

Chapter 3

Bounding the Error of Path Loss Prediction

Despite the large quantity of work done on modeling path loss, there is an important shortcoming that this chapter begins to address: there have been relatively few comparative evaluations of path loss prediction models using a sufficiently representative data set as a basis for evaluation. Those studies that do exist only make comparisons between a small number of similar models. And, where there has been substantial work of serious rigor done, for instance in the VHF bands where solid work in the 1960's produced well validated results for analog television (TV) propagation, it is not clear how well these models work for predicting propagation in different types of systems operating at different frequencies. The result is that wireless researchers are left without proper guidance in picking among dozens of propagation models. Further, among the available models it is not clear which is best or what the penalty is of using a model outside of its intended coverage. In [44], for instance, Camp *et al.* show that a wireless mesh network planned with a given path loss model can be massively under- or overprovisioned as a result of small changes to model parameters. For the purpose of this thesis, it is crucial to put practical bounds on the performance of existing methods in order to define a clear benchmark of success.

This chapter analyzes 30 propagation models spanning 65 years of publications using five novel metrics to gauge performance. Although many of these models are quite different from one another, they all make use of the same basic variables on which to base their predictions: position (including height and orientation) of the transmitter and receiver, carrier frequency, and digital elevation model and land cover classification along the main line-of-sight (LOS) transmit path. These models utilize a mix of approaches:

⁰ Work in this chapter has appeared in [173, 174, 167, 172]. Data collected for the experiments in this chapter has been made publicly available at [166, 170].

empirical, (purely) analytical, stochastic or some combination thereof. They are tested in this analysis without starting bias as to which should perform best. Active-measurement models (e.g., [200] and the geostatistical approach advocated by this thesis), which make use of directed *in situ* measurements to correct their predictions are not considered here, as they are the focus of later chapters in the thesis.

The focus in this chapter is the efficacy of the models studied at the task of *predicting* median path loss values in environments with representative terrain and a large range of equipment and link lengths. Many authors have considered the problem of predicting outdoor path loss in uncluttered environments to be solved. We will see this is far from true—making accurate *a priori* predictions about path loss, without *in situ* measurements, with the models available, is a very difficult task even in “simple” environments.

In the end, the results show that no single model is able to predict path loss consistently well. Even for the seemingly simple case of long links between well-positioned antennas in a rural environment, the available models are unable to predict path loss at an accuracy that is usable for any more than crude estimates. Indeed, no model is able to achieve a RMSE of less than 14 dB in rural environments and 8–9 dB in urban environments—a performance that is only achieved after substantial fine tuning. Explicit data-fitting approaches do not perform better, producing 8–9 dB RMSE as well. This conclusion motivates the work on more rigorous measurement based approaches that forms the remainder of this thesis.

3.1 Measurement

This section describes data sets collected to address the goals of this chapter. These measurements were collected over the course of several years in multiple environments and with differing (but consistent) hardware. They range from “clean” measurements taken in rural New Zealand, to “noisy” measurements collected in the urban center of a large US city along with some special measurements to investigate points of particular interest, such as measurements with phased array and directional antennas, and some in suburban environments. Overall, these data sets combine to paint a unique picture of the real-world wireless radio environment at varying levels of complexity. Table 3.1 provides a summary of these data sets.

Campaign	Name	Environment	Type	Frequency	Method	Transmitters	Measurements
A	wart	Campus	Point-to-Point	2.4 GHz	Packet	7	33,881
A	wart/snow	Campus	Point-to-Point	2.4 GHz	Packet	7	24,867
B	pdx	Urban	Urban Mesh/Infrastructure	2.4 GHz	Packet	250	≈ 117
B	pdx/stumble	Urban	Urban Mesh/Infrastructure	2.4 GHz	Packet	59,131	200,694
C	boulder/ptg	Campus	Infrastructure/Downstream	2.4 GHz	Packet	1,693	1,693
C	boulder/gtp	Campus	Infrastructure/Upstream	2.4 GHz	Packet	329	329
D	cost231	Urban	Infrastructure/Downstream	900 MHz	Continuous Wave (CW)	2,336	2,336
E	wmp/a	Rural	Point-to-Point/Infrastructure	5.8 GHz		368	2,090,943
E	wmp/g	Rural	Point-to-Point/Infrastructure	2.4 GHz	Packet	368	20,314,594
F	tfa	Suburban	Mesh/Infrastructure	2.4 GHz	Packet	22	389,401
G	google	Urban/Suburban	Mesh/Infrastructure	2.4 GHz	Packet	168	75,101

Table 3.1: Summary of data sets

3.1.1 Packet-Based Measurements

With the exception of the COST-231 data, discussed in section 3.1.3.3 below, all data sets used in this thesis were collected using commodity hardware and packet-based measurements were used to determine received signal strength. This approach differs from some prior work on path loss modeling that uses continuous wave (CW) measurements [91, 48]. When using packet-based methods to collect information about received signal strength and path loss, a transmitter is configured to transmit “beacon” frames periodically. A (often mobile) receiver records these beacon frames. Using an open source driver, such as MadWifi [12], and a compatible chipset, frames can be recorded in their entirety to the harddisk in real-time using any number of userspace software tools (e.g., tcpdump). If these frames are recorded with the optional Radiotap header [8] (or equivalently, the more archaic Prism II header) then the record will include information about the physical layer, such as the received signal strength of the frame, any Frame Check Sequence (FCS) errors, and a noise floor measurement. Using this approach, inexpensive commodity hardware can be used to make extensive passive measurements of a wireless network.

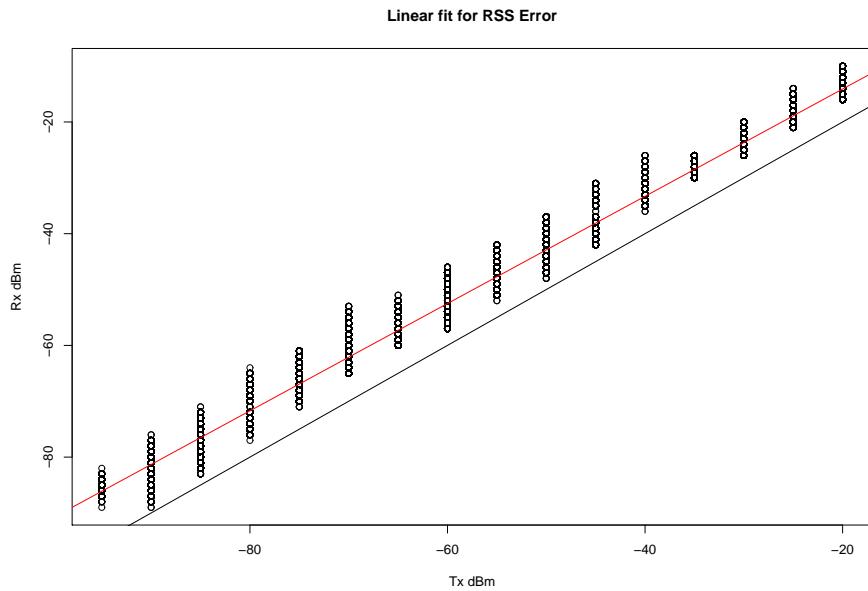


Figure 3.1: Linear fit to RSS error observed from commodity cards during calibration.

To get an idea of how accurate commodity radios are in measuring Received Signal Strength (RSS),

some calibration experiments were performed in a conductive setting. Each of four radio cards was directly connected to an Agilent E4438C Vector Signal Generator (VSG). The cards were all Atheros-based Lenovo-rebranded Mini-PCI Express, of the same family (brand and model line) chipset to those used for all of our packet-based measurements. The VSG was configured to generate 802.11 frames and the laptop to receive them. For each of the four cards many samples were collected while varying the transmit power of the VSG between -20 dBm and -95 dBm (lower than the receive sensitivity threshold of just about any commodity 802.11 radio) on 5 dB increments. Finally, a linear least squares fit was performed, finding a slope of 0.9602 and adjusted R-squared value of 0.9894 (indicating a strong fit to the data). Figure 3.1 shows this data and the fit line. The commodity radios perform remarkably well in terms of RSS measurement. To correct for the minor error they do exhibit, the slope of this fit can be used to adjust our measurements, dividing each measurement by the slope value.

However, there is a drawback to this approach. Packet-based methods necessarily “drop” measurements for packets that cannot be demodulated. All receivers have fundamental limits in their receive sensitivity that are a function of their design. However, because packet-based measurement techniques rely on demodulation of packets to determine the received signal strength, they have a necessarily lower sensitivity than receivers that calculate received power from pure signal (i.e., continuous wave measurements). Additionally, without driver modification, commodity receivers generally update noise floor measurements infrequently. For the purpose of analyzing accuracy of median path loss prediction (as is done in chapter 3), these limitations are not problematic. In one sense, commodity hardware “loses” only the least interesting measurements—if we are unable to decode the signal at a given point, we are at least aware that the signal is *below* the minimum detectable signal for basic modulation schemes, and is as a result, unlikely to be usable for many applications.

It should be noted that packet-based measurement methods are not appropriate for all modeling tasks—the tradeoff between convenience and affordability of commodity hardware versus the completeness of the measurements must be considered. For instance, if the goal of a measurement campaign is to sense signals or interference near the noise floor in order to predict capacity for next generation protocols, or if the goal is to model delay spread or Doppler shift, then packet-based measurements will not be suffi-

cient. However, the work in this chapter has less demanding data requirements than these applications. For the purpose of measuring median SNR at a given point in space from the perspective of a typical receiver, packet-based measurements made with commodity hardware are both sufficiently accurate and convincingly representative.

3.1.2 Rural Measurements

In cooperation with the Waikato Applied Network Dynamics (WAND) research group at the University of Waikato [13] and the RuralLink wireless internet service provider (WISP) [9], a large set of measurements was acquired from a commercial network in rural New Zealand. These measurements were collected for the Wireless Measurement Project (WMP) [185]. Rural environments are simpler both in the sense that there are fewer obstacles to cause fading, and those obstacles that do exist are typically large and constant (e.g., mountains and terrain features) which produce only large scale shadowing and minimal small scale (fast) fading. Moreover, the isolated nature of rural networks result in less interference from neighboring competing networks, which can create random fades that are difficult to predict and model. Hence, the measurements here are intended to form a comparative baseline for the measurements in more complex environments.

The network used in this study is a large commercial network that provides Internet access to rural segments of the Waikato region in New Zealand (as well as some in other regions). The overall approach to measurement involves periodically broadcasting measurement frames from all nodes and meanwhile recording any overheard measurement frames. Every two minutes, each device on the network transmits a measurement frame at each supported bitrate. Meanwhile, each device uses a monitor mode interface to log packets. Because this is a production network, privacy concerns are of clear importance, which is why all measurements are made with injected packets and a Nondisclosure Agreement (NDA) was required for use of parts of the data that contained sensitive information (principally client locations).

The network is arranged in the typical hub and spoke topology as can be seen in figure 3.2. The backhaul network is composed of long distance 802.11a links operating at 5.8 GHz. Atypically liberal power regulations in New Zealand and Australia around 5.8 GHz allow for much longer links than can be seen in

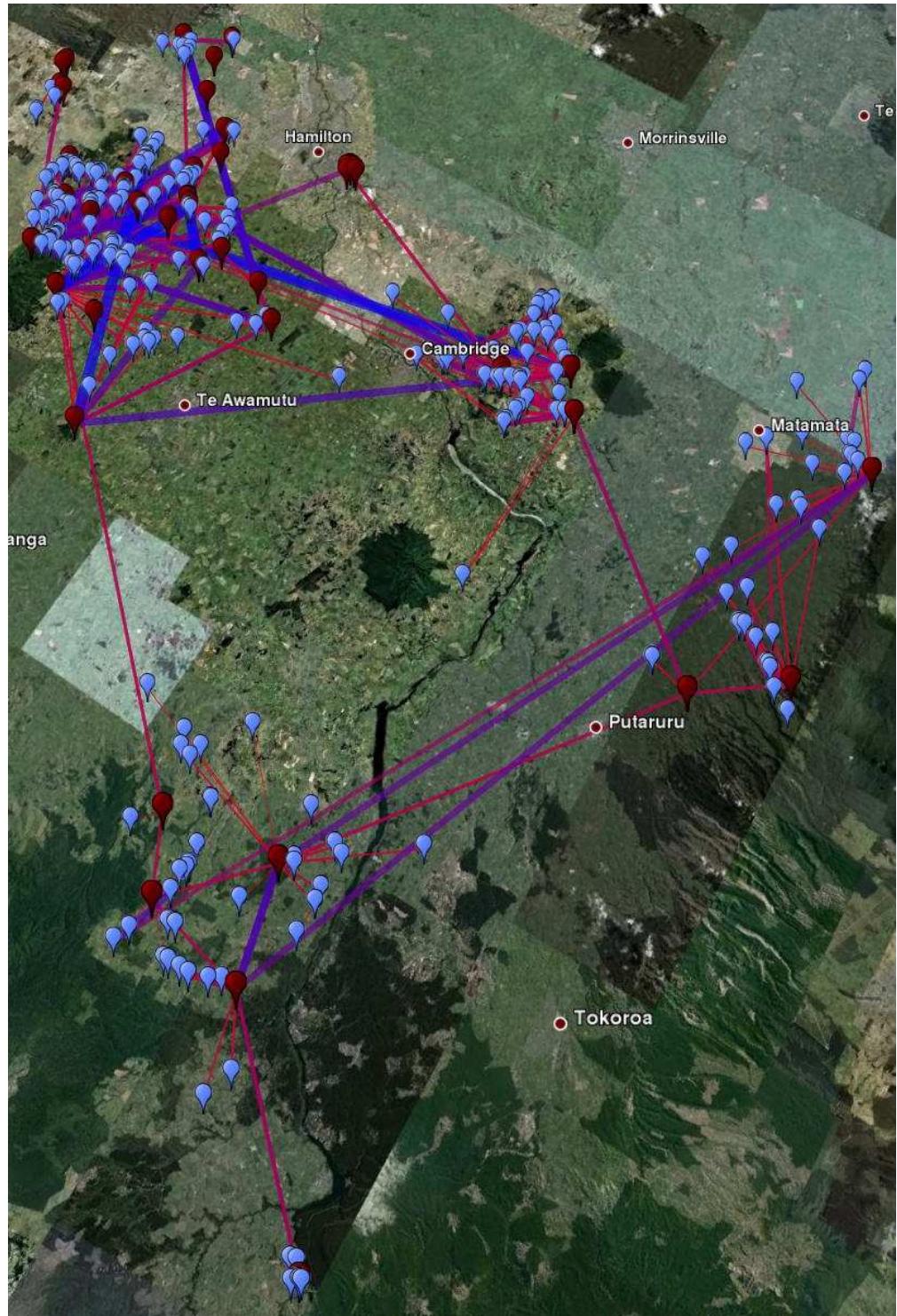


Figure 3.2: The largest of three disconnected sections of the network (80x100km). Link color indicates strength: blue implies strong, red implies weak. Backhaul nodes (mainly 5.8 GHz) are red and CPEs are light blue.

most other places in the world—40 km is a typical link length in this network¹. These are commonly point-to-point links that use highly directional antennas that are carefully steered. The local access network is composed of predominantly 802.11b/g links that provide connectivity to Client Premises Equipment (CPE). Often, an 802.11g AP with an omnidirectional or sector antenna will provide access to a dozen or more CPE devices that have directional (patch panel) antennas pointing back to the AP. With few exceptions, each node in the network is an embedded computer running the Linux operating system that allows us to use standard open source tools to perform measurement and monitoring. All nodes under measurement use an Atheros-brand radio and the MadWifi driver [12] is used to collect frames in monitor mode and record received signal strengths using the radiotap extension to libpcap [8].

After collection, the data requires scrubbing to discard frames that have arrived with errors. Because there is substantial redundancy in measurements (many measurements are made between every pair of participating nodes), discarding some small fraction of (presumably randomly) damaged frames is unlikely to harm the integrity of the data overall. As a rule, any frame that arrives with its checksum in error or those from a source that produces less than 100 packets is discarded. For the work in this thesis, one representative week of data collected between July 25th, 2010 and August 2nd, 2010 is used. Because detailed documentation about each node simply did not exist, some assumptions were made for analysis. The locations of nodes for which there is no specific GPS reading are either hand coded, or in the case of some client devices, geocoded using an address. Antenna orientations for directional antennas are assumed to be ideal—pointing in the exact bearing of their mate. All nodes are assumed to be positioned 3 m off the ground, which is correct for the vast majority of nodes. While these assumptions are not perfect, and are clearly a source of error, they are reasonably accurate for a network of this size and complexity. Certainly, any errors in antenna heights, locations, or orientations are on the same scale as those errors would be for anyone using one of the propagation models analyzed to make predictions about their own network of interest.

In the end, the scrubbed data for a single week constitutes 19,235,611 measurements taken on 1,328 links (1,262 802.11b/g links at 2.4 GHz and 464 802.11a links at 5.8 GHz) from 368 participating nodes.

¹ Fixed radio links (Unlicensed National Information Infrastructure (U-NII) devices) operating between 5.725 and 5.825 GHz that use wide band digital modulation are allowed an EIRP of 200 W [1].

Of these nodes, the vast majority are clients and hence many of the antennas are of the patch panel variety (70%). Of the remaining 30%, 21% are highly directional point-to-point parabolic dishes, and 4.5% each of omnidirectional and sector antennas.

3.1.3 Urban Measurements

In addition to the “baseline” measurements in a rural setting, measurements were collected in three additional environments to complete the picture of the urban/suburban wireless propagation environment. Figure 3.3 provides a schematic of the three urban data sets and table 3.1 provides further details. The three campaigns cover the three transceiver configurations that are most important in the urban wireless environment. The first, **A**, concerns well-positioned (i.e., tower or rooftop) fixed wireless transceivers. This sort of link is typically used for backhaul or long distance connections (e.g., [20]). The second, **B**, concerns propagation between a single fixed ground-level node (i.e., on a utility pole) and mobile ground-level client devices. Finally, **C**, concerns infrastructure network configurations where one fixed well-positioned transmitter (AP) is responsible for serving multiple ground-level mobile nodes.

3.1.3.1 Backhaul

The first data set, **A**, was collected using the University of Colorado at Boulder (CU) Wide Area Radio Testbed (WART), which is composed of six 8-element uniform circular phased array antennas [29]. Figure 3.4 shows the layout of this testbed. The devices are mounted on rooftops on the CU campus and in the surrounding city of Boulder, Colorado. These devices can electronically change their antenna pattern, which allows for them to operate as a directional wireless network with a main lobe pointed in one of 16 directions or as an omnidirectional antenna whose gain is (approximately) uniform in the azimuth plane. To collect this data, an “NxN scan” is done of the sort proposed in [41], which results in RSS measurements for every combination of transmitter, receiver, and antenna pattern. In short, this works by having each AP take a turn transmitting in each state while all other nodes listen and log packets. Identical measurements were collected during the winter (no leaves), during a snowstorm, and during the summer of 2010. These network measurements are applicable to rooftop-to-rooftop communication systems, including cell networks, and

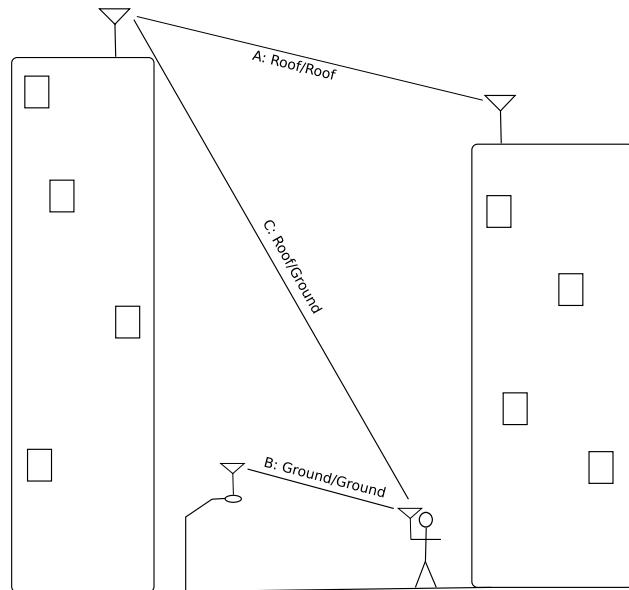


Figure 3.3: Visual schematic of three urban data sets. A: roof to roof measurements from CU WART (Wide Area Radio Testbed), B: ground (utility poles) to ground (mobile node) measurements in Portland, Oregon, C: roof to ground and ground to roof measurements from CU WART.

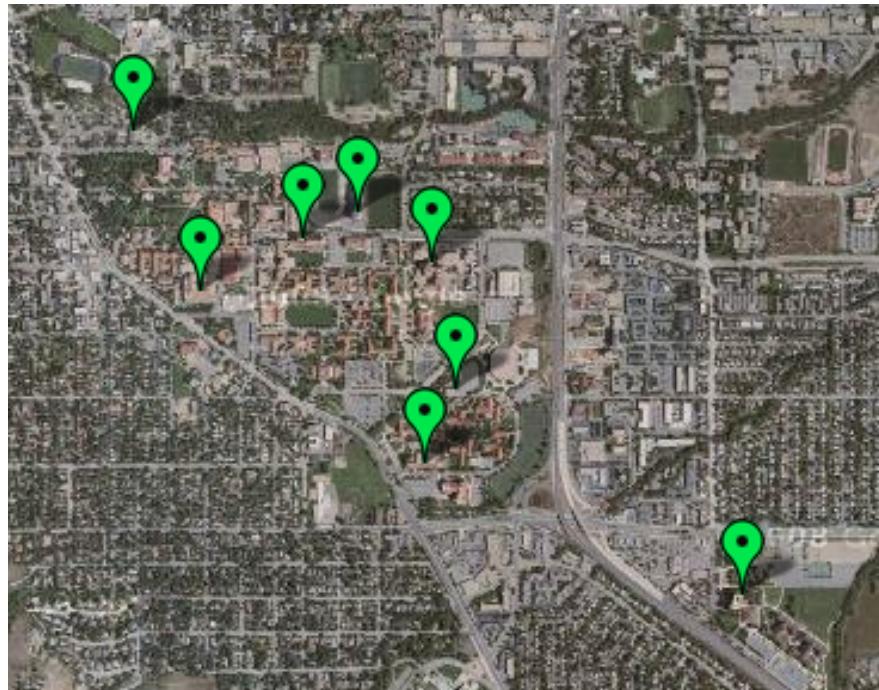


Figure 3.4: University of Colorado Wide Area Radio Testbed (CU-WART)

point-to-point or point-to-multipoint wireless backhaul networks both with directional antennas and with omnidirectional antennas. Although this is a reasonably small network, the representativeness of the environment (a typical urban/suburban campus) and the large number of effective antenna patterns (17^6 unique combinations) that can be tested provide a strong argument for the generalizability of this data.

3.1.3.2 Street Level Infrastructure

The second set of measurements, **B**, involves three data sets from three urban municipal wireless networks: a (now defunct) municipal wireless mesh network in Portland, Oregon, the Google WiFi network in Mountain View, California, and the TFA network in Houston, Texas. All three data sets involve data collected with a mobile client. As a standard practice, the precision of the GPS coordinates is truncated to five significant digits, which has the effect of averaging measurements within a 0.74m (≈ 6 wavelength) circle (a conservative averaging by the standard of [132]).

Portland, Oregon

In this network, 70 APs are deployed on utility poles in a 2 km by 2 km square region. Each AP has a 7.4 dBi omnidirectional antenna that provides local coverage in infrastructure mode. These measurements were collected during the summer of 2007. This data set, which consists of both laborious point testing and extensive war-driving data is most representative of ground-to-ground links in urban environments. The data collection method for this data set is outlined in section 4.2. In short, collection involved a two-stage process. First, a mobile receiver was driven on all publicly accessible streets in the 2 km by 2 km region. The receiver was a Netgear WGT-634u wireless router running OpenWRT linux [17] and the open-source sniffing tool Kismet [15]. The Kismet tool performs channel-hopping to record measurements on all 11 802.11b/g channels which imposes a uniform random sampling (in time) on the observed measurements. The receiver's radio is a Atheros-brand chipset, with an external 5 dBi magnetic roof-mount "rubber duck" antenna and a Universal Serial Bus (USB) GPS receiver. Passive measurements of management frames (beacons) were recorded to a USB compact flash dongle. This results in a large set of measurements that is referred to as "pdx/stumble" here. After this initial stage, 250 additional locations were selected at random from within the region and tested more rigorously with a state-based point tester. At each of these points

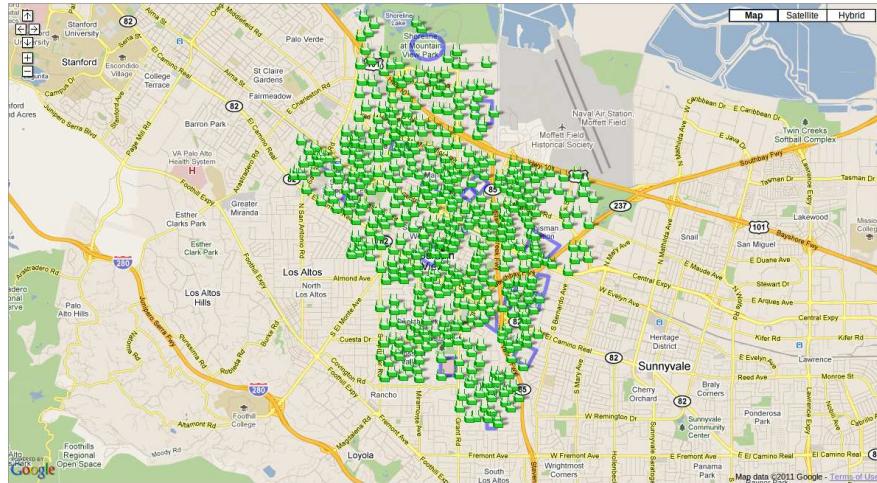


Figure 3.5: Google WiFi Network in Mountain View, California

physical layer information was recorded (i.e., SNR) along with results from higher layer tests. This smaller data set is called “pdx” in the remainder of the thesis and the data collection procedure is described in more detail in section 4.2.1.

Mountain View, California

The Google WiFi network [83], deployed in Mountain View, California covers much of the city (31 km^2) with 447 Tropos-brand [226] 2.4 GHz 802.11 mesh routers. Figure 3.5 provides a basic layout of the network and gives an idea of the extent and density of the deployment. The measurements used here were collected by Robinson *et al.* between October 3rd and 10th in 2007 for their work in [200]. These measurements were made publicly available at [191] and involve passive measurements over a subset of the coverage area (12 km^2) encompassing 168 mesh nodes. These nodes are mounted on light poles as in the Portland measurements and have a 7.4 dBi omnidirectional antenna for local coverage in addition to the backhaul network. The measurements were made with an IBM T42 laptop with a 3 dBi antenna and GPS receiver running the NetStumbler sniffing software [16]. As with the Portland measurements, these are all passive measurements of management frames (beacons) and the sniffer employs channel-hopping to make a uniform random sample (in time) of all 11 channels. The Received Signal Strength Indicator (RSSI) and noise values are recorded for each packet overheard along with a time-stamp and GPS location. Some

minor anonymization of the data has been done to remove unique identifiers (Basic Service Set Identifier (BSSID)s). RSSI is converted to RSS by subtracting 149 from each value [199]. Precise height and transmit power control information was not recorded for this data, so in our application we use the reasonable constant values of 20 dB (100 mW) transmit power (as extracted from Tropos product white-paper specifications) and 12m for the utility pole height.

Houston, Texas

The final set of street level infrastructure measurements comes from the community wireless mesh network constructed by Rice University and the TFA non-profit organization in Houston, Texas [222]. Figure 3.6 shows a heatmap of the measurements. These measurements were collected by Robinson, Camp *et al.* for their work in [44] and [200]. The measurements have been made publicly available at [192]. This network involves 18 wireless nodes in a residential area in Southeast Houston, providing coverage to approximately 3 km^2 and more than 4,000 users. In the data collection, the NetStumbler software was used on a laptop with an a GPS device and Orinoco Gold 802.11b wireless interface (Atheros chipset) connected to a car-roof mounted 7 dBi omnidirectional antenna. As with the other measurements, all data collection is passive and the software channel-hops to record a random sample of overheard management frames (beacons) on each of the 11 channels. The drive-test covers all city streets in the region and was carried out 15 times between the hours of 10am and 6pm between December 15th 2006 and February 15th, 2007. Although this is a winter data collection, Houston has a tropical climate, so it is presumed that the fading due to foilage is constant throughout the year. The measurements contain signal strength, noise, and location values as well as the vehicle's average velocity at the point of measurement.

3.1.3.3 Wide Area Infrastructure

The final data set, **C**, involves two sets of measurements: one carried out at CU of the WART and one set of published measurements from a well-placed transmitter in Munich, Germany.

Boulder, Colorado

The first data set was collected using a mobile node (a Samsung brand “netbook”) with a pair of diversity antennas. In this experiment, the six rooftop CU WART nodes were configured to transmit 80

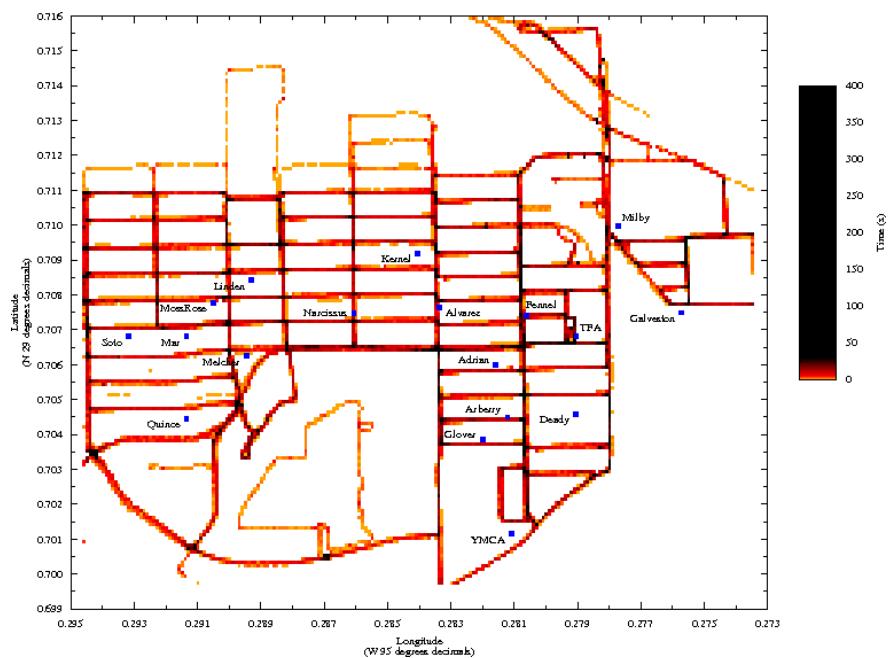


Figure 3.6: TFA-Wireless Network measurements in Houston, Texas

byte “beacon” packets every $0.5 + U(0.0, 0.5)$ seconds, where $U(X, Y)$ is a uniformly distributed random number between X and Y . Beacons are configured to transmit at 1 Mbps, so that possible effects of Doppler spread on higher datarate waveforms are avoided. Similarly, the mobile device was configured to transmit beacons at the same rate. Meanwhile, each rooftop testbed node was configured to its 9 dBi omnidirectional antenna pattern.

All nodes, including the mobile node, were configured to log packets using a second monitor mode (promiscuous) wireless interface. The mobile node was additionally instrumented with a USB GPS receiver that was used both to keep a log of position and to synchronize the system clock so that the wireless trace was in sync with the GPS position log. These measurements were collected during the summer of 2010. During the experiment, the mobile node was attached to an elevated (nonconducting) platform on the front of a bicycle. The bicycle was pedaled around the CU campus on pedestrian paths, streets, and in parking lots. This data set is most representative of an infrastructure wireless networks where a well-positioned static transmitter must serve mobile clients on the ground. This data set is subdivided into the upstream part (“boulder/gtp”) and the downstream part (“boulder/ptg”).

Munich, Germany

The second group of measurements is from a reference data set collected by the COST-231 group at 900 MHz [48] in Munich in 1996. This data set, which provides path loss measurements collected by a mobile receiver from three well-placed (rooftop) transmitters is closest in intent to our data set C, but does not include upstream measurements.

3.2 Implementation Details

Table 2.1 in section 2.2 provides details of the models evaluated in this study. Each of the 30 models is implemented from their respective publications in the ruby programming language. Section C.2 in appendix C provides the source code for the implementations. Only one of the models, the ITM [98], has a reference implementation. Hence, there are fundamental concerns about correctness. To address this basic issue, sanity checking of model output is performed. However, without access to the data sets on which the models were derived, or their reference implementations, it is impossible to make a more rigorous verification than

this.

3.2.1 Terrain Databases

Terrain Models require access to a DEM, and in the case of ITU-R 452, a Landcover Classification Database (LCDB) as well. The DEM used for the networks in the United States is a publicly available raster data set from the United States Geological Survey (USGS) Seamless Map Server, providing 1/3 arcsecond spatial resolution. The US LCDB is also provided by the USGS as a raster data set, which is generated by the USGS using a trained decision tree algorithm. For the New Zealand data sets, DEM and LCDB data are provided by the Environment Waikato organization. The DEM has a vertical precision of 1 m and an estimated accuracy of 5-6 m RMSE. The GDAL library [75] is used to perform coordinate conversions and data extraction to generate path profiles for the terrain algorithms.

3.2.2 Corrections for Hata-Okumura

In the implementation of Hata-Okumura used in this analysis, and its derivative models, a few crude corrections are made to antenna heights in the event that they fall outside of the models' coverage (and would therefore produce anomalous results). First, the minimum of the two heights is subtracted from both so that they are relative. For instance, antenna heights of 30 and 40 m become 0 and 10. Then, heights are swapped if necessary so that the transmitter height is always higher than the receiver height (at this point the receiver height will be zero). Next, one meter is added to the receiver height and subtracted from the transmitter height, keeping the relative difference but setting the receiver height to 1 m. For instance 0 and 10 m would become 1 and 11 m. Finally, the transmitter height is decreased or increased as necessary so that it is above the minimum (30 m) and below the maximum (200 m) permissible values for the Hata-Okumura model.

These corrections are necessary to use the Hata-Okumura model with transmitter or receiver heights that would otherwise produce meaningless (infinite) results. It is not certain what the impact is on the model performance by making these corrections. However, it stands to reason that even if the performance is negatively impacted, an inaccurate prediction will still be closer to the true answer than an anomalous

(infinite) prediction.

3.3 Method

The approach taken is to ask each model to offer a prediction of median path loss for each link in the data. The model is fed whatever information it requires, including DEM and LCDB information. The model produces an estimate of the loss \hat{L} that is combined with known values to calculate the predicted received signal strength P_r :

$$P_r = P_t + G_t(\theta) + G_r(\phi) - \hat{L} \quad (3.1)$$

Where G_t is the antenna gain of the transmitter in the azimuthal direction (θ) of the receiver and G_r is the antenna gain of the receiver in the azimuthal direction (ϕ) of the transmitter. These gains are drawn from measured antenna patterns. The antenna patterns were derived for each antenna empirically, using the procedure described in appendix A. The transmit power (P_t) is set to 18 dBm for all nodes, which is the maximum transmit power of the Atheros radios that all measurement nodes use. For a given link, the median received signal strength value is calculated across all measurements (\bar{P}_r). Then, the prediction error, ϵ , is the difference between this prediction and the median measured value:

$$\epsilon = \bar{P}_r - P_r \quad (3.2)$$

Some models come with tunable parameters of varying esotericism. For these models, a range of reasonable parameter values are tried without bias towards those expected to perform best.

This entire process requires a substantial amount of computation, but is trivially parallelizable. To make the computation of results tractable, the task of prediction is subdivided into a large number of simultaneously executing threads and the results are merged upon completion. Figure 3.7 shows a schematic of the process. Parallel computation must occur in two sequential stages. During the first stage, path profile information is extracted and prepared for each link in parallel, and during the second stage this information is fed to each algorithm for each link, which can also be done in parallel. With the merged data in hand, each

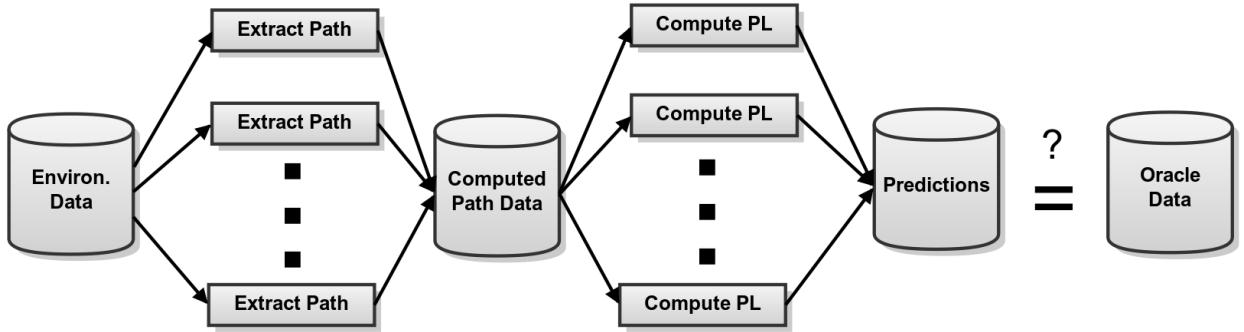


Figure 3.7: Schematic showing the process of parallel computation of path loss predictions using many models and many links.

prediction is compared with an oracle value for the link. This oracle value is computed from the measured received signal strength for the link as well as known values for the transmitter power and antenna gain.

It is worth noting that among the models studied, only very few were designed with the exact sort of networks studied here in mind. Indeed, some are very specific about the type of environment in which they are to be used. In this study both appropriate and “inappropriate” models are given an equal chance at making predictions for our network—there is no starting bias about which should perform best.

The next section describes the process of explicitly fitting the data to a theoretical model and looking at the number of measurements required for a fit. This gives an initial estimate of expected error for direct (naïve) fits to the collected data. Then, to analyze the performance of the algorithms, five domain-appropriate metrics of decreasing stringency are proposed. The performance of each model with respect to these metrics, as well as general trends and possible sources of systematic error, are described in section 3.6. Finally, in an attempt to put a lower bound on model error, explicit parameter fitting of the best models is performed and this best-case performance is compared to the naïve approach of straight line fitting.

3.4 Simple Log-Domain Data Fitting

Consider equation 2.8 in section 2, which describes the fundamental power law relationship between path loss and distance. It is common in the literature to show this relationship as a straight line on a log/log plot. When this equation is modified to have a flexible exponent and error term, it is possible to do a linear

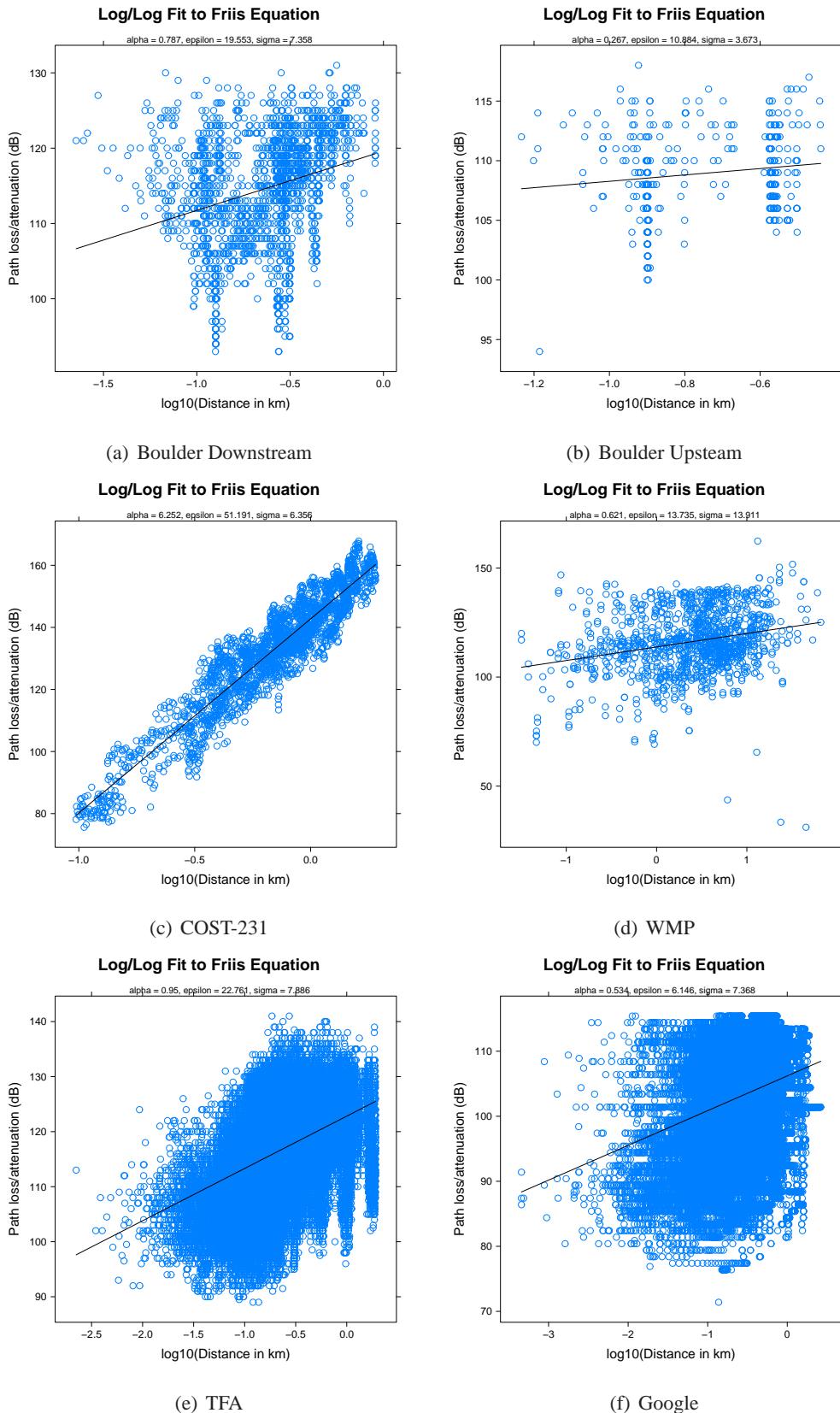


Figure 3.8: Explicit power law fits to Data. Fit parameters are provided on the plots.

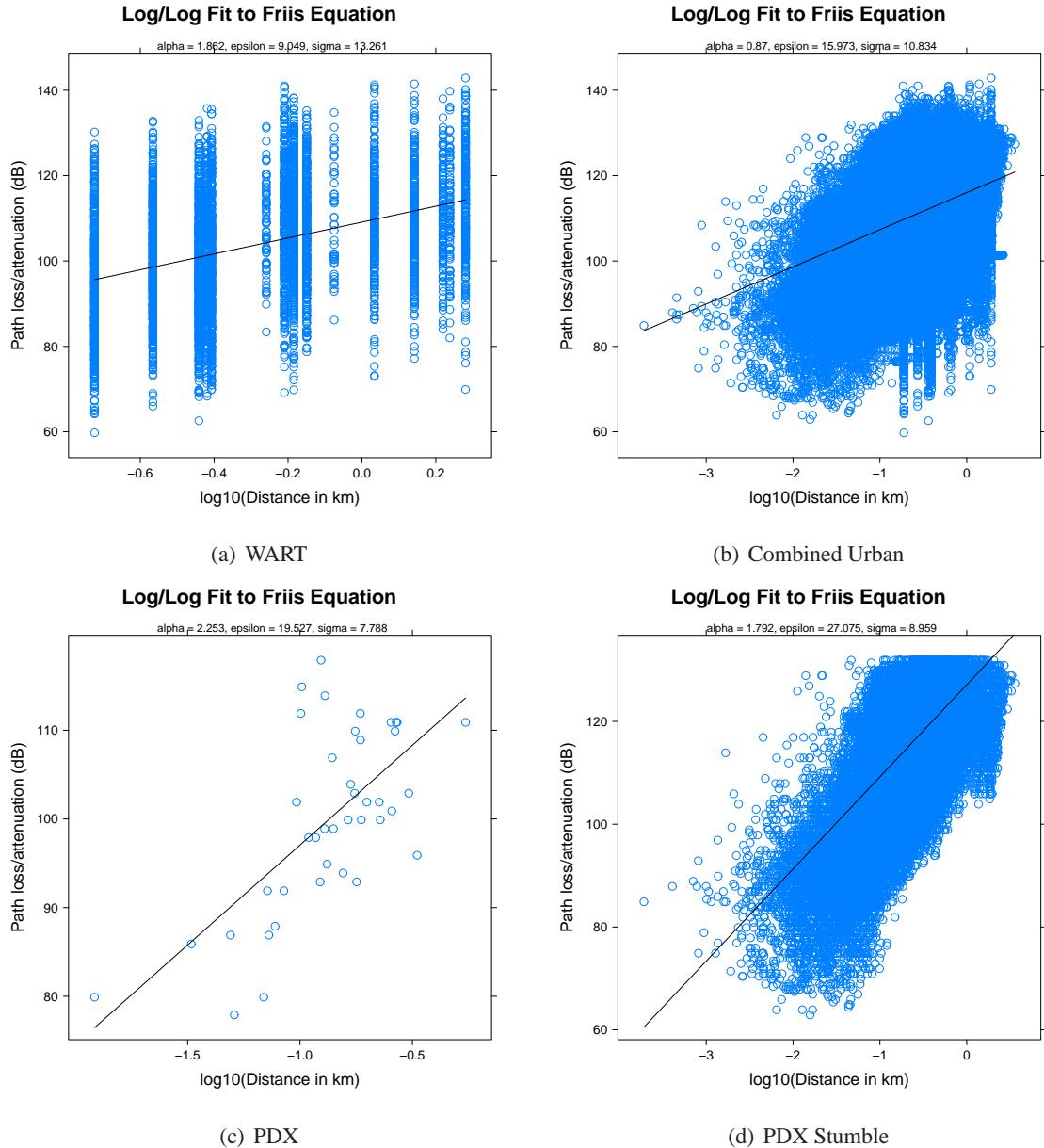
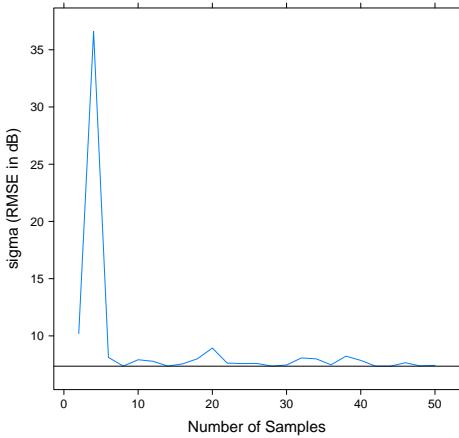
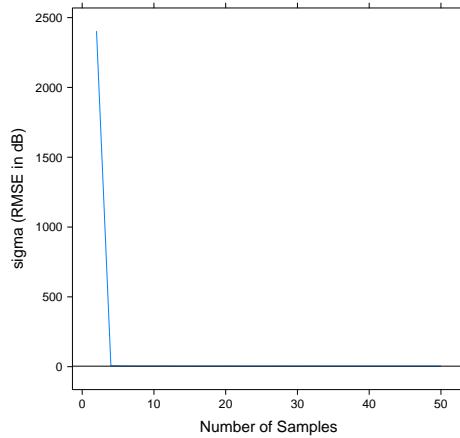


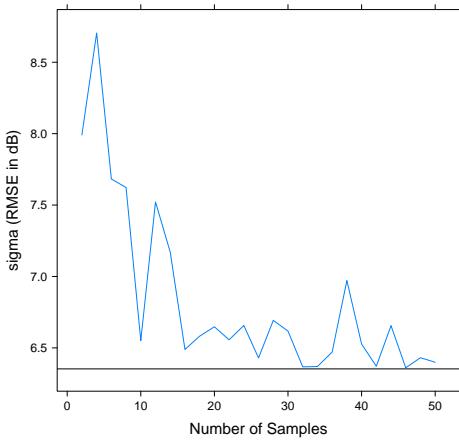
Figure 3.9: Explicit power law fits to Data. Fit parameters are provided on the plots.

Error of Fitted Model for Increasing Sample Size

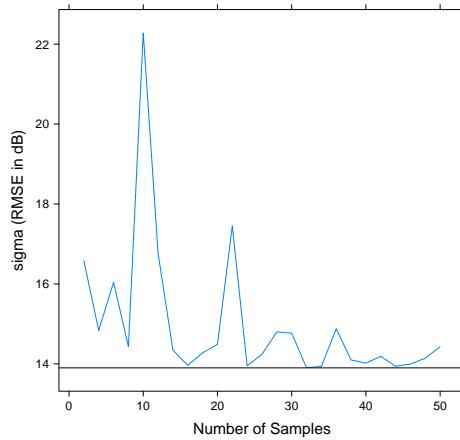
(a) Boulder Downstream

Error of Fitted Model for Increasing Sample Size

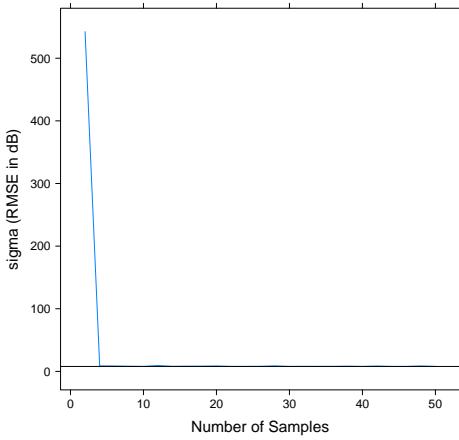
(b) Boulder Upstream

Error of Fitted Model for Increasing Sample Size

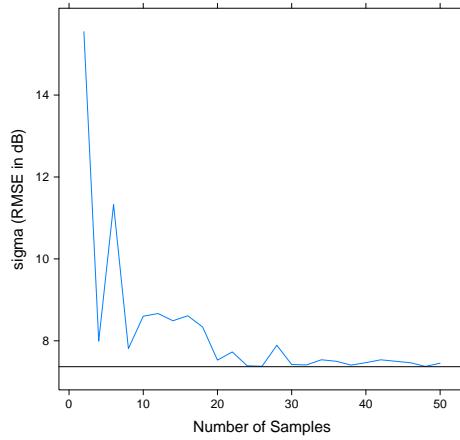
(c) COST-231

Error of Fitted Model for Increasing Sample Size

(d) WMP

Error of Fitted Model for Increasing Sample Size

(e) TFA

Error of Fitted Model for Increasing Sample Size

(f) Google

Figure 3.10: Number of samples required for naïve fit. Plots show fit standard error for fits increasing random samples and a horizontal line is given at the RMSE obtained for all points.

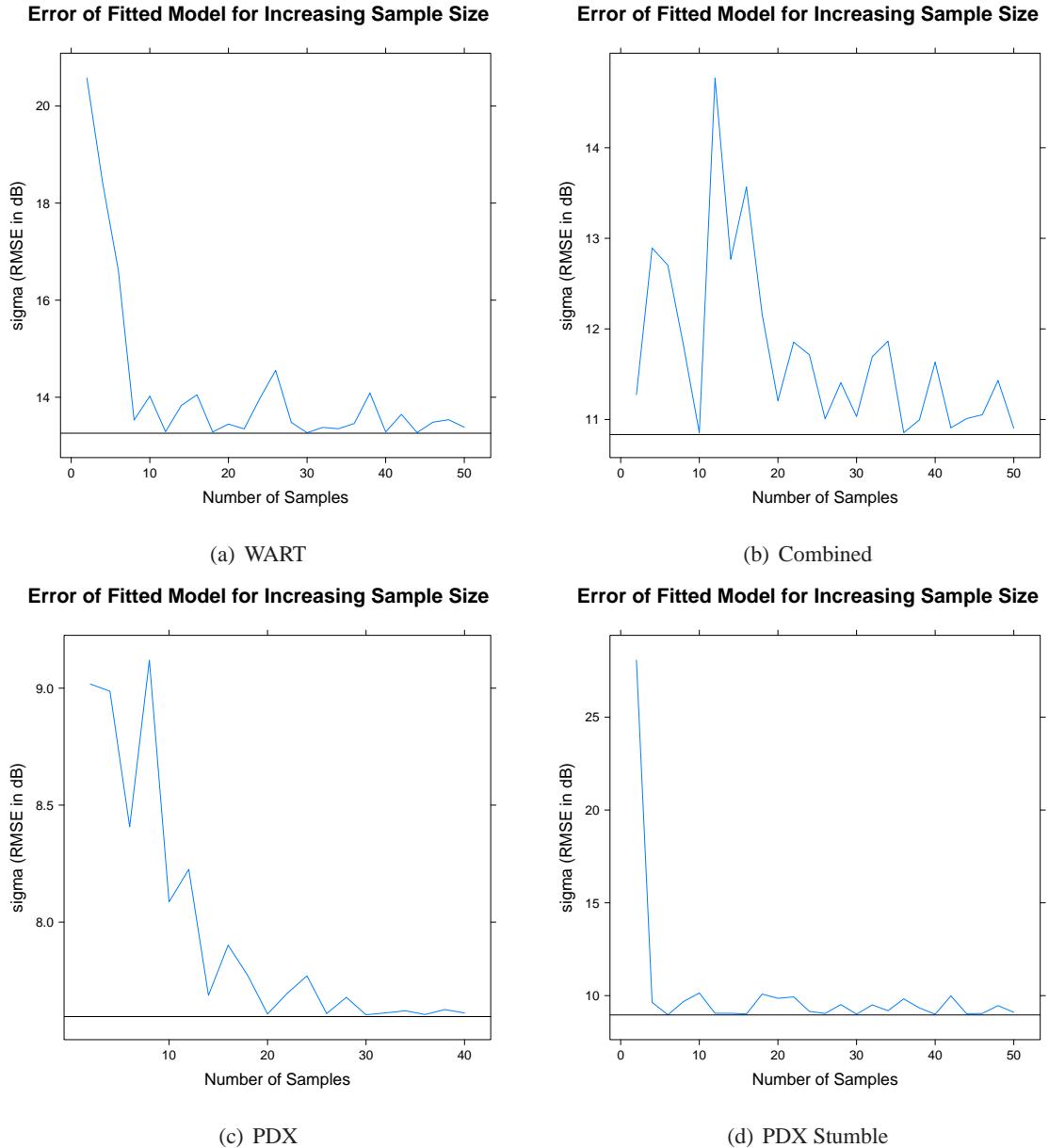


Figure 3.11: Number of samples required for naïve fit. Plots show fit standard error for fits increasing random samples and a horizontal line is given at the RMSE obtained for all points.

fit in the log/log domain and come up with empirical estimates of the exponent (α) and offset (ϵ):

$$P_r = P_t - (\alpha 10 \log_{10}(d) + 20 \log_{10}(f) + 32.45 + \epsilon) \quad (3.3)$$

Figures 3.8 and 3.9 show the resulting fits using this method for each data set and one superset that includes the combination of all urban measurements. One unavoidable side effect of packet-based measurements is that it is impossible to record SNR values for packets that fail to demodulate. Hence, because the 2.4 and 5.8 GHz data is derived from packet-based measurements, low SNR values (and therefore high path loss values) are underrepresented here, which leads to “shallow” fits and unrealistically low values of α . As a result, while it is safe to make comparisons between the 2.4/5.8 GHz data sets, it is not safe to directly compare the slope of the 900 MHz and 2.4/5.8 GHz fits.

Fits are computed using linear least square regression. Table 3.2 lists fitted parameters (α, ϵ) and residual standard error (σ)². Between the 2.4 GHz data sets, there is little consensus about the slope or intercept of this power-law relationship, except that it should be in the neighborhood of $\alpha \approx 2$ and $\epsilon \approx 15$. All fits are noisy, with standard error around 8.68 dB on average for the urban data sets. This residual error tends to be Gaussian, which is also in agreement with previously published measurements (e.g., [183]). However, the size of this error is almost two orders of magnitude from the 3 dB that Rizk *et al.* suggest as an expected repeated measures variance for outdoor urban environments (and hence the expected magnitude of the error due to temporally varying fast fading) [198]. Looking at figure 3.8, it is easy to see that the 2.4 GHz measurements are substantially less well behaved than the COST-231 data, even in comparable environments.

In order to understand how many measurements are needed to create a fit of this sort, successively increasing random samples of the data sets are taken to generate a fit. The rate that residual error of the model (with respect to the complete data set) converges as the subsample size increases can be studied from these results. Figures 3.10 and 3.11 show this plot for each data set. All plots follow a similar trend: the eventual model is closely matched with approximately 20, or at most 40, data points. Table 3.2 gives an approximate minimum sample size for each data set in the column labeled N derived from these plots.

² For all intents and purposes, standard error (σ) and RMSE are interchangeable.

Name	α	ϵ	σ	N	Top Three Performing Models by SC-RMSE						Ideal RMSE
wart	1.86	9.05	13.26	15	flatedge	13.73	itu.terrain	13.89	hatao	14.03	1.96
wart/snow	1.92	9.25	13.36	15	itu.terrain	13.93	flatedge	14.16	hatao	14.19	1.87
pdx	2.25	19.53	7.8	5	allsebrook200	8.38	hatal	8.97	davidsons	9.37	1.14
pdx stumble	1.79	27.08	8.96	40	allsebrook400	8.34	itur25	10.50	hatam	10.51	1.02
boulder/ptg	0.79	19.56	7.36	20	allsebrook400	7.90	ecc33m	9.38	hatam	10.47	0.94
boulder/gtp	0.27	10.88	3.67	5	allsebrook400	5.45	hatal.fc	7.15	edwards200	8.51	1.01
cost231	6.25	51.19	6.36	15	edwards200	9.23	hatam	9.99	itur25	10.55	1.23
wmp	0.62	13.74	13.92	15	flatedge	15.34	alsebrook200	16.72	egli	16.83	5.98
tfa	0.95	22.76	7.89	20	herring.atg	8.90	allsebrook200	9.03	flatedge	10.83	1.43
google	0.54	6.15	7.37	30	davidsons	13.56	itu.terrain	16.12	hatal	16.83	2.93

Table 3.2: Summary of results by data set

3.5 Performance Metrics for Path Loss Prediction

The performance of the models is analyzed with respect to several metrics in order of decreasing stringency:

- (1) RMSE and Spread-Corrected Root Mean Square Error (SC-RMSE)
- (2) Competitive Success
- (3) Individual Accuracy Relative to Spread
- (4) Skewness
- (5) Rank Correlation

3.5.1 RMSE and SC-RMSE

RMSE is the most obvious and straightforward metric for analyzing the error of a predictive model of this sort. As discussed above, for a given model we compute an error value (ϵ as in equation 3.2) for each prediction of each link in each data set. For a given set of links l , in a given data set D and a given model m , the overall RMSE for a given model for a given data set is:

$$RMSE_{m,D} = \sqrt{\frac{\sum_{l \in D} \epsilon_{m,l}^2}{|D|}} \quad (3.4)$$

where $\epsilon_{m,l}$ is the error of model m for link l and $|D|$ is the number of links in the data set D . SC-RMSE is a version of RMSE that subtracts off the expected spread in the measurements from the RMSE. This way, if a given link has large variation in the measurements, then the error a model obtains on that link is reduced by a proportional amount. This has the effect of reducing the error associated with especially noisy links. Figure 3.12 provides an explanatory diagram comparing normal error (ϵ) and spread-corrected error (ϵ'). The spread-corrected error for a given model m and link l is the absolute value of the error, reduced by the standard deviation (σ_l) of measurements on link l :

$$\epsilon'_{m,l} = |\epsilon_{m,l}| - \sigma_l \quad (3.5)$$

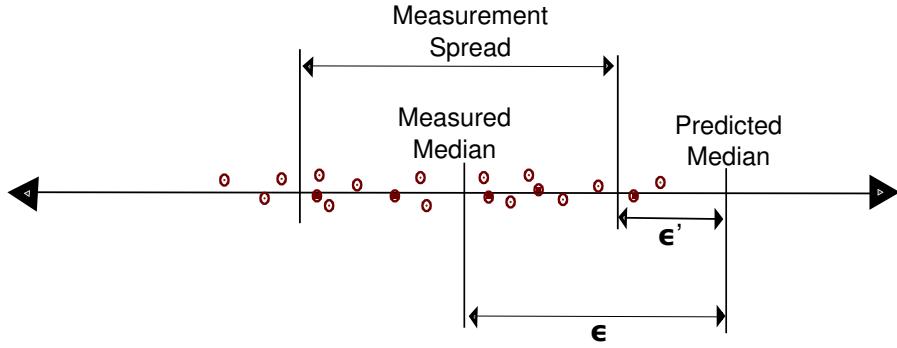


Figure 3.12: Schematic explaining error (ϵ) and spread-corrected error (ϵ') in terms of measurement spread and measured and predicted median values.

Computing SC-RMSE is identical to RMSE as shown in equation 3.4, except ϵ' is substituted for ϵ .

3.5.2 Competitive Success

The competitive success metric is the percentage of links in a given data set that a given model has made the best prediction for. For each link, the model that makes the prediction with the smallest ϵ is recorded. The percentage is computed by counting the number of best predictions for each model and dividing by the total number of links:

$$CS_{m,D} = 100 \frac{N_{best,m,D}}{|D|} \quad (3.6)$$

When analyzing many models, if one model (or a set of related models) is dominant for a given environment then it would score near 100 on this metric. Because the percentage points are divided evenly between all models tested, if a large number of models are tested, this metric may become spread too thinly to be useful for analysis (i.e., too many similar models share the winnings and no single model comes out on top).

3.5.3 Individual Accuracy Relative to Spread

The individual accuracy metric is the percentage of links where the given model is able to make a prediction within one or two standard deviations of the measured spread:

$$IA_{m,D} = 100 \frac{\sum_{l \in D} \begin{cases} 1 & |\epsilon_{m,l}| < k\sigma_l \\ 0 & o.w. \end{cases}}{|D|}; k = 1, 2, \dots \quad (3.7)$$

where k is how many standard deviations to use for the metric. In the following analysis, results for $k = 1$ and $k = 2$ are used.

3.5.4 Skewness

The fourth metric is skewness, which is simply the sum of model error across all links:

$$S_{m,D} = \sum_{l \in D} \epsilon_{m,l} \quad (3.8)$$

This metric highlights those models that systematically over- and underpredict. Some applications may have a particular cost/benefit for under or overpredictions. Models that systematically overpredict path loss (and therefore underpredict received signal strength) score a high value on this metric. Models that systematically underpredict score a large negative value. Models that make an equal amount of under- and overpredictions will score a value of zero.

3.5.5 Rank Correlation

The final metric is rank correlation using Spearman's ρ^3 . In some applications, predicting an accurate median path loss value might not be necessary so long as a model is able to put links in a correct order from best to worst (consider, for instance, the application of dynamic routing). Spearman's ρ is a nonparametric measure of statistical dependence and in this application describes the relationship between ranked predictions and oracle values using a value between -1.0 (strong negative correlation) and 1.0 (strong positive correlation).

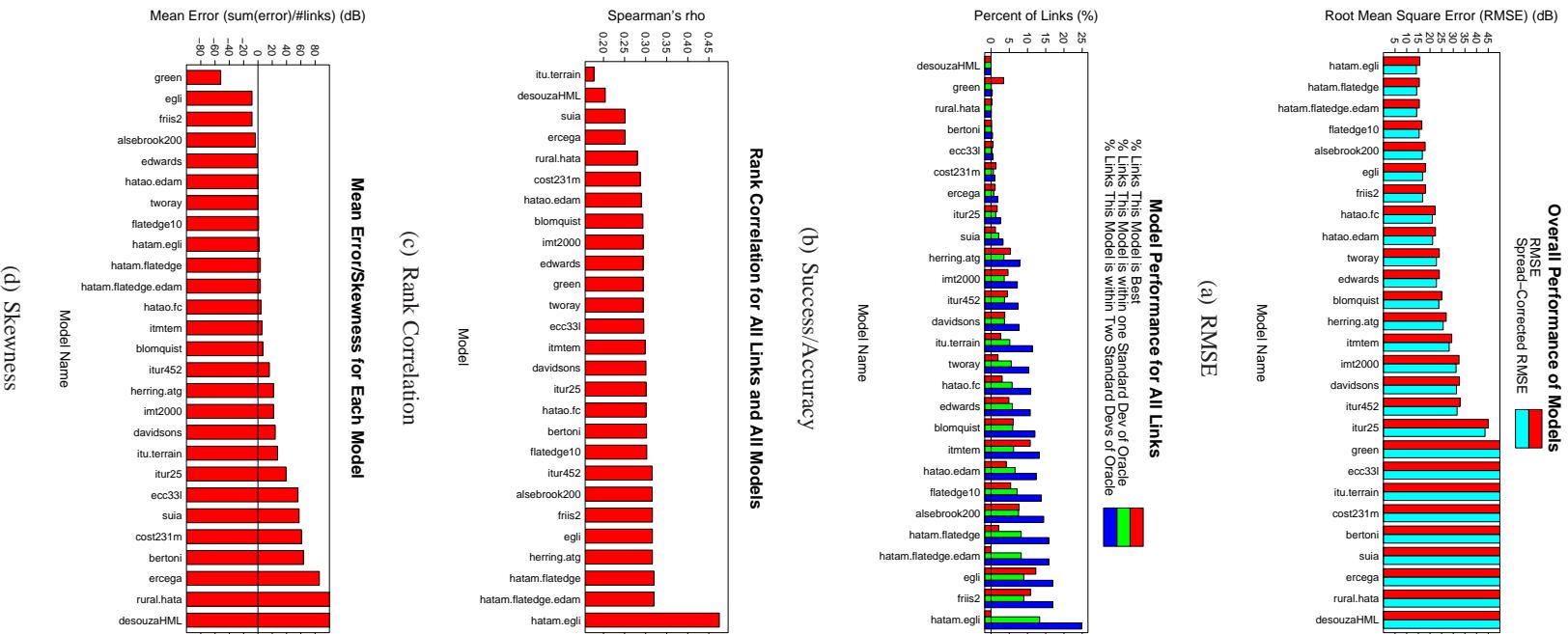


Figure 3.13: Five metric results for WMP data set

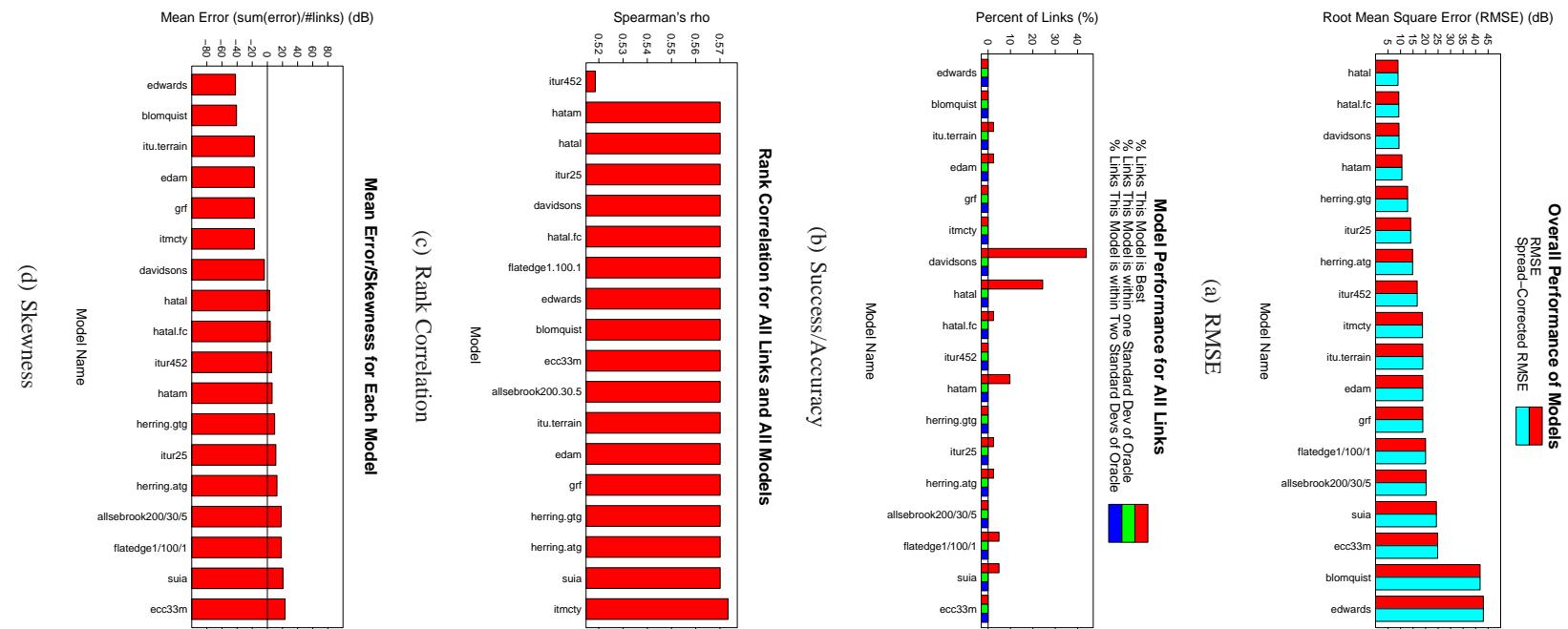


Figure 3.14: Five metric results for PDX data set

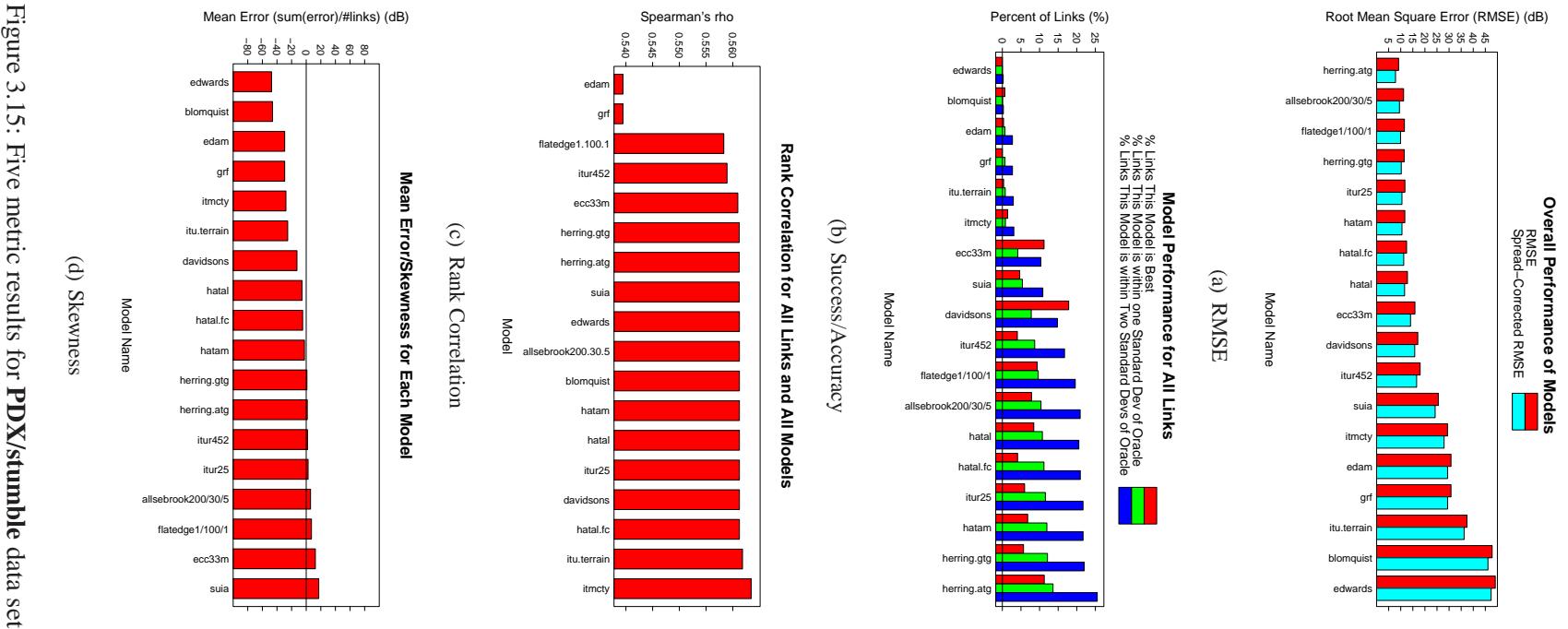


Figure 3.15: Five metric results for PDX/stumble data set

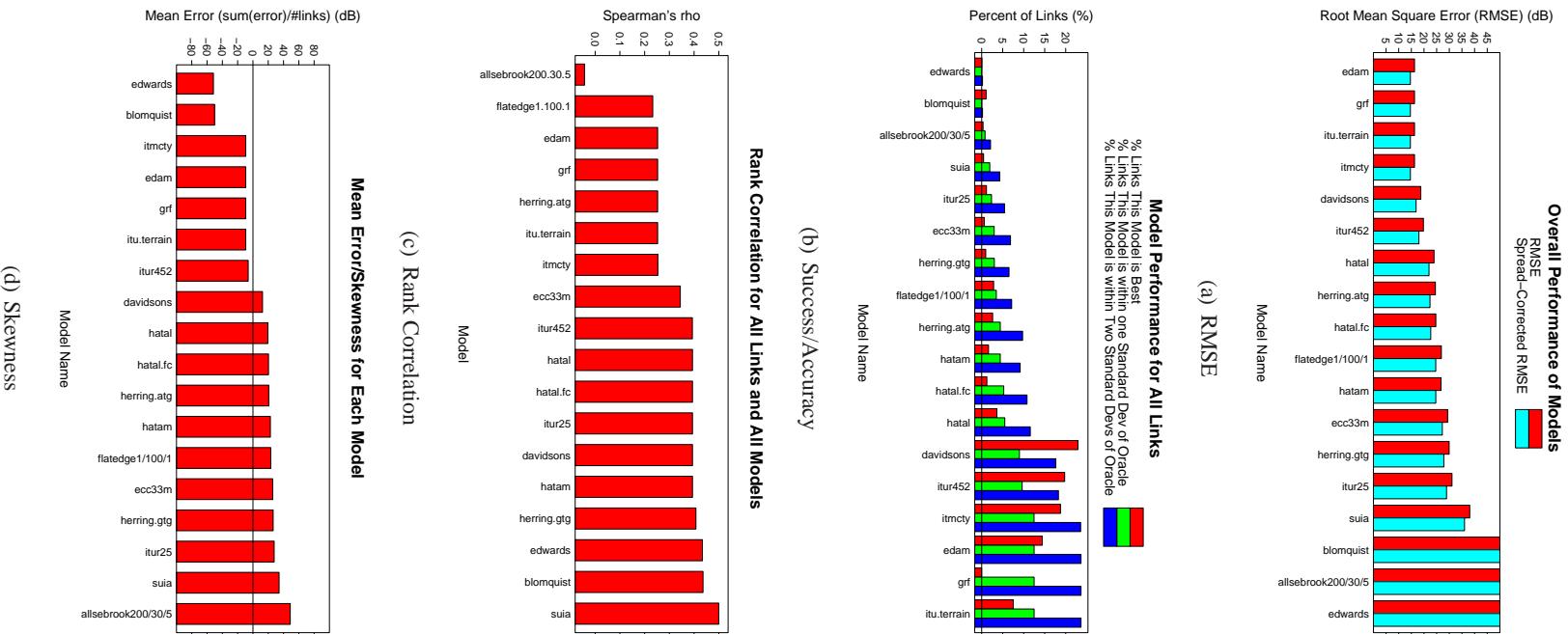


Figure 3.16: Five metric results for WART data set

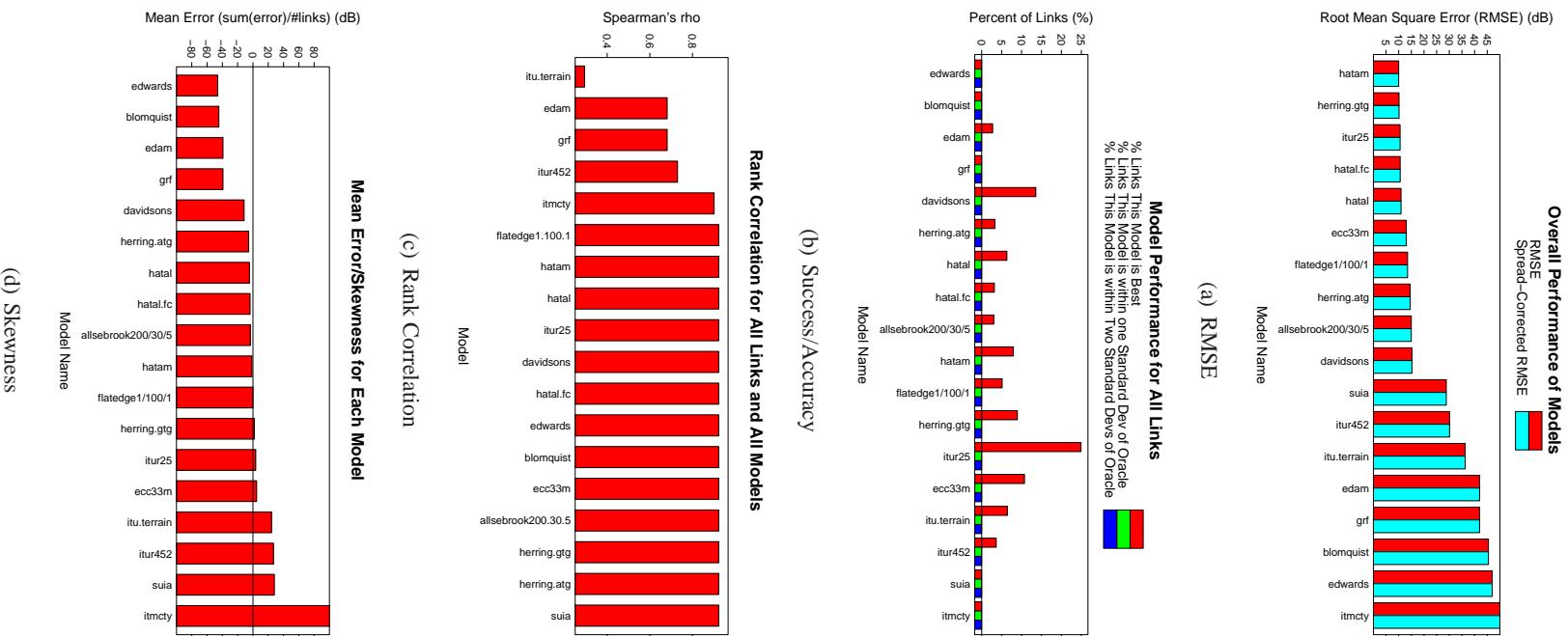


Figure 3.17: Five metric results for COST231 data set

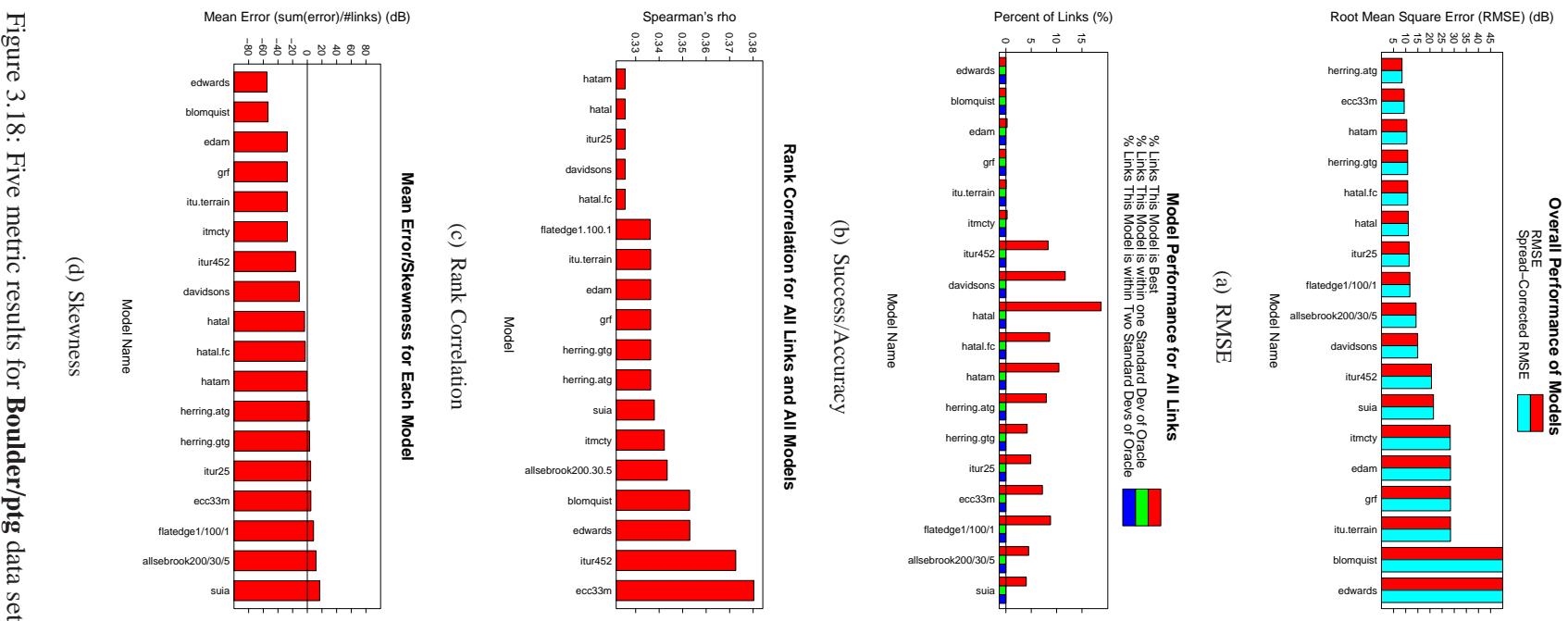


Figure 3.18: Five metric results for Boulder/ptg data set

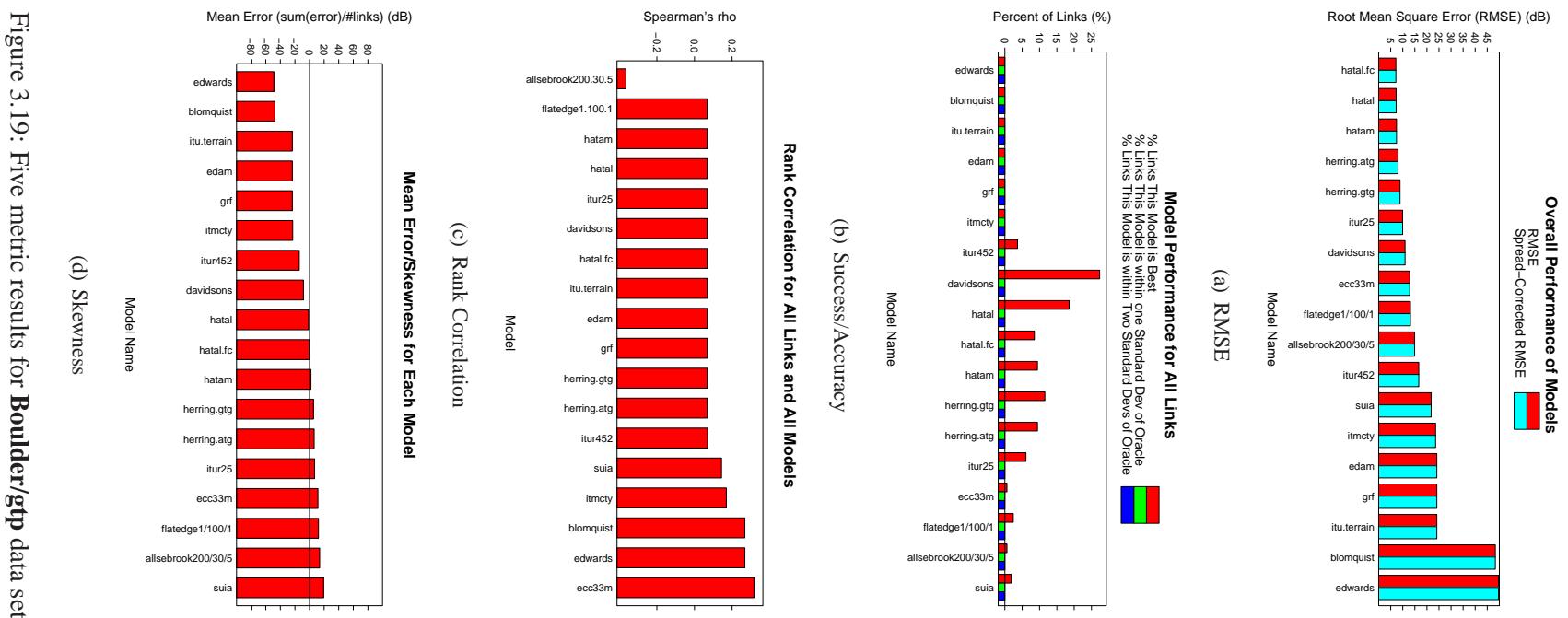


Figure 3.19: Five metric results for Boulder/gtp data set

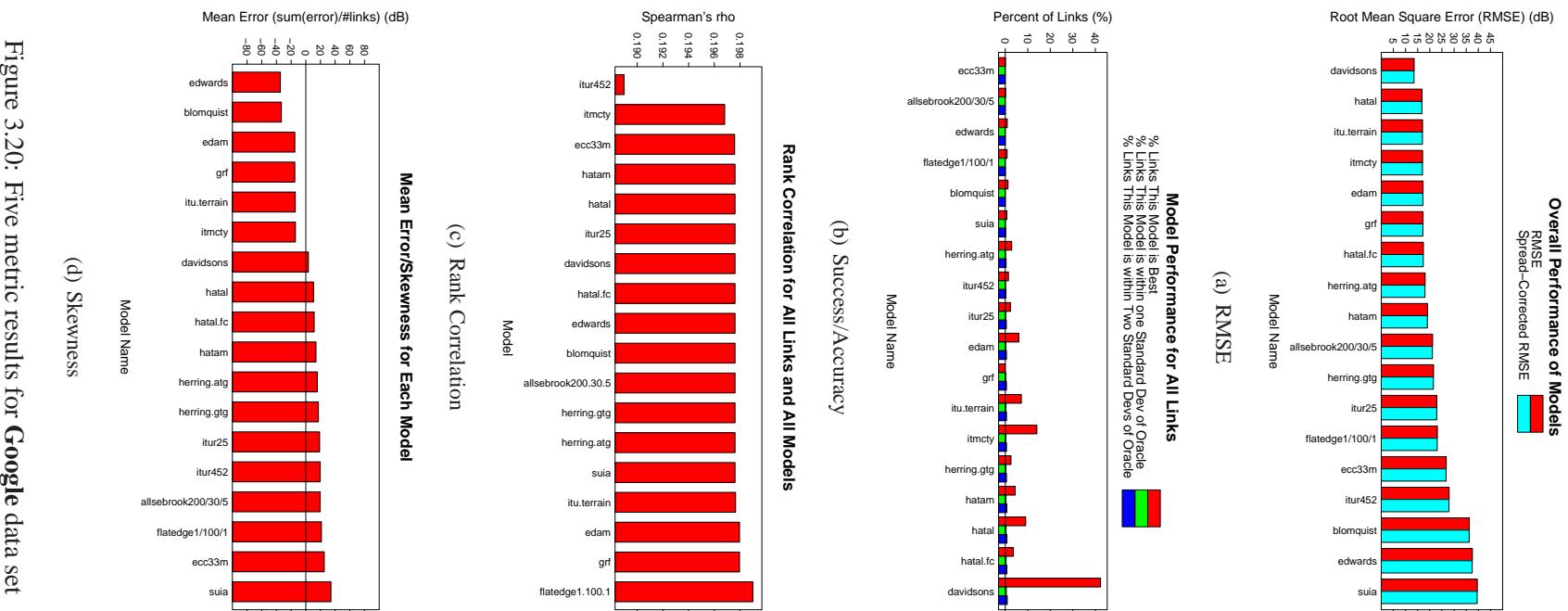


Figure 3.20: Five metric results for Google data set

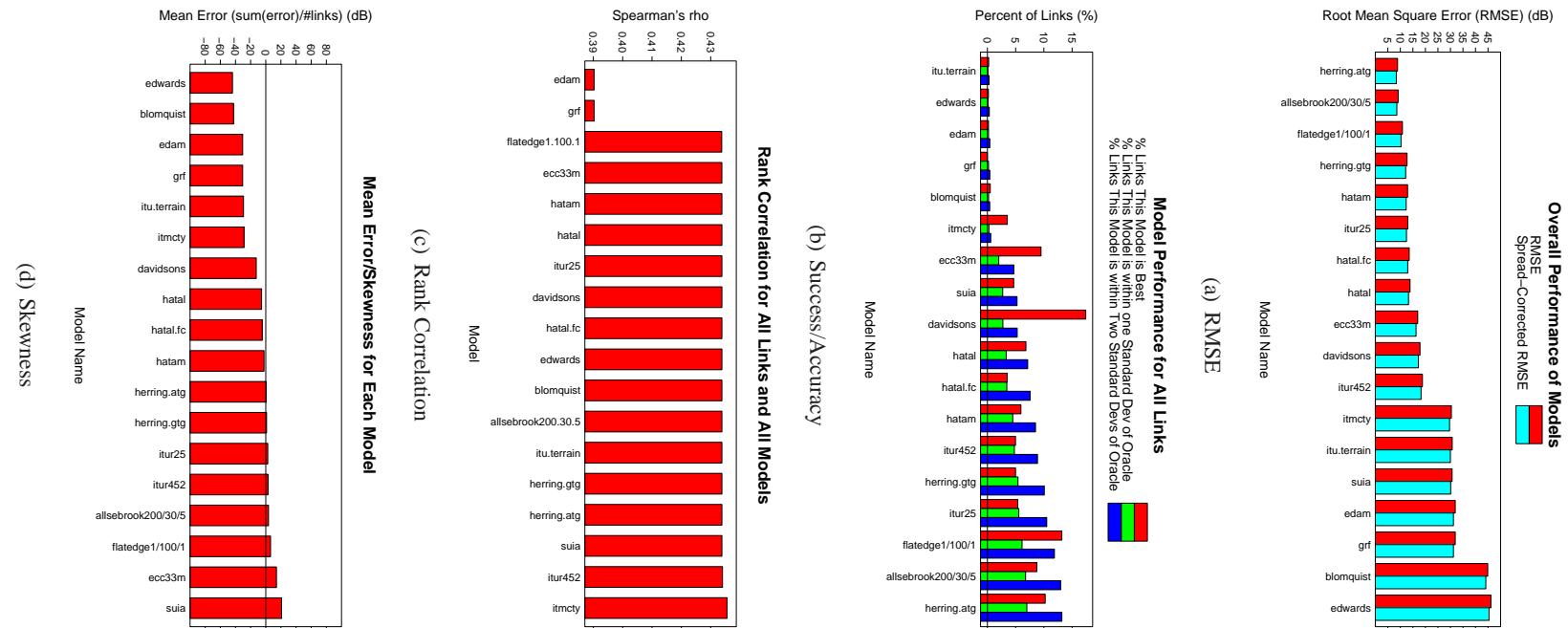


Figure 3.2.1: Five metric results for TFA data set

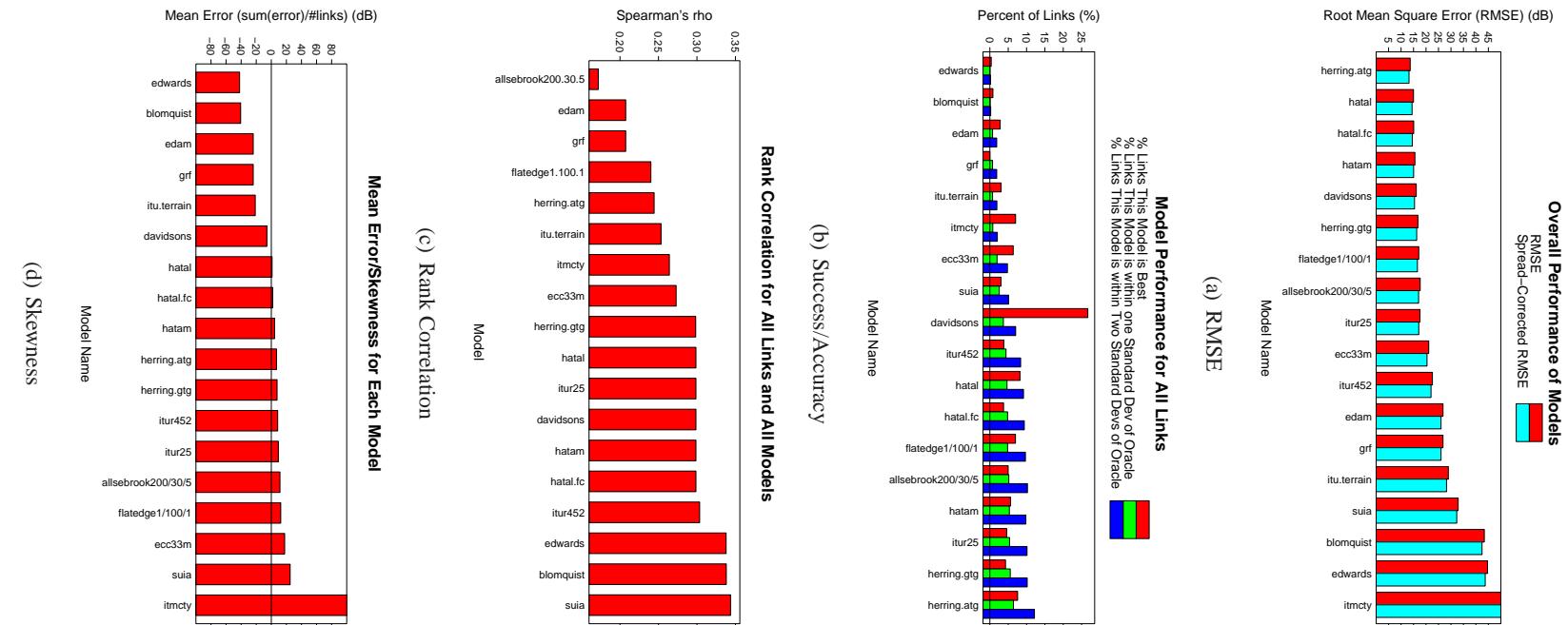


Figure 3.22: Five metric results for all Urban data sets

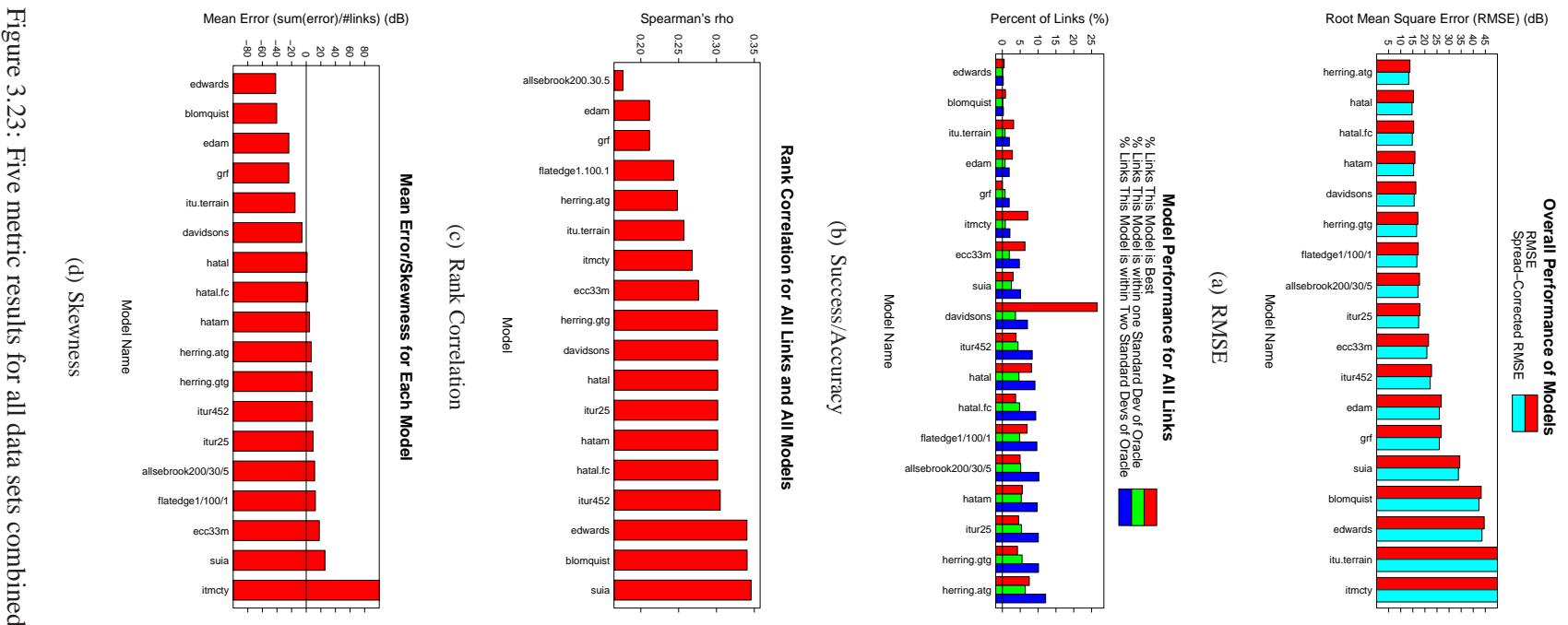


Figure 3.23: Five metric results for all data sets combined

3.6 Results

Figures 3.13-3.23 show the results of these metrics for each data set and all (urban) data sets combined. To simplify the plots, only results from the 18 best-performing models (30 for the rural data) are included. Because the urban and rural data sets were best modeled by different algorithms, a slightly different set of models is shown for each of these. However, the urban data sets present results from the same subset of models so that results are easily comparable.

Looking first at the results for the rural (WMP) data, the best-performing models achieve an RMSE on the order of 15 dB. The best models are the Alsebrook model (with its terrain roughness parameter set to 200m) at just under 18 dB RMSE (16.7 dB when corrected), and the Flat-Edge model (with 10 “buildings” presumed) at 16.5 dB RMSE (15.3 dB when corrected). In the urban data sets, the urban models do much better in terms of RMSE. The best models achieve an RMSE on the order of 10 dB, and the worst (of the best) approach more than 50 dB. The overall winners are the Hata model, the Allsebrook-Parsons model, the Flat-Edge model, and the ITU-R model. This follows from expectations because all of these models were derived for predicting path loss in urban environments. The Hata model and Allsebrook-Parsons models are based on measurements from Japanese and British cities respectively. The Flat-Edge model is a purely theoretical model based on the Walfisch-Bertoni model, which computes loss due to diffraction over a set of uniform screens (simulating buildings separated by streets). Table 3.2 provides the top three models by SC-RMSE for each data set and their corresponding error.

The second metric, competitive success, is shown with the leftmost (red) bar in the second of each set of plots. For most of the data sets, there is no clear winner, with the best models sharing between 10 and 15 percent of the winnings. This indicates that there is no single model that outperforms all others. However, there are a few exceptions. For the PDX data set, the Davidsons model takes 40% of the winnings. In the COST-231 data set, the ITU-R 25 model takes 30%. In the Google data set, the Davidson’s model takes more than 30%. And, in the downstream Boulder measurements (boulder/gtp), the Davidon’s model again takes 25% of the winnings. There are not, however, one or two models that outperform all others in a large subset

³ Kendall’s τ would be an equally appropriate metric, but is slower to compute.

of the data. Hence, we can conclude that the choice of the most-winning model is environment-dependent.

The third metric is percentage of predictions within one (or two) standard deviations of the true median value. This metric requires multiple measurements at each point in order to estimate temporal variation in the channel. Of the data sets, six have this data available: WMP, COST-231, PDX/Stumble, Google, TFA, and WART. For the WMP data the best-performing models (Allsebrook-Parsons, Flat Edge, Herring Air-to-Ground, and ITU-R) score between 10% (for within one standard deviation) and 20% (for within two standard deviations) on this metric. Similar results can be seen for our other data sets, but with different winners. For the PDX/Stumble data, the winners are Herring Air-to-Ground, Hata, and ITU-R 25. For the WART data set, the winners are the ITM, ITU-Terrain, and Blomquist. For the COST-231 data set, the winners are Herring Air-to-Ground, Hata, and Allsebrook-Parsons. Again, the best-performing model appears to be largely environment-dependent.

The fourth metric is skewness. The interpretation of this metric is largely application-dependent, i.e., it is hard to know in advance whether over- or underestimates are more harmful. If a model makes an equal amount of over- or underestimates (resulting in zero skewness), but has a large RMSE, is it better than a model that systematically overestimates but has a small RMSE? The Hata model is particularly well behaved by this metric, producing a value near zero for all data sets. As one would expect, the Hata-derived models perform similarly (i.e., ITU-R 25, Davidsons, etc.). The rest of the models seem to vary largely from data set to data set, although ITU-R 452 performs well for some data sets.

The final metric is rank correlation. For just about all of the models a rank correlation around 0.5 is observed, which indicates a moderate (but not strong) correlation between measured and predicted rank orderings. Models that perform particularly poorly by this metric achieve values much lower on occasion. A result near zero indicates that there is no noticeable correlation between rank orderings. The COST-231 rank correlations are substantially higher than all other data sets. This may be related to the fact that the COST-231 data more closely fits theoretical expectations of the relationship of path loss to distance. Hence, models that use something like Friis equation at their core will produce rank values that are closer to data in this data set. Overall, however, there does not seem to be a consensus about which model performs best at rank ordering—the winners are different for each data set.

3.6.1 Explicit Parameter Fitting

In order to get an idea of minimum obtainable error with these models, two well-performing models that have tunable parameters are used: Allsebrook-Parsons and Flat-Edge. The experiment proceeds by searching the parameter space to find the best-possible parameter configuration for each⁴. The Allsebrook-Parsons model takes three parameters (besides carrier frequency, which is common to nearly all the models): Δh , a terrain roughness parameter (in m), h_0 , the average height of buildings (in m), and d_2 , the average width of streets (in m). The Flat-Edge model also takes three parameters: n , the number of buildings between the transmitter and receiver, h_0 , the average height of these buildings (in m), and w , the street width (in m). After sweeping the parameter space, an Analysis of Variance (ANOVA) is used to determine the parameters that best explain the variance in the data.

For the Allsebrook-Parsons model, the Δh and h_2 parameters are both important. For the Flat-Edge model, h_0 is the only significant parameter. Figure 3.24 shows the response (in terms of RMSE) for tuning these parameters. The optimal values can be determined from the minima of these plots and a similar approach could be carried out with any subset of the data. However, *the optimal parameters for one datum are not usually in agreement with others, forcing a compromise in terms of accuracy and specificity*. Even with cherry-picked parameters, the RMSE is still in the neighborhood of 9-12 dB, which is too large for most applications.

If 9 dB is considered to be the minimum achievable error of a *well-tuned* model, it is interesting to note that approximately the same performance can be achieved with a straight line fit through a small number (≈ 20) of measurements as was shown in section 3.4. In [73], the authors found similar bounds on error (6-10 dB) attempting to fit a single model to substantial measurement data at 1900 MHz. If the domain of interest is network planning, and it is not possible to make measurements of a network (because it does not yet exist), then tuning an *a priori* model may be the right approach to take. However, if the goal is modeling the path loss of a network that can be directly studied, and taking 20 (randomly distributed) measurements is reasonably cheap, then this approach seems easy to advocate by comparison.

⁴ Data from the Boulder, WART, and PDX data sets were used for this experiment.

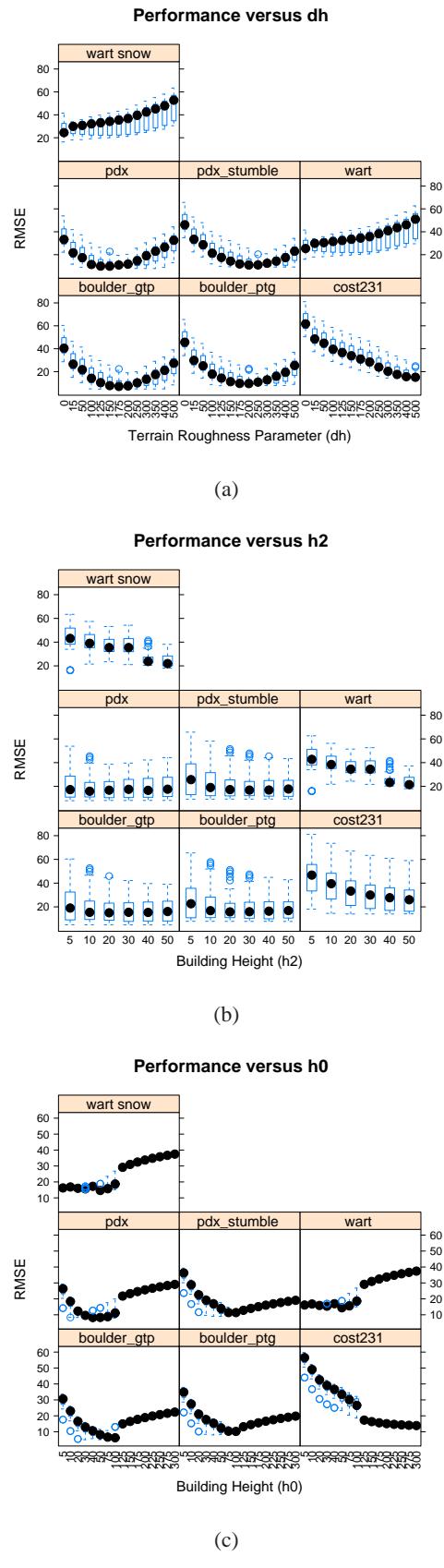


Figure 3.24: Explicit parameter fitting for the Allsebrook-Parsons and Flat-Edge model parameters.

3.6.2 Factors Correlated with Error

Overall, these results are not terribly impressive. Even in the mean case, the best models with their best parameter settings cannot achieve an error of less than 15 dB for the rural measurements and approximately 9 dB for the urban data sets—three to five orders of magnitude from the correct value. More permissive performance metrics show the models are unable to widely succeed at seemingly simple tasks of rank-ordering links, or making predictions within two standard deviations of the measured value. This begs the question: is there some common source of error that is affecting all models?

In order to understand which variables may serve to explain model error, a factorial ANOVA was performed using spread-corrected error as the fitted value and transmitter height, receiver height, distance, line-of-sight (a boolean value based on path elevation profile), and data set. Although all of these variables show moderate correlations (which speaks to the fact that many models add corrections based on these variables), some are much better explanations of variance than others. Perhaps not surprisingly, distance and data set name are the biggest winners with extremely large F-values⁵ (16,687.34 and 52,375.54, respectively, and 14,156.54 when combined). Figures 3.25-3.27 plot the relationship between error and link distance for each of the best-performing models—the relationship is plain to see. This leads to the conclusion that the best results can be obtained when an appropriate model is known for a given environment, and when the model is designed for the same distances of links being modeled. *Using models outside of their best environment and best distance coverage will result in substantial error.* This conclusion motivates hybridized models that change their approach based on the environment or length of links being modeled.

3.6.3 Distance-Hybrid Models

To understand the possible benefit of hybridized models, three hybrid models were implemented and applied to the WMP data. The WMP data was chosen because it includes the largest variety of link lengths. The first uses the Hata model (for medium cities) for links under 500m (where it is well-performing) and the Flat-Edge model (with 10 “buildings”) for longer links (hatam.flatedge10). This new model performs

⁵ The F-value is a statistic that describes the ratio between explained variance and unexplained variance. Or, put differently, the ratio of between-group variability to within-group variability.

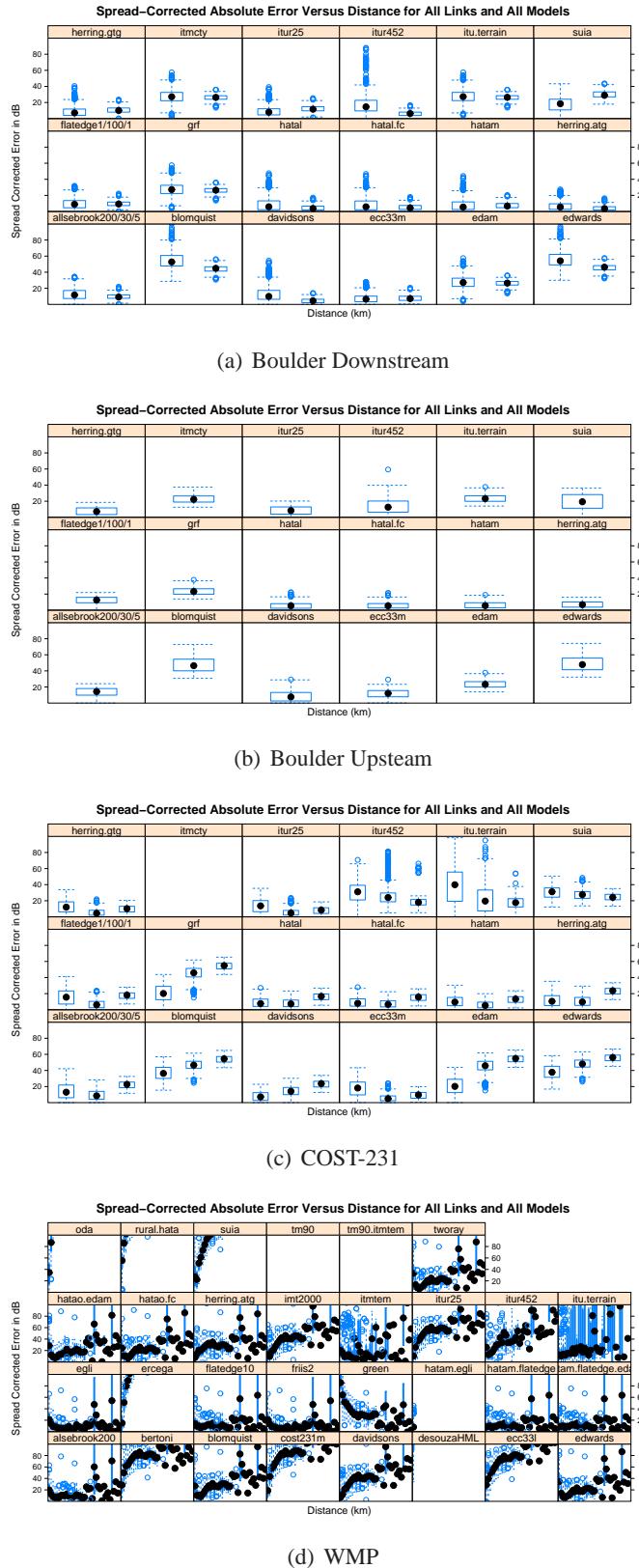


Figure 3.25: Correlation between model accuracy and link distance for each data set. Distance is bucketed by kilometer.

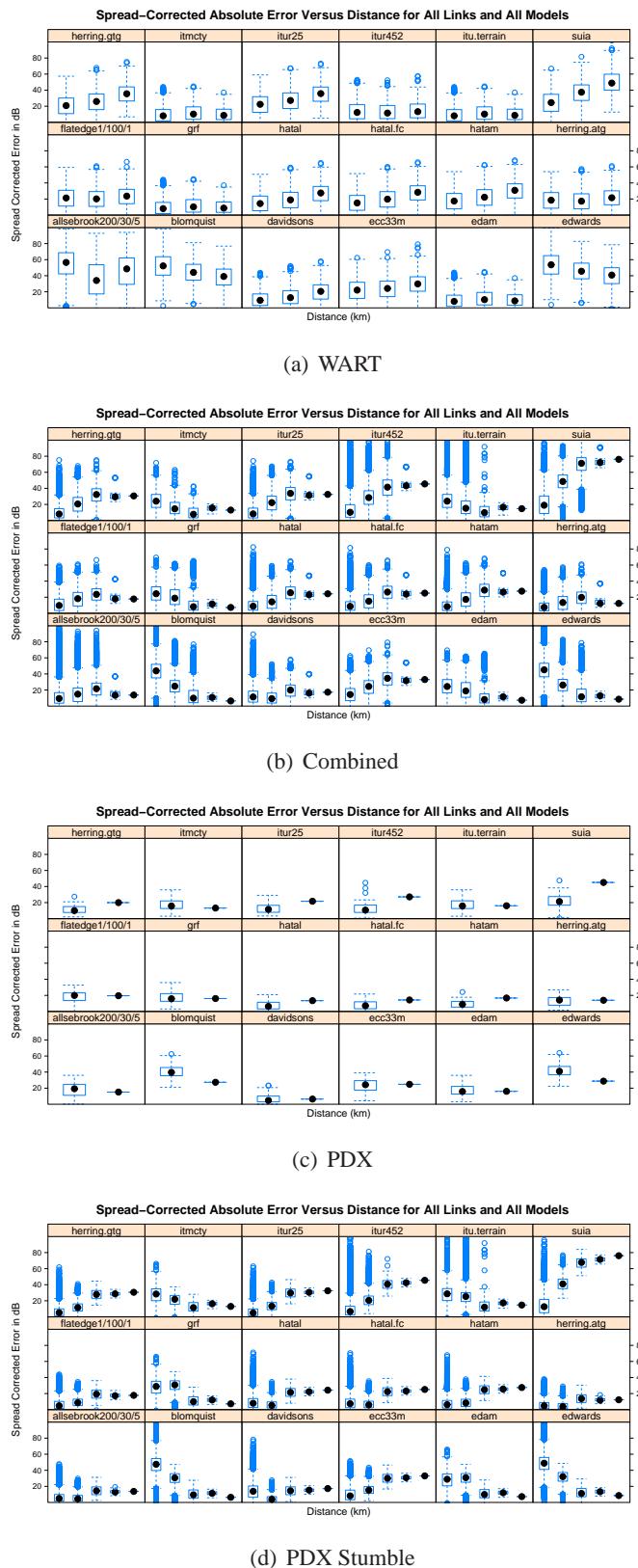


Figure 3.26: Correlation between model accuracy and link distance for each data set. Distance is bucketed by kilometer.

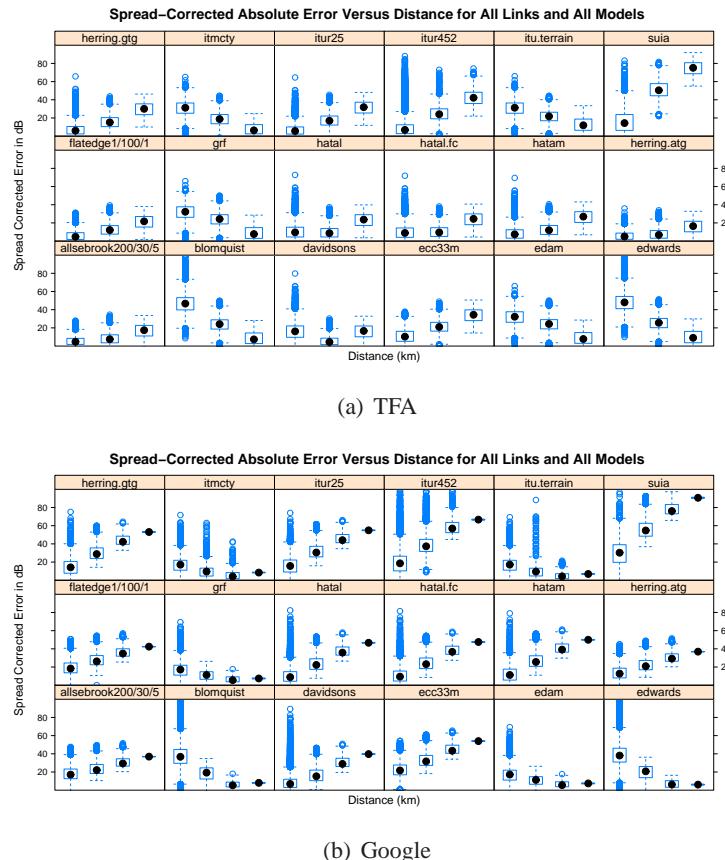


Figure 3.27: Correlation between model accuracy and link distance for each data set. Distance is bucketed by kilometer.

marginally better than all other models, producing a corrected RMSE of 14.3 dB. Very slightly better performance is achieved by combining the Hata model with the Egli Model (14.2 dB RMSE). The third combination uses the TM90 model for links less than 10 miles and the ITM for longer links (tm90.itmtem). However, this combination is not well-performing with respect to the measurements⁶. Treating this tuning and hybridization as an optimization problem with the goal of producing the best-performing configuration of existing protocols is a project for future work. Taking this approach however, one must be careful to avoid overfitting a model to the data available.

3.6.4 Practical Interpretation

As an example of what these performance results mean for real applications, consider figure 3.28, which shows a predicted coverage map for the Portland MetroFi network using two well-performing models *tuned to their best-performing configurations*. Maps with zero-mean 12 dB Gaussian noise, which approximates the expected residual error from these models, have also been included. To generate these maps, the 2 km by 2 km coverage area was divided into a 500x500 raster and each pixel is colored based on predicted received signal strength, linearly interpolated between red (at -95 dBm) and green (at -30 dBm). For each pixel, the predicted path loss from all 72 APs is computed and the maximum value is used to color the pixel.

Comparing these maps to the empirical and operator-assumed coverage maps in figure 4.1, it is clear to see that there is no consensus on what the propagation environment looks like. The Hata model may produce the picture that is closest to the measurements, but the results show that it is not the best-performing model overall. Moreover, the Allsebrook-Parsons model, which is well-performing overall, and has been tuned to its best configuration, produces a map that is in stark disagreement with reality.

Yet, the future holds promise. Consider the final column in Table 3.2, which gives the RMSE for each data set if we choose to take only the best prediction among all of the predictions made by the 30 models and their configurations. This represents one version of a minimal achievable error in a world with a perfectly hybridized model that always knows which model to use when. In this scenario, we can see a very

⁶ This approach is of special interest because it is the one advocated by the FCC in recent discussions about whitespaces transmissions in 3 GHz and below 900 MHz bands. In [141] in particular, the FCC suggests the use of the ITM for long distance predictions and the TM90 model for shorter (less than 10 mile) predictions.

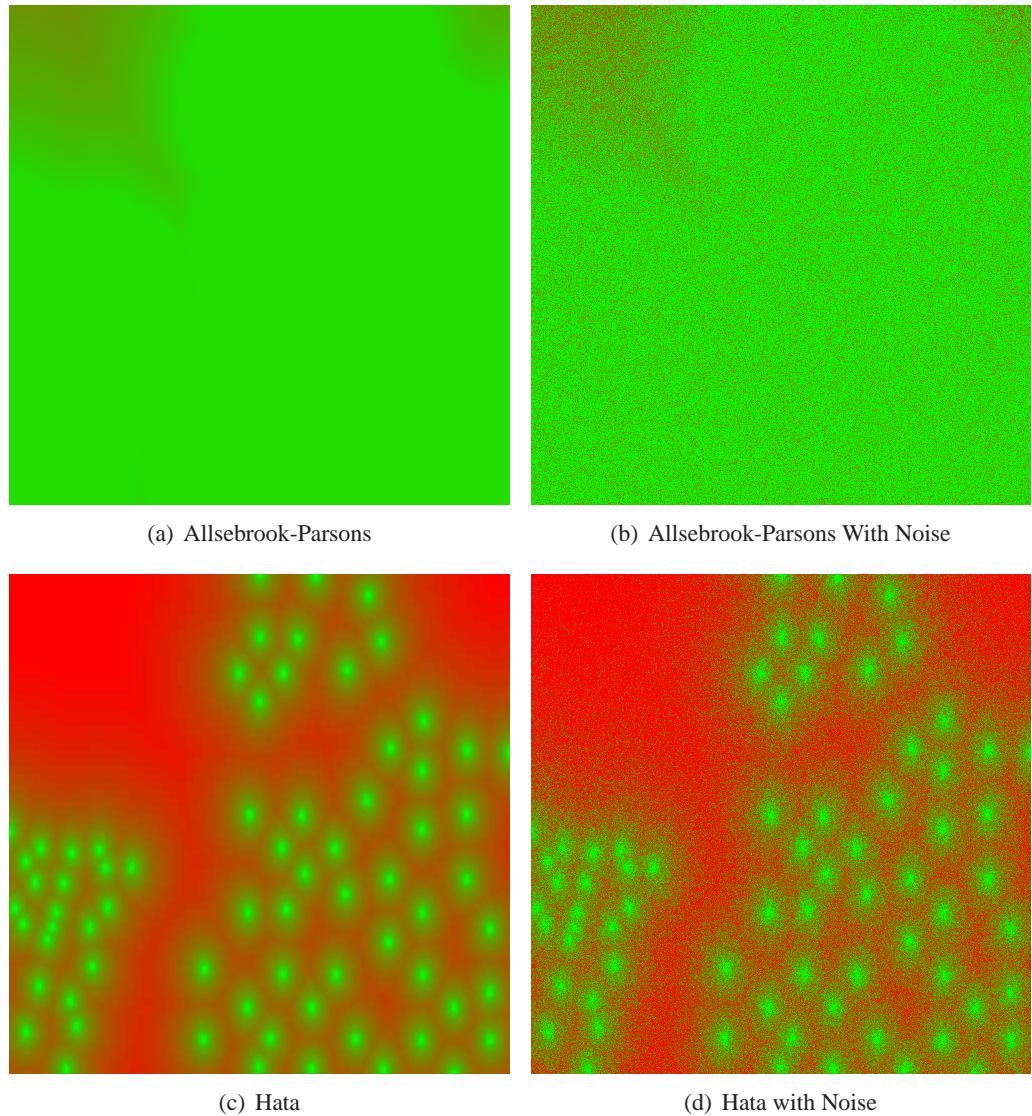


Figure 3.28: Comparison of predicted coverage maps for Portland, Oregon using two well-performing models, with and without same scale Gaussian error included. True green indicates predicted received signal at -30 dBm and true red indicates predicted received signal the noise floor (-95 dBm). Intermediary values are linearly interpolated between these two color values.

attractive bound on error—as low as 1 dB. This indicates that there is still room for improvement. If we were able to determine the situations when each model is likely to succeed, then it is reasonable to assume that it is possible to construct a single hybrid model that is more accurate than the sum of its parts. This thesis takes the perspective that an approach that marries appropriate (possibly hybridized) modeling techniques with directed measurements, will result in a better complete system than can be accomplished with either

measurement or model-tuning alone.

3.6.5 Miscellaneous Observations

This section discusses a few important miscellaneous observations based on the results above.

3.6.5.1 Modeling Directional Antennas is Challenging

One interesting additional observation from this data is that modeling path loss from directional transmitters is especially difficult. This can be seen in the fact that the data from the directional CU-WART testbed is particularly noisy. There has been at least one attempt to model this phenomenon explicitly in the past [85], but even using this correction, the error in prediction of directional propagation is still much greater than for omnidirectional transmitters. To this end an empirical supplementary model was derived from an extensive set of measurements. This model is called the EDAM and is described in detail in appendix A. Although this model is not particularly winning in the analysis here, in prior work it was shown to be better than simple models found in common simulators in at least one application [30]. While not a complete solution, EDAM is a solid first step in the direction of an appropriate modeling strategy for antenna directivity.

3.6.5.2 Models that Generate Errors

It is worth noting that some algorithms will generate error conditions when used outside of their intended coverage. If these models are given the benefit of the doubt and only used where no errors or warnings were generated, the overall performance looks better. For instance, the corrected RMSE for ITM (with parameters for a temperate environment) on the WMP data set improves from 28.2 dB to 23.1 dB if the most egregious errors are discarded (which stem from problems predicting refraction over terrain for certain terrain types, and is only 290 of 2,492 predictions) and down to 17.3 dB when only those predictions that generate zero warnings are used (which usually stem from links that are too short and are only 696 of 2,492 predictions). This is a substantial improvement—at 17.3 dB corrected RMSE, the ITM is performing on par with the best of the other models.

3.6.5.3 Prediction in Rural Environments is Challenging

In a result that appears completely counterintuitive, the rural data set is much more difficult to model than the urban data sets. To look for sources of systematic error, covariance (correlation) between “best prediction error” (the error of the best prediction from all models) and various possible factors was analyzed. There appears to be no significant correlation between carrier frequency (and therefore neither modulation scheme nor protocol) or antenna geometry. However, there is a large correlation between error and distance. It is hypothesized that the reason the WMP data is especially difficult to model may have to do with two factors: (1) Because researchers have assumed that rural environments are “easy” or “solved”, there has been substantially more work in developing (empirical) models for urban environments. The majority of state-of-the-art Rural models, on the other hand, are largely analytical and were mostly developed 30 or more years ago (i.e., the ITM) (2) This data set has an exceptionally large variety of link lengths, and as has been shown, prediction error is strongly correlated with distance for many models. However, more work is needed to confirm or deny these hypotheses.

3.7 Evaluation of Raytracing Systems

Ray-tracing (or many-ray) models, which compute the interactions between many rays and obstacles using the UTD or Finite Difference Method (FDM), are considered by many to be the state-of-the-art in path loss prediction. These models differ from the comparatively simple models discussed so far in that they consider the combined effect of constructive and destructive interference along many competing paths. These models have been widely integrated into commercial wireless planning software (e.g., [51, 187, 239]). Because of the large licensing cost of this software and significant data requirements (building models are required for outdoor prediction, and often architectural floorplans are required for indoor prediction), their use is generally excluded from all but the most demanding (and well-funded) applications. Indeed, individuals who design wireless planning tools often find that while ray-tracing methods are the highest powered models in their software, they are typically used seldomly as compared to more simplistic (often probabilistic empirical and data corrected) models such as those investigated in this section [238].

In order to understand how well ray-tracing solutions to wireless planning work, trial licenses from two well-regarded software vendors were obtained: EDX Wireless [239], and REMCOM [187]. The aim here is to predict the path loss at points on the CU campus and compare them to measurements from a set of fixed WiMax BSs at those same points. Chapter 5 describes this ground-truth data and how it was collected. This data set was chosen because of the availability of building vector data on the CU campus, while similar data was not available for the environments in which the other measurements were collected. Because ray-tracing software is clearly very dependent on the environmental data used to make predictions, three environmental data sets of increasing fidelity are used:

- Buildings as Rectilinear Shapes: in this data set, each building on the CU campus is modeled as a single polygon of an approximately correct height and footprint. This data set was manually created by EDX engineers and hence is stored in a proprietary format, which prevents use with the other (REMCOM) software. This dataset is typical of what a customer would use in planning a network [238].
- Crowdsourced Building Vector Data: for this dataset, building data was extracted from the Google Earth 3D warehouse [209]. Because the Google Sketchup software was developed in Boulder, Colorado, the 3D warehouse data for the CU campus is particularly good as it was designed by the Sketchup company itself for internal testing. This data is available through the 3D warehouse website and can be downloaded as a set of several hundred Collada files [102]. With some care, this Collada format is converted to the Stereo Lithography (STL) format, which is more widely useful. The STL format describes the building extents as a set of positioned facets (2D polygons) and their normal vectors. As part of the REMCOM Wireless Insight API [189], this STL format can be converted to a standard Shapefile, which describes buildings with polygonal shapes. The dataset constitutes a high level of fidelity which has been obtained via a large amount of work by many individuals, yet its accuracy has not been independently verified.
- Light Detection and Ranging (LiDaR) Data: High resolution LiDaR data was obtained from geography researcher Shane Grigsby at CU, who collected the data in collaboration with the CU En-

vironmental Center [155] and National Science Foundation (NSF) National Center for Airbourne Laser Mapping (NCALM) [74]. This dataset contains more than 200 million points which describe the height (and to some extent the “hardness”) of all obstacles on the CU campus in very high resolution. The dataset is described in detail at [87]. This data set constitutes the highest resolution data obtainable for an outdoor environment, and is very costly to collect.

After some discussion with engineers at EDX and REMCOM [238, 188], it became clear that state-of-the-art wireless planning software is simply incapable of working with data at the fidelity offered by the LiDaR data set. Indeed, converting such a point-data set to a raster data set is a complicated task, and converting a raster data set to a vector format is an open research question. The middle data set, derived from crowd-sourced data, was able to be used in the REMCOM software after some work to reformat it. However, this data set is far too complex for use in the EDX software [238]. Hence, results are only available for the REMCOM software using the second data set and the EDX software, using the most simplistic data. Although this substantially hinders the power of the results here, it is not feasible at present to perform a factorial analysis of the sensitivity of ray-tracing results to the fidelity of input data. Instead, this question is left for future work.

3.7.1 Case Study: REMCOM Wireless Insight and Crowd Sourced Building Models

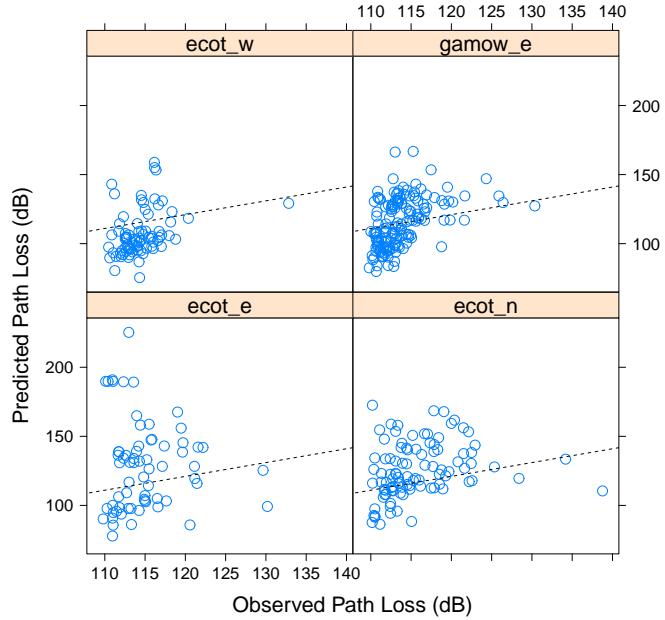
For this scenario, the WiMax measurements for the five BSs described in section 6.1 are compared to predictions at the same points. This data set contains 653 WiMax measurements from the five BSs. Using the REMCOM Wireless Insight software, a scenario that models the transmitter antennas as generic sectors with the correct beamwidth, transmission power, location, and orientation is defined. The buildings are modeled using the STL data extracted from the Google 3D warehouse and placed on a flat terrain (the CU campus has little actual elevation change). REMCOM’s proprietary “Full 3D” prediction method is used to predict a path loss value. Antenna models used are generic 120-degree sector antenna patterns without downtilt, and rotated to the correct position in the azimuth. Results are recorded in a proprietary, but parseable output format by the software.

In sum, the predictions have little bearing to the observed values. The overall RMSE is 58.27 dB. The absolute error appears to be normally distributed, with a mean of 38.34 and standard deviation of 43.91. Figure 3.29 shows a point by point comparison of the predictions to the observations, and clearly only a weak correlation is present ($\rho = 0.253$ with $p - value \approx 0$), ruling out the hypothesis that the error could be from a systematic shift. If predictions where the REMCOM software refused to make a prediction and returned the noise floor value (presumably due to an error with knife-edge diffraction computation) are removed, the RMSE is reduced to 24.09 and the mean absolute error to 18.56. Despite being a well-regarded tool for wireless prediction and planning in general, the REMCOM software performs poorly in this scenario. Although, we cannot claim that this is necessarily a representative application (and, indeed the complexity of the building data may have negatively affected results), this does demonstrate the sort of errors that might be observed in a typical application of ray-tracing software using building models derived from crowd-sourced data. In this case, the ultimate performance is on the same order or worse than much simpler path loss prediction methods described above.

3.7.2 Case Study: EDX SignalPro and Rectilinear Building Models

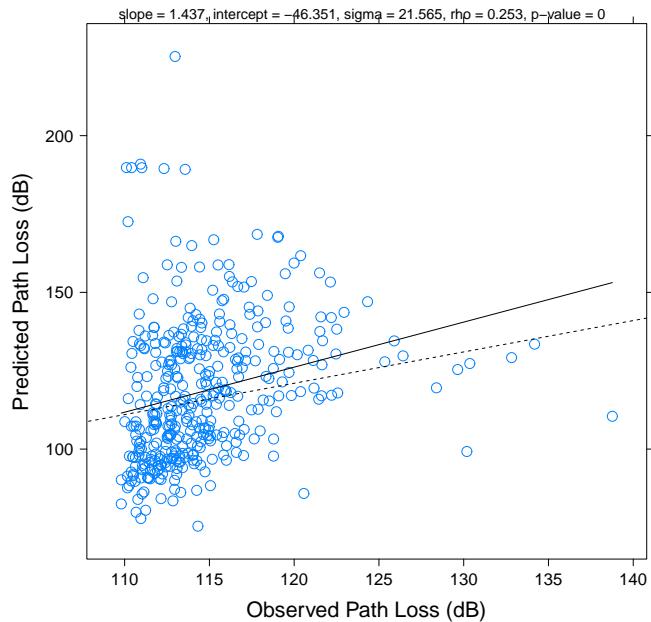
In this scenario, the EDX SignalPro software was used with simple rectilinear building models. These building models were provided by EDX engineers, who have used the CU campus for testing their software, and were described by the engineers as typical of the building models many of their clients would use [238]. Unfortunately, the EDX software was unable to make use of the crowd-sourced building models, so a direct comparison between the results from the two software packages is not possible. Figure 3.30 plots the measurements versus the predictions. Compared to the previous scenario, the predictions here are actually better correlated with the observations. A Pearson's correlation coefficient of $\rho = 0.271$ with a $p - value \approx 0$ is observed. Similarly, a linear fit has a slope of 0.838 and an intercept of 6.019 indicating that the predictions fall roughly along the same line of the observations (with a 6 dB systematic shift). However, there is still substantial residual error, with a standard error (RMSE) of 11.372. As compared to the results with the REMCOM software, this is actually quite good, and on the order of the best untuned basic propagation models. With some tuning and correction from a few measurements, it is easy to imagine

Comparison between Predictions and Observations



(a) Each AP

Comparison between Predictions and Observations



(b) All APs Combined

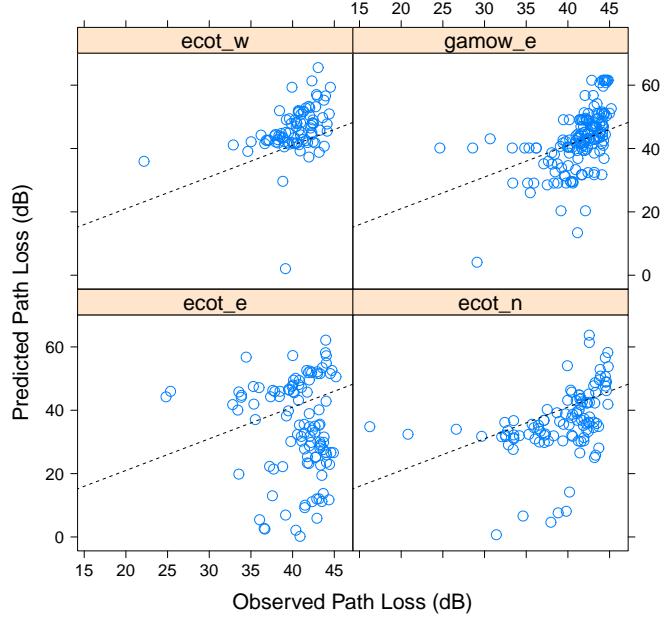
Figure 3.29: Correlation between predicted values and observed values using REMCOM ray-tracing software and WiMax data. The dotted line has a slope equal to 1, which the data points would fall upon if the predictions were perfect. Deviations from this line indicate the magnitude of error. Fit and correlation statistics are given for the aggregate (all APs) predictions. To simplify the plot, points where the prediction software refused to make a prediction have been censored, as well as locations where there was no signal observed.

that this software could produce results of the same order of accuracy as those with an explicit data fitting or model tuning approach described in previous sections. However, this accuracy comes at the cost of increased data requirements. Determining the shape and height of every building in a study area to create a rectilinear building model may be very time consuming, even as compared to the time required to make some number of measurements.

3.7.3 Summary of Results

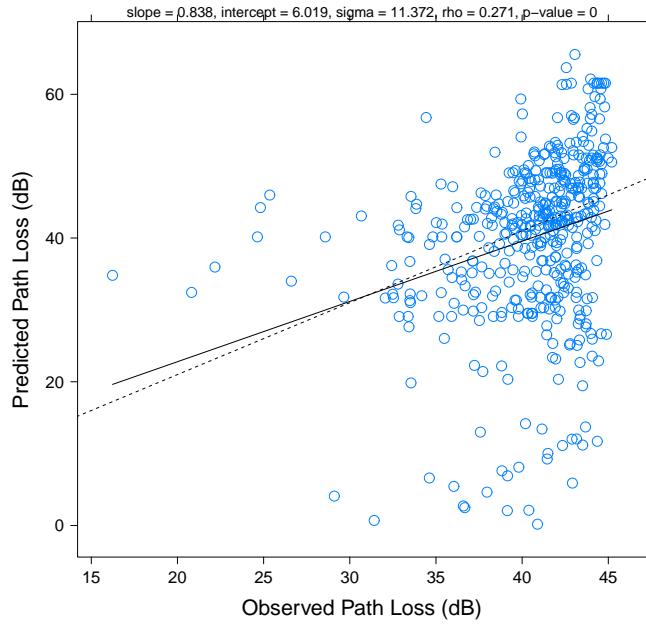
In the two scenarios studied, an unintuitive result was produced: the scenario with lesser building data fidelity outperformed that with higher resolution building models. This may be due to the fact that the complex polygons produced by the crowd-sourced building data produced many diffraction errors that were not present in the simpler data. Unfortunately, state-of-the-art ray-tracing tools are simply incapable of using high resolution building data collected from a LiDaR scan, and hence need substantial improvement in their efficiency and preprocessing algorithms to work with data of this fidelity. In an application where it is more costly to make direct measurements of the radio propagation than it is to gather data about the obstacles and buildings in the environment, then the use of these complex ray-tracing models may be justified. For basic planning purposes, their fidelity is likely sufficient. However, they do not appear to be substantially more accurate than a well-chosen simple path loss model (e.g., something from the Hata family). Some experts would argue that their fidelity is pendulously tied to decisions about how to model the diffraction and absorption of building construction materials (a classic example being buildings with radio-opaque glass, which may act as Faraday cages), and the choice of which subset of rays are used for calculation [142]. Although a great deal more work is needed to generally understand the relationship between the performance of ray-tracing approaches to path loss modeling and the fidelity of input data, these two case studies makes a compelling argument for the value of an increased focus on empirical coverage mapping as opposed to greater complexity in computation and environment modeling, whose performance in the general case is not well understood.

Comparison between Predictions and Observations



(a) Each AP

Comparison between Predictions and Observations



(b) All APs Combined

Figure 3.30: Correlation between predicted values and observed values using EDX ray-tracing software and WiMax data. The dotted line has slope equal to 1, which the data points would fall upon if the predictions were perfect. Deviations from this line indicate the magnitude of error. Fit and correlation statistics are given for the aggregate (all APs) predictions. To simplify the plot, points where the prediction software refused to make a prediction have been censored, as well as locations where there was no signal observed.

3.8 Discussion

This chapter has presented the first rigorous evaluation of a large number of path loss models from the literature using a sufficiently representative data set from real (production) networks. Besides providing guidance in the choice of an appropriate model when one is needed, this work was largely motivated by a need to create baseline performance values. Without an existing well-established error bound for these approaches, it is impossible to evaluate the success (or failure) of more complex approaches to path loss modeling (and coverage mapping). For the models implemented here, and the data sets analyzed, it is possible to say that *a priori* path loss modeling will achieve, at least, 8-9 dB RMSE in urban environments and \approx 15 dB RMSE in rural environments. This is true almost regardless of the model selected, how complex it is, or how well it is tuned. And, this bound seems to agree with prior work at other frequencies in similar environments that have also produced results with RMSE in the neighborhood of 9 dB (e.g., [73, 60]).

Direct approaches to data fitting, such as a straight line fit to the log/log relationship between path loss and distance, produce a similar level of error: 8-9 dB for urban environments and \approx 15 dB for rural environments. Fits of this quality can be obtained after only 20-40 measurements. Hence, whether a network operator does a small random sampling and basic fit, or carefully tunes an *a priori* model to their environment, they can still expect predictions that are only accurate to within 3 to 5 orders of magnitude. This result motivates continued work on more advanced methods and creates a well-defined measure of success for these more advanced models in terms of overall prediction accuracy: if a model can produce a coverage map where the variation (error) between the measurements and model is less than 12 dB, than we can say with confidence that it is outperforming an equivalent map generated using state-of-the-art *a priori* modeling routines. Moreover, a map with less than 8-9 dB error can be said to be better than can be expected with either hand tuned per-environment modeling or exhaustive measurement and explicit (straight line) fitting.

Among the most important outcomes of this work is a set of guidelines for researchers, which can help provide direction in the complicated landscape of path loss prediction models. As a general rule, when it is feasible to make direct measurements of a network, one should do so. It has been shown that a small

number of measurements can have substantial power in terms of tuning the models studied and in fitting parameters for basic empirical models. When it is not possible to make measurements of a network, the careful researcher should choose from standard well-accepted models such as Okumura-Hata or Davidson, which generally have the least systematic skew in predictions, and are among the best-performing models overall. In simulation studies, a repeated-measures approach is advocated, where stochastic models are used in a repeated-measures/Monte Carlo experimental design, so that a realistic channel variance can be modeled. For this application, the recent proposal of Herring appears to be a good choice, or for the greatest comparability, the Hata model with stochastic lognormal fading. Although there are a large number of models from which to choose, this work shows that in many cases the most important factors that a researcher should consider are having a realistic expectation of error, and choosing a model that enables repeatability and comparability of results.