim, J. Park, A. Molisch, K. Haneda, and M. Peter, “28 ghz channel modeling using 3d ray-tracing in urban environments,” in *Antennas and Propagation (EuCAP), 2015 9th European Conference on*, April 2015, pp. 1–5.

[8] Y. Chang, S. Baek, S. Hur, Y. Mok, and Y. Lee, “A novel dual-slope mm-wave channel model based on 3d ray-tracing in urban environments,” in *Personal, Indoor, and Mobile Radio Communication (PIMRC), 2014 IEEE 25th Annual International Symposium on*, Sept 2014, pp. 222–226.

[9] REMCOM, “Wireless InSite,” http://www.remcom.com/wireless-insite, [Online; accessed 16-Oct.-2015].

[10] M. Samimi and T. Rappaport, “3-d statistical channel model for millimeter-wave outdoor mobile broadband communications,” in *Communications (ICC), 2015 IEEE International Conference on*, June 2015, pp. 2430–2436.

[11] S. Hur, S. Baek, B. Kim, Y. Chang, A. Molisch, T. Rappaport, K. Haneda, and J. Park, “Proposal on millimeter-wave channel modeling for 5g cellular system,” 2016, *accepted to* IEEE Journal of Selected Topics in Signal Processing.